Uncertainty quantification in the simulation of road traffic and associated atmospheric emissions in a metropolitan area

Ruiwei Chen

UNIVERSITÉ PARIS-EST

ÉCOLE DOCTORALE SCIENCES, INGÉNIERIE ET ENVIRONNEMENT

THÈSE

présentée pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ PARIS-EST

Spécialité: Sciences et Techniques de l'Environnement

Uncertainty quantification in the simulation of road traffic and associated atmospheric emissions in a metropolitan area *

Ruiwei CHEN

Thèse soutenue le 25 mai 2018 devant le jury composé de :

Dr Ludovic LECLERCQ	IFSTTAR	*Président & Rapporteur*
Dr Carlos CANUDAS DE WIT	CNRS	*Rapporteur*
Dr Carlos BORREGO	University of Aveiro	*Examinateur*
Dr Karine KATA SARTELET	ENPC	*Directrice de thèse*
Dr Vivien MALLET	INRIA	*Encadrant*
Dr Vincent AGUILÉRA	DIRIF	*Co-encadrant*
M. Jérémy VIGNERON	AIRPARIF	*Invité*

∗. Quantification d'incertitude en simulation du trafic routier et de ses émissions atmosphériques à l'échelle métropolitaine

This thesis is dedicated to my beloved parents, for their love, endless support, encouragement and sacrifices. To my mother, who gave me life and has prepared me to face challenges with faith and optimism. To my father, who is always in the deepest part of my heart. I miss him so much.

Remerciements

Je remercie l'Agence Nationale de la Recherche (ANR) d'avoir financé mes travaux de thèse.

Je remercie dans un premier temps les membres du jury pour avoir accepté d'évaluer mes travaux, pour avoir assisté à ma soutenance et pour leurs critiques et suggestions pertinentes lors de la soutenance. Je remercie professeur Carlos Canudas de Wit et professeur Ludovic Leclercq d'avoir accepté d'être les rapporteurs de ma thèse et de m'avoir fait part de remarques constructives. Je remercie expressément Ludovic Leclercq de s'être déplacé à Paris et d'avoir présidé ma soutenance. Je remercie professeur Carlos Borrego et monsieur Jérémy Vigneron pour avoir accepté d'être les examinateurs. I thank Prof. Borrego for sparing a time from his busy schedule and coming to Paris from Portugal.

Je voudrais remercier Christian Seigneur et Pietro Bernardara de m'avoir accueillie au sein du CEREA. Je n'oublierai pas le tiramisu de novembre pendant trois ans, et je n'oublierai pas non plus mon premier Escape Game à Paris. Je remercie Karine Sartelet pour avoir assuré la direction de cette thèse. J'apprécie beaucoup ses attitudes rigoureuses envers la recherche scientifique. Je lui suis reconnaissante pour ses conseils et encouragements non seulement pour ma thèse mais aussi pour l'après-thèse.

Je tiens à remercier mes encadrants de thèse, Vincent Aguiléra et Vivien Mallet. C'est un grand plaisir de travailler avec eux. Je les remercie pour la confience qu'ils m'ont toujours accordé. Je remercie Vincent pour ses qualités scientifiques et techniques sur l'affectation dynamique du trafic. Je lui suis reconnaissante de ses conseils pertinents pour l'ensemble de la rédaction des articles et de la thèse. Je le remercie également pour ses conseils et son aide pour l'après-thèse. Je remercie Vivien pour m'avoir proposé cette thèse et pour l'avoir dirigée. Je tiens particulièrement à souligner sa disponibilité et ses qualités pédagogiques, scientifiques et techniques. Je lui suis très reconnaissante de son expertise, ses conseils, son esprit critique et son attitude positive. Je le remercie pour ses encouragements tout au long de la thèse, et surtout pendant les moments les plus difficiles. Je le remercie pour sa patience. Je n'oublierai pas ses corrections dans les moindres détails de mes articles et de mon manuscrit, même s'il me connaît très bien comme une experte des *phrases compliquées*. J'apprécie également son enthousiasme pour la recherche scientifique, qui m'encourage à poursuivre des travaux de recherche après la thèse.

Ma thèse a pu être financée grâce au projet ANR ESTIMAIR. Je remercie la Ville de Clermont-Ferrand pour les données brutes de comptage du trafic. Je remercie également le Syndicat Mixte des Transports en Commun (SMTC) de l'agglomération clermontoise

pour nous avoir fourni la demande du trafic et le réseau modélisé pour Clermont-Ferrand, et pour leur aide et compétence sur l'affectation statique du trafic. De plus, je remercie l'ensemble des chercheurs et ingénieurs du projet. Je remercie Florian Cohn et David Poulet pour avoir préparé et traité les données du trafic et le réseau modélisé de Clermont-Ferrand. Je remercie Fabien Brocheton et Lionel Soulhac pour les échanges scientifiques lors des réunions du projet, ainsi que leurs conseils et remarques pertinents sur mes présentations, mes articles et ma thèse. Je remercie spécialement Fabien de son aide pour avoir assuré l'obtention de données de bonne qualité.

Je remercie mes collègues du CEREA pour les échanges conviviaux au cours des pauses déjeuners et des goûters du 16h. Je remercie Véronique pour avoir assuré les démarches administratives pendant ma thèse. Je remercie Sylvain pour son soutien informatique et son aide technique sur le calcul distribué. Je remercie Youngsoeb pour son soutien informatique, ses conseils et ses encouragements continus. De plus, ses makis coréens faits maison me manqueront sans doute. Je remercie Yelva pour ses conseils et aides concernant le calcul des émissions du trafic. Je remercie Alban, Anthony, Jean-Matthieu, Marc, Mounir et Valentin, pour leur soutien et compréhension quand j'avais besoin de clusters supplémentaires pour mes simulations. Je n'oublierai pas non plus les gâteaux du chef pâtissier Valentin. Je remercie Shupeng pour ses encouragements et son soutien non seulement pendant mes trois ans et demi de thèse mais également pendant l'après-thèse. Je remercie Stéphanie, Juliette, et Marwa pour les moments joyeux ou difficiles partagés ensemble dans le même bureau. Je remercie Laëtitia, Cécile, Lya, Palmira, Marwa et Carole, pour les moments inoubliables partagés lors des pauses, du sport, ou des sorties entre filles. Je remercie en particulier Carole, pour ses accompagnements et encouragements continus, pendant les moments les plus heureux ou bien les plus durs. J'ai eu la chance de pouvoir partager mes sentiments avec elle lors des repas, ou bien au tour d'un gâteau délicieux.

Je remercie Raphaël, Janelle, Bao, Nourhène, Frédéric et Antoine pour m'avoir accueillie lors de mes passages à l'INRIA. Je n'oublierai pas les moments conviviaux partagés, notamment pendant mes derniers mois de thèse.

Je pense à remercier les personnes qui m'ont introduite dans l'univers de la recherche sur le trafic routier et ses émissions. Je remercie mon encadrant de master, Prof. Qiyuan Peng, qui m'a encouragée à poursuivre une thèse de doctorat, pour ses soutiens et encouragements. Je remercie l'ensemble du laboratoire LICIT pour m'avoir accueillie lors de mon stage d'ingénieur en 2011. Je remercie notamment Nicolas et Ludovic, qui m'ont introduite à la carrière de recherche scientifique sur le trafic routier, par leurs qualités scientifiques et leur dynamisme. Je remercie Thamara pour les moments joyeux partagés

à l'ENTPE, à Paris et lors des conférences TAP. Je remercie spécialement Xiaoyan pour ses conseils, ses encouragements et son soutien depuis toujours. Elle est comme ma grande soeur en France. Je pense aussi au LVMT et je voudrais remercier Fabien Leurent et Nicolas Coulombel pour leurs remarques sur la modélisation et l'affectation du trafic, ainsi que leurs conseils concernant l'après-thèse. Je tiens à remercier particulièrement Luc Charansonney pour ses aides et conseils sur le modèle LADTA et LTK, ainsi que d'autres outils (QGIS, VISUM, etc).

Je n'oublie pas mes amis du groupe de semi-marathon. Un grand merci à Xiao, Yifan et Yi, non seulement pour les moments joyeux partagés lors de nos sorties du week-end, mais également pour leur soutien, leurs conseils et leurs encouragements pour le travail et la vie quotidienne en France. Je remercie Pierre, Sherry, Alex et Fanny, pour leurs accompagnements et les moments joyeux partagés depuis nos études à Centrale Lyon. Je remercie spécialement Fanny, qui me soutient depuis toujours même si nous ne vivons pas dans la même ville. Je la remercie pour ses relectures et corrections du français quand j'en avais besoin. Je la remercie pour les échanges sur des sujets divers comme la recherche, l'écologie, le sport, la culture, etc. J'apprécie beaucoup ses attitudes rigoureuses et positives envers la recherche scientifique. Je pense aussi à mes amies en Chine, Ye, Yi, Juan, Xiaojiao, Zao, et Le, avec qui je peux tout partager malgré la distance. Je les remercie pour leur soutien et accompagnement depuis plus de dix ans.

J'adresse mes plus sincères remerciements à ma famille: mes parents, mes beaux parents, tous mes proches, qui m'ont accompagnée, aidée, soutenue et encouragée sans condition, tout au long de la réalisation de cette thèse. Je remercie en particulier ma mère, qui est venue depuis la Chine pour assister à ma soutenance. Maman, sans toi, je n'aurais pas pu devenir ce qui je suis. Je suis fière d'être ta fille.

Enfin, je tiens particulièrement à remercier mon mari, Sicong, avec qui j'ai commencé nos aventures depuis plus de dix ans. Ta patience, ta compréhension, tes encouragements et ton amour m'ont soutenue, et particulièrement durant ces trois ans et demi de thèse où j'ai passé peut-être plus de temps avec mon ordinateur qu'avec toi. Tu as toujours été mon inspiration. Je te remercie d'être toujours à mes côtés, pendant les moments les plus joyeux et les périodes les plus sombres. Nous avons tous les deux réussi nos thèses et nous allons découvrir notre nouvelle vie ensemble. J'ai hâte de continuer nos aventures dans l'avenir.

Du fond du coeur, merci à tous!

Abstract

This work focuses on the uncertainty quantification in the modeling of road traffic emissions in a metropolitan area. The first step is to estimate the time-dependent traffic flow at street-resolution for a full agglomeration area, using a dynamic traffic assignment (DTA) model. Then, a metamodel is built for the DTA model set up for the agglomeration, in order to reduce the computational cost of the DTA simulation. Then the road traffic emissions of atmospheric pollutants are estimated at street resolution, based on a modeling chain that couples the DTA metamodel with an emission factor model. This modeling chain is then used to conduct a global sensitivity analysis to identify the most influential inputs in computed traffic flows, speeds and emissions. At last, the uncertainty quantification is carried out based on ensemble simulations using Monte Carlo approach. The ensemble is evaluated with observations in order to check and optimize its reliability.

Keywords : Dynamic traffic assignment, traffic emissions, metamodeling, global sensitivity analysis, uncertainty quantification, ensemble simulation

Résumé

Ce travail porte sur la quantification d'incertitude dans la modélisation des émissions de polluants atmosphériques dues au trafic routier d'une aire urbaine. Une chaîne de modélisation des émissions de polluants atmosphériques est construite, en couplant un modèle d'affectation dynamique du trafic (ADT) avec un modèle de facteurs d'émission. Cette chaîne est appliquée à l'agglomération de Clermont-Ferrand (France) à la résolution de la rue. Un méta-modèle de l'ADT est construit pour réduire le temps d'évaluation du modèle. Une analyse de sensibilité globale est ensuite effectuée sur cette chaîne, afin d'identifier les entrées les plus influentes sur les sorties. Enfin, pour la quantification d'incertitude, deux ensembles sont construits avec l'approche de Monte Carlo, l'un pour l'ADT et l'autre pour les émissions. L'ensemble d'ADT est évalué et amélioré grâce à la comparaison avec les débits du trafic observés, afin de mieux échantillonner les incertitudes.

Mots clés: Affectation dynamique du trafic, émission du trafic routier, méta-modélisation, analyse de sensibilité globale, quantification d'incertitude, simulation d'ensemble

Contents

Preamble

Context and background

Increasing attention has been paying to air quality at urban area in order to study the exposure of city dwellers to air pollution. To monitor air quality, forecast pollution episodes, inform citizens, and assist policy maker in controlling air pollution in the short and long terms, urban scale air quality models down to street resolution are developed and used by atmospheric environmental scientists and agencies. With the help of air quality models, different scenarios can be designed in order to understand what kind of measures can be proposed to decision makers, medias, and citizens in order to decrease air pollutant concentrations, and more importantly, to prevent pollution episodes from happening. However, there are uncertainties in the estimated pollutant concentrations from the models, due to uncertainties in input data, parametrizations, and numerical limitations of the models.

Emissions are one of the most important inputs of an urban air quality model, because they can directly impact the model performance in estimating atmospheric pollutant concentrations at urban scale. The atmospheric pollutant emissions due to on-road traffic are one of the key uncertainty sources in urban air quality simulations [Soulhac et al., 2012]. In fact, on-road traffic emissions are one of the main sources of atmospheric pollution at urban scale. According to the 2017 European Environment Agency (EEA) report [EEA, 2017], road transportation contributes 39 % of NOx emissions in Europe, as the first source among all sectors. In addition, road transportation is also the second biggest contributor of CO and BC (black carbon) emissions, representing 20 % and 29 %, respectively. Road transport is also responsible for 11 % of the particulate matter (PM) emissions in Europe. In Paris Region (*Région Île-de-France*), on-road traffic contributes 56 % of NOx emissions, 28 % of PM_{10} and 35 % of $PM_{2.5}$ emissions [Airparif, 2014]. This is the main reason why many countries and regions impose legal requirements for on-road vehicles and engines, and introduce *emission standards*. Those emission standards set emission thresholds for vehicles and engines, which may require better emission control technology to comply. In Europe for example, emissions from road vehicles have been controlled by European legislation since the 1970s. In order to meet the increasingly stringent requirements of the legislation, vehicle manufacturers have continually improved engine technologies and have introduced various emission-control systems. Road vehicles are usually classified according to their level of emission

control technology (pre-Euro, Euro1, Euro2, ..., Euro6c, etc.), which is actually defined in terms of the emission legislation with which they are compliant [EEA, 2016].

Modeling of atmospheric pollutant emissions

In fact, on-road traffic emissions are estimated from numerical models. On-road emission is the production of (i) activity data and (ii) emission factor. In the on-road transportation sector, *activity data* refers to total vehicle-kilometers (in veh·km) on the area that we want to study, during a given period [Boulter and McCrae, 2007; EEA, 2016]. The *emission factors* (in $\mathrm{g\,km^{-1}\,veh^{-1}}$) for road traffic are estimations of vehicle average real-world emissions, according to different type of vehicle, emission standard technologies, type of road and traffic conditions. In order to analyze the uncertainty of estimated on-road traffic emissions, it is therefore necessary to analyze the uncertainty due to both categories of inputs: (i) the vehicle-kilometers and (ii) the emission factors. For an annual estimation of emissions through out a country, vehicle-kilometers can be roughly obtained from a country's statistics. However, for the emission estimation at urban scale down to street resolution, a traffic assignment model should be coupled with an emission model in order to compute traffic flow at street-level at a required temporal resolution. As reviewed in [Boulter and McCrae, 2007], "uncertainty in emission estimates remains strongly associated to the traffic-related parameters, which are themselves often highly uncertain." On the one hand, traffic volume information is required for estimating emissions. At urban scale, traffic volumes are modeled by traffic models, for which the uncertainty analysis is a new topic. On the other hand, emission factors are modeled as a function of vehicle kinematics (vehicle speed), which can also be simulated by traffic models. Therefore, uncertainty quantification in on-road traffic models is an important part for analyzing uncertainty in traffic emission estimations.

At urban scale, macroscopic traffic assignment models (TA models) are used to compute traffic flow and vehicle travel speed at street resolution. The traffic flow is defined as the number of vehicles passing through a point during a unit period of time. Traffic assignment aims at determining the network traffic flows and travel times according to network users' route choices when they travel from their origin to their destination. It can also be considered as an economical equilibrium between the demand and the supply. The *traffic demand* is represented as an Origin-Destination matrix (O-D matrix), describing the total flux from original zones to destination zones. The *supply* is the road network with limited capacities to absorb all the demand. One of

the main hypotheses for traffic assignment problems is that every network user makes their route choice by minimizing their own travel cost, such as travel time, toll price, etc. At equilibrium when every traveler succeeds in finding such a route, all used routes associated with the same origin and destination should have the same minimum cost, so that there is no possibility for travelers to shift to another route [Wardrop, 1952]. There are static traffic assignment models and dynamic ones. The static models compute average traffic flow and vehicle travel time during a given period, without taking into account time-dependent interactions of traffic flows. The dynamic models add the time dimension into the model. Dynamic models can give time-varying traffic flows and speeds with a finer temporal resolution compared to static models. That is the main reasons why we chose a dynamic traffic assignment (DTA) model in this PhD study. The uncertainty lying in the inputs of the DTA models may then propagate to the estimated traffic outputs (traffic flow, vehicle kinematics, etc.), and then influence the simulated on-road traffic emissions from emission models. Therefore, an uncertainty analysis on DTA simulation needs to be carried out first.

Then in order to estimate time-dependent street resolution emission due to on-road traffic at urban area, the DTA model is coupled with an emission model to build a modeling chain. The uncertainties lying in traffic flows and vehicle speeds, due to inputs of traffic models, are propagated to the final emission simulation, via the modeling chain. In addition, the calculation of emission factor requires other inputs and there is uncertainty lying in these inputs, too. For example, the emission factors for a given pollutant vary based on vehicle fleet composition: vehicle type (passenger car, light commercial car, heavy duty vehicles, motorcycles), motorization type (diesel, petrol, hybrid, electric), vehicle technology (Euro standard), etc. This fleet composition should be taken into account in the distribution of traffic volume, too. The estimated emissions at urban area are very sensitive to the fleet composition [Grassot et al., 2012; Carteret et al., 2014]. Therefore, it is important to analyze the uncertainty lying in the final estimated on-road traffic emissions, due to inputs of traffic assignment model, as well as the inputs of the emission model. Eventually, the uncertainty in these inputs is propagated to the uncertainty of the final computed on-road emissions.

Uncertainty quantification

The uncertainty quantification aims to quantify the uncertainty of outputs resulting from numerical models. Brief definitions of the *uncertainty* and *uncertainty quantification* are presented below. Unlike the deterministic simulation with known inputs, the

inputs and outputs are both considered as random variables U and X in the process of uncertainty quantification. Let us denote the model as $X = f(U)$. A *deterministic* simulation can be denoted as $x = f(u)$ where x and u are *realizations* of X and U. There are many criteria to evaluate the performance of the model f. Let us assume that the *true* value of X is an observed value o. If we consider that there is uncertainty in the observation, the observation can also be considered as a random variable O. However, here we assume that the observation is perfect and we ignore the error lying in observations. Therefore, in deterministic simulation, we can evaluate the model f with the error e, which is defined as the distance between x and the observation o:

$$e = d(x, o), \tag{1}$$

where $d(\cdot, \cdot)$ is a measure of distance (Mean Squared discrepancy for example). Another criterion to evaluate the performance or the robustness of the numerical model is the *a priori uncertainty*: $\sigma_X = \sigma$. It is defined only from the *a priori* probability distribution of X without taking into account the observations. There is also the *a posteriori uncertainty* which is estimated given the observation o. In our study, we only focus on the *a priori uncertainty* and the term "uncertainty" in this dissertation refers to *a priori uncertainty*. Therefore, the objective of uncertainty quantification is to determine the probability distribution of X. If the model f is used to forecast a target value, then we may calculate the expectation of X and denote it as $E(X)$. The measure of the uncertainty can then be expressed as its variance $Var(X) = E((X - E(X))(X - E(X))^T)$. To do so, we need to use a probabilistic approach and introduce uncertainties in inputs U, in order to obtain the probabilistic distributions of output X computed by $f(U)$.

In practice, an ensemble of deterministic simulations is generated in order to get the distribution of the output variable X. With a Monte Carlo approach of size n for example, a sequence of inputs $(u_i)_{i=1,2,\ldots,n}$ is drawn according to a given probabilistic distribution. Therefore, n deterministic simulations are carried out with each of the input u_i: $x_i = f(u_i), i = 1, \ldots, n$. Therefore, the distribution of the output sequence $(x_i)_{i=1,\ldots,n}$ can be considered as samples from the random variable X. Its variance or the standard deviation can be considered as a measure of the uncertainty of X, due to the given uncertainty lying in the input U.

If the *error* can already evaluate the performance of the model f with deterministic simulations, why do we make use of a probabilistic approach to quantify the uncertainty which is computationally costly? In fact, in the context of estimating uncertainty in DTA simulation and emission estimation in urban area, the main outputs are traffic flows, vehicle speeds, and atmospheric pollutant emissions at street resolution. On the

one hand, there are no available observations of the traffic flows/speeds in all the streets. On the road network, there may be loop detectors that measure the number of vehicles passing a point of a lane during a unit of time and therefore give traffic flow of the street. However, not all the streets on the network are equipped with detectors. On the other hand, the emissions due to on-road traffic cannot be measured directly in atmosphere. There are interactions between the pollutant emitted by road traffic and other chemical species in the environment. The concentrations measured by air quality monitoring stations cannot directly measure the emission. Therefore, the *errors* can not be measured on all the streets of the studied urban area. Instead, the measure of *a priori* uncertainty does not depend on the observation and the uncertainty estimation can be carried out on all the roads of a network. In addition, the deterministic simulations do not take into account the uncertainty in inputs. With probabilistic approach, we are able to give a possible distribution of the inputs U, and then quantify the uncertainty in output X by estimating its probability distribution, on all the roads of the studied urban area.

Objectives

The main motivation of this PhD work is to quantify the uncertainty in the simulation of on-road emissions due to the inputs of a dynamic traffic assignment model and the inputs of the emission factor model, at urban scale with high spatial and temporal resolutions. Within the context of emission estimation and uncertainty quantification, the objectives of this dissertation are as follows.

1. To apply a DTA model at street resolution to a real-world network and evaluate the model performance with loop detector observations.
2. To build a modeling chain of on-road emission estimation at metropolitan scale, at high spatial and temporal resolutions. This involves evaluating a state-of-the-art dynamic traffic assignment (DTA) model applied in an urban area. The DTA model computes time-varying traffic flows and speeds at metropolitan scale at high spatial and temporal resolution. Then the time-varying traffic information is coupled with an emission model in order to compute on-road emissions at street level with the same spatial-temporal resolution.
3. To identify the most influential factors in the DTA model and the modeling chain that affect the final computed traffic flow and its emissions at urban scale down to street resolution. This involves qualitative and quantitative sensitivity analysis on (i) the computed traffic flows and speeds with respect to the DTA model inputs

and (ii) the computed on-road traffic emissions with respect to both inputs of the DTA model and emission factor model in the modeling chain.

4. To quantify the uncertainty of the computed traffic flows and its emissions for a full city scale down to street resolution. To this end, two ensembles of simulations are to be generated: (i) an ensemble of dynamic traffic assignment simulations in order to quantify the uncertainty of the DTA simulation in an urban area during a long period (one month), with time resolution down to every 15 minutes. The inputs of DTA models are to be perturbed with given probabilistic distributions. Therefore, the distributions of the computed traffic flows and speeds can be deduced based on the ensemble. (ii) An ensemble of on-road traffic emissions at street resolution, based on the built ensemble of traffic assignment during the same period of time with the same temporal resolution. It is then possible to analyze the propagation of the uncertainty in the modeling chain from traffic assignment model to the final computed on-road traffic emissions, due to the uncertainty of traffic assignment inputs.

In order to achieve the objectives 3 and 4 listed above, a large number of model evaluations are required. Nevertheless, the additional time dimension in the traffic assignment model makes it more accurate in modeling time-varying traffic features, yet more computationally costly. The DTA model used in this PhD work is proved to be tractable in an operational context for a large-size traffic network with a model evaluation time around several hours [Aguiléra and Leurent, 2009]. However, the model evaluation time still needs to be reduced. For example, in the quantitative global sensitivity analysis, the order of magnitude of the number of model evaluations is typically $10^4 \times k$, where k is the dimension of inputs. In order to tackle with the computational burden, this dissertation proposes a metamodeling framework in order to build a metamodel (or surrogate model) of the DTA model applied to a metropolitan area. The metamodel computes similar traffic assignment results (traffic flows and speeds) at the street resolution with high temporal resolution when compared with those computed by the original model, but with a much lower computational cost.

Organization of the thesis

The dissertation is organized as follows:

Chapter 1 presents literature review, and gives introduction of models and methods used in this PhD work. Firstly, a review of traffic assignment models is given. Secondly, a review of metamodeling method is presented, especially its application in the field of

traffic assignment model on large-scale network. Thirdly, a review of state-of-the-art emission models is given. Fourthly, a brief review regarding main methods of sensitivity analysis and uncertainty quantification is given, especially their applications in traffic modeling and on-road emission estimations.

Chapter 2 presents the most important definitions, assumptions and formulations of the dynamic user-equilibrium (DUE) traffic assignment problem at first, followed by the presentation of a DUE-based traffic assignment model named LADTA (for Lumped Analytical DTA) [Leurent, 2003] with its inputs, outputs and assumptions. Then, the framework of the implementation of LADTA model, LTK (for ***L****ADTA* ***T****ool* ***K****it*) [Aguiléra and Leurent, 2009; Aguiléra, 2014], is presented. Finally, a DTA simulation using LADTA is presented with an example of a two-link network.

Chapter 3 presents the application of LADTA model to a metropolitan scale network, and qualitative sensitivity analysis of the computed results with respect to LADTA inputs. The case study is for the whole agglomeration of Clermont-Ferrand (France). It has about 200,000 residents in an area of about 300 km^2. LADTA computes continuous traffic flow at street resolution during one day. The traffic flows computed by LADTA are compared with those computed by the widely-used static traffic assignment model VISUM, during evening peak hour (17:00 - 18:00) of a working day. The computed traffic flows by both models are also compared with loop detector measurements in order to evaluate the models performance.

Chapter 4 presents a metamodeling method for DTA model applied to a real-world network. A case study is carried out for LADTA model applied in the agglomeration of Clermont-Ferrand. Uncertain inputs in DTA simulation with LADTA model are identified: (i) time-varying traffic demand, (ii) spatial uncertainty in the traffic demand, (iii) spatial uncertainty in network parameters. Then a metamodel based on these uncertain inputs is built. Thirdly, a one-month DTA simulation during November 2014 is carried out with both the metamodel and the complete model. The performance of the metamodel is evaluated by comparing its computed traffic flows with those of the complete model at street resolution for all time intervals of 15 minutes during one month. The performance of the metamodel is also evaluated by comparing its computed traffic flows with the loop detector measurements during a month, at detector resolution at all time intervals.

In **Chapter 5**, the metamodel built in Chapter 4 is used. A global sensitivity analysis (GSA) with Sobol' method is carried out in order to study the sensitivity of the computed traffic flows and speeds with respect to the uncertain inputs of LADTA model. Sobol' sensitivity indices are computed at street resolution throughout the whole

agglomeration of Clermont-Ferrand.

Chapter 6 builds the modeling chain for estimating on-road by coupling the DTA model LADTA with the emission model COPERT IV, in the whole agglomeration of Clermont-Ferrand at street resolution. Then a qualitative sensitivity analysis is carried out on the modeling chain, in order to study the sensitivity of the computed on-road traffic emissions with respect to the LADTA inputs as well as the inputs of COPERT IV model: (i) total traffic demand, (ii) speed limits of the road network, (iii) vehicle fleet composition. Then COPERT IV model is also coupled with the metamodel built in Chapter 4. Then the GSA with Sobol' method is carried out to study the sensitivity of on-road traffic emissions computed by the modeling chain in the whole agglomeration at street resolution, with respect to both (i) the uncertain inputs of LADTA model, and (ii) the vehicle fleet composition inputs for COPERT IV.

Chapter 7 generates two ensembles: (i) an ensemble of DTA simulations and (ii) an ensemble of emission simulations, for the agglomeration of Clermont-Ferrand at street resolution at all time intervals of 15 minutes during November 2014. Then the uncertainty in the computed traffic flows/speeds/emissions with respect to the uncertainty lying in the inputs of DTA model is quantified based on the generated ensemble. In addition, by comparing the traffic flow results from the DTA simulation ensemble with the traffic flows measurements obtained by loop detectors, the reliability of the ensemble of DTA simulation is evaluated in order to know whether the ensemble well represents the uncertainty lying in the computed traffic flows. Preliminary results for the ensemble calibration are also presented in this chapter.

Finally, conclusions of the PhD work are presented at the end of this dissertation. Perspectives and future research lines that have raised from the results of this thesis are also presented.

Chapter 1

Introduction

Summary

This chapter introduces to traffic and emission models. It reviews the associated literature and the various methods used in this PhD work. Firstly, a review of macroscopic static and dynamic user-equilibrium based traffic assignment models is given. Secondly, a review of the metamodeling approach is given, especially for its application in the field of traffic assignment modeling and in on-road traffic emission modeling. Thirdly, a review of state-of-the-art emission models is given. At last, a brief review regarding main methods of sensitivity analysis and uncertainty quantification are given, in view of their application to traffic modeling and on-road emission estimations.

Résumé

Ce chapitre présente l'état de l'art des modèles de trafic et des méthodes utilisées dans cette thèse. Tout d'abord, différents modèles d'affectation du trafic sont présentés. On souligne notamment (i) les différences entre modèle statique et modèle dynamique, et (ii) les différences entre les modèles dynamiques à l'échelle mésoscopique et à l'échelle macroscopique pour une agglomération entière. Deuxièmement, une présentation de l'état de l'art sur la méta-modélisation est proposé, en particulier pour l'application dans le domaine de la modélisation dynamique du trafic et de ses émissions. Ensuite, différents modèles d'émission sont brièvement présentés. Le modèle COPERT est choisi pour estimer les émissions dues au trafic routier à l'échelle métropolitaine. Enfin, les différentes méthodes d'analyse de sensibilité et de quantification d'incertitude sont présentées. On présente aussi l'originalité du travail de cette thèse, par rapport à l'état de l'art.

Contents

1.1 Traffic assignment

The traffic assignment (TA) aims at determining the network traffic flows according to users route choices when they travel from origins to destinations. It can be considered as an economical equilibrium between the *traffic demand* and the *supply*. The traffic demand is often represented by an Origin-Destination matrix (O-D matrix), filled with traffic flow between each Origin-Destination pair (O-D pair). In metropolitan area, the O-D matrix summarizes the total number of users traveling between the Origin-Destination zones (O-D zones), predefined for the simulation area. The supply is a modeled network for the simulation area, represented by an oriented graph with nodes and links. In general, nodes are used to model intersections of roads on the real-world network. Links are used to model roads/streets/freeways of the real-world network. We talk about *demand-supply equilibrium* since there are various constraints on the network for users to choose their travel routes. Capacity and speed limit are two main constraints of roads for a network. The *capacity* is defined as the maximum traffic flow (in $\mathrm{veh\,h^{-1}}$) that can pass through a road. There are two categories of traffic assignments: static one and dynamic one. There are two main modeling problems in traffic assignment problem: (i) modeling the route choices of the network users throughout the network and (ii) the flowing model: a function that maps the traffic flow to travel time for a path or for a road, especially when there are congestions on the network. In the static TA problem, the demand and parameters for the network supply are constant during a given simulation period, usually during peak hours. There is no temporal dimension in static TA problem. In dynamic traffic assignment (DTA) problem, traffic demand, road capacities and network parameters can depend on the *clock time* during the simulation period.

1.1.1 Route choice model

One of the most widely used assumptions for modeling users' route choice is that they choose their routes between O-D zones by minimizing their perceived travel cost, such as travel times, monetary costs (toll, fuel, etc.). The *generalized cost* of users route choice is then modeled by converting all kinds of costs into one general criterion (often the time). Then, it is assumed by the *user equilibrium (UE) condition* [Wardrop, 1952] that at equilibrium when every user succeeds in finding a route with a minimum generalized cost, all *used* routes associated with the same O-D pair should have the same minimum cost, so that there is no possibility for users to shift to another route. Different individuals may have different perceptions of their *least cost* route. Different

user classes (e.g., according to vehicle types such as passenger cars, utility vehicles, trucks, etc.) can be considered when modeling route choice. *Stochastic effects* can also be considered in modeling route choice since travelers may not have the full information on the network so that they choose their *best routes* according to a perceived probability distribution. Detailed reviews concerning different kinds of route choice models can be found in de Dios Ortúzar and Willumsen [2011] (page 356-357).

In static TA problem with fixed O-D matrix and network parameters, the equilibrium status does not depend on the *clock time*. The static assignment problem is formulated as a classical optimization problem with unique and stable equilibrium [Beckmann et al., 1956; Smith, 1979; Sheffi, 1985]. This optimization problem can be formulated as

$$find\, \boldsymbol{x}^* \in \boldsymbol{X},\ s.t.\ \langle \boldsymbol{C}(\boldsymbol{x}^*), (\boldsymbol{x} - \boldsymbol{x}^*) \rangle \geq 0, \forall \boldsymbol{x} \in \boldsymbol{X}, \tag{1.1}$$

where $\boldsymbol{x}$ is the path flow between origins and destinations. $\boldsymbol{X}$ is the set of traffic flows satisfying the traffic demand constraints of the O-D matrix. $\boldsymbol{C}$ is the generalized cost of all paths. At metropolitan scale with large-size network, static TA model is well studied. Most of the transportation planning software are based on static TA formulations such as VISUM, TransCAD, etc.

In dynamic traffic assignment (DTA) models however, the traffic demand, travel costs, link parameters (capacity and speed limits) and route choice strategies are time-dependent [Chiu et al., 2011; Peeta and Ziliaskopoulos, 2001]. For the DTA problem, the Wardrop equilibrium principle of route choice can be extended to the *dynamic user equilibrium (DUE)* principle as follows:

Under equilibrium conditions in networks where congestion varies over time, traffic arranges itself so that at each instant the costs incurred by drivers on those routes that are used are equal and no greater than those on any unused route [de Dios Ortúzar and Willumsen, 2011].

The DUE represents the travel time needed to reach any point in the network and its variations as a function of *clock time* over the simulation period. It gives an accurate description of spatio-temporal propagation of traffic through the network [Meunier and Wagner, 2010].

With the additional time dimension, DTA models are more complex than static ones in both analytical and algorithmic aspects. On the one hand, the analytical models are more complex regarding the dynamic user equilibrium (DUE) condition and analytical formulations of the DTA problem. Researches on the existence, uniqueness and stability of the DUE solution are challenging tasks among DTA studies [Zhu and Marcotte, 2000; Mounce, 2006, 2007; Meunier and Wagner, 2010; Wagner, 2012]. The DUE-based

traffic assignment problem can be considered as a dynamic optimization problem (DOP), and the formulations of the analytical model are complex. As reviewed in Ran and Boyce [1996]; Chen and Hsueh [1998]; Peeta and Ziliaskopoulos [2001], the variational inequality (VI) approach, firstly proposed by Friesz et al. [1989, 1992, 1993], is a general approach that provides analytical flexibility and convenience in addressing various DTA problems. On the other hand, the computation of dynamic minimum cost path is an NP-hard problem [Orda and Rom, 1990]. The application of DTA to real-word network with limited (computational) budget is still challenging. In fact, it requires to compute for each node in the network, the path who has the minimum (generalized) travel costs from all other nodes as a function of *clock time* during the whole simulation period. DTA models are far more computationally demanding than static ones, and may exceed in many cases the capability of current hardware.

1.1.2 Flowing model

The *offer* side of a traffic assignment problem is often constrained by the infrastructure of the simulation area. The flowing model aims at modeling the travel speed or travel time on a road in function of traffic flow, with constraints of road capacity. In a static TA model, the effect of road capacity on travel time is modeled by a volume-delay function (VDF): the travel cost (or time) is a strictly increasing function of the traffic volume on a road. For example, Bureau [1964] proposed the most commonly used function

$$t = t_0[1 + \alpha_{VDF}(\frac{V}{C})^{\beta}], \tag{1.2}$$

where t is the link travel time to be determined, t_0 is the link free flow travel time, V is the volume on the link, C is the capacity, α_{VDF} and β are coefficients to be adjusted. Figure 1.1 illustrates this type of VDFs with $\alpha_{VDF} = 1$ and a varying β.
It is observed that VDFs are very sensitive to β. To represent the fact that the link travel time increases with the volume loaded on a link in congested cases, the volume may increase indefinitely and exceed the link capacity. This is not possible in reality, and it is also unrealistic to assume that the travel cost depends only on the flow of the link. In other words, in congested condition, the travel time calculated by VDF in static models does not depend on physical features of congestion (such as travel speed, density, or queue) [de Dios Ortúzar and Willumsen, 2011].

Modeling dynamics of traffic flow can better represent the congestion phenomenon. Time-dependent dynamic flowing models are used in DTA problems for describing time-dependent relation between traffic loaded on a path (or on a link) and travel cost

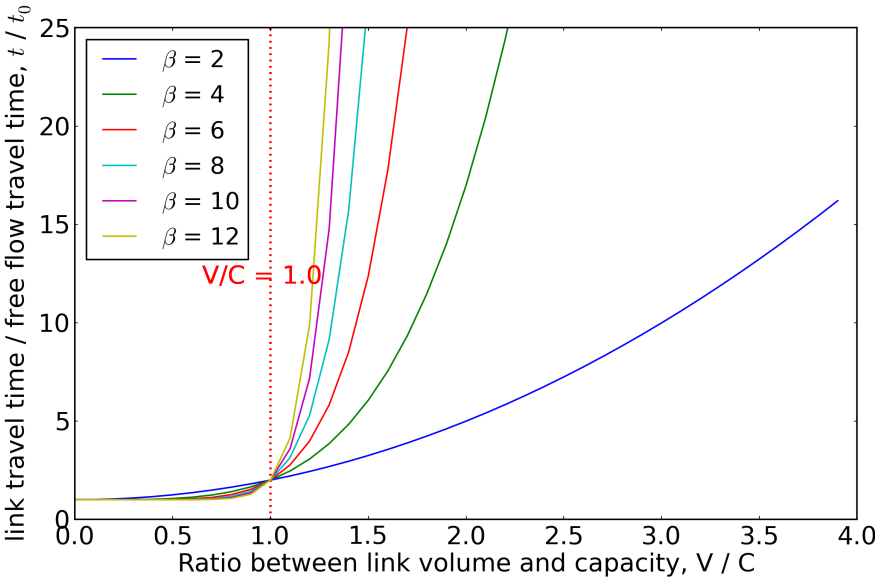


Figure 1.1: An example of volume-delay functions in static traffic assignment models. When volume loaded on a link exceeds the capacity, the link travel time (t) can be calculated with $\alpha_{VDF} = 1$ and different β.

(time). Some DTA models use a microscopic traffic simulator [Mahut and Florian, 2010] as flowing model which provides detailed behavior of each traveler on the network. Other traffic flowing models use a simpler macroscopic approach. For instance, LWR model of Whitham [1955] and Richards [1956] based on hydrodynamic principles of fluid mechanics (see in Appendix A). It is widely used in macroscopic traffic modeling but it involves calibrating the *fundamental diagram*, which gives a relation between the traffic flow (in veh h^{-1}) and the traffic density (in veh km^{-1}), according to characteristics of each street. This might be computationally costly for a metropolitan-scale network with more than 10,000 links. Another simpler way to represent the relation between travel time and traffic flow on links is the use of a first-in-first-out (FIFO) point queue model [Kuwahara and Akamatsu, 1993; Mounce, 2006]. A point queue is associated with each link and it replaces the volume/delay functions used in static TA problem. A point queue model allows to map the time-varying traffic demand on a link to its travel time. The time-varying route travel times are then computed by composing the individual link travel time functions along each route. The point queue indeed ignores the spillback of queues. However, it can model the congestion condition where the travel time is bigger than the free-flow travel time when the entering traffic flow of a link exceeds link capacity. It can better model the travel time in congested condition, when compared with volume/delay functions in static models.

1.1.3 Choosing between static and dynamic models

The choice of the modeling approach depends on the required application: the accuracy required, the decision-making context, the level of details required, the availability of data and resources, etc. [de Dios Ortúzar and Willumsen, 2011] (page 10 - 12). For emission estimation at urban scale, it is important to choose appropriate models in order to build the modeling chain from traffic modeling to emission modeling. Microscopic simulators coupled with emission model have been already used for estimating on-road traffic emissions at district scale on a limited-dimension network [da Rocha et al., 2015; Shorshani et al., 2015a,c; Thouron et al., 2018]. At urban scale and metropolitan scale however, dynamic microscopic simulators are computational costly if we want to compute aggregated traffic flows and speeds at link resolution for a high-dimensional network. In this case, macroscopic traffic assignment models are widely used. Compared to static macroscopic traffic assignment models [Fellendorf et al., 2000], DUE-based assignment models compute time-dependent traffic flow and average speed at link resolution. This allows us to estimate on-road traffic emissions with detailed time resolution. Moreover, unlike in static models where averaged traffic flows are computed for each chosen route during an hour, the DUE-based assignment model takes into account the continuous spatial-temporal evaluation of traffic flows throughout the network. The spatial-temporal feature is also important for estimating on-road traffic emissions.

The traffic model we use in this PhD work is a DUE-based dynamic traffic assignment model called LADTA. The traffic demand in O-D matrix is a continuous function of time, expressed by cumulated traffic demand for each O-D pair during a given simulation time interval. The outputs of LADTA are cumulated traffic volumes at link resolution, as a function of the clock time at which users enter into the link, during the same simulation period. Link capacities and speed limits can be time-depending, too. Detailed descriptions of LADTA model, and its implementation LTK are presented in Chapter 2. It has already been applied to a real-world network [Leurent, 2003; Aguiléra and Leurent, 2009]. It has also been coupled with an emission model in order to estimate the atmospheric pollutant emission due to on-road traffic in the metropolitan area of Paris region [Aguiléra and Tordeux, 2011].

1.2 Model evaluation and performance criteria

This section presents the statistical criteria to compare two sequences of traffic flows. On the one hand, the criteria are used to compare the results computed by two models: (i) static model v.s. dynamic model in traffic assignment, (ii) metamodel v.s. original

model to evaluate the performance of the metamodel. On the other hand, in order to evaluate the performance of models, the computed traffic flows on the network are also compared with loop detector measurements.

Traffic measurement data

At urban area, inductive loop detectors are installed beneath the road surface in order to detect whether a metallic object is above them. Figure 1.2 illustrates the basic structure and function of a loop detector [Treiber et al., 2012].

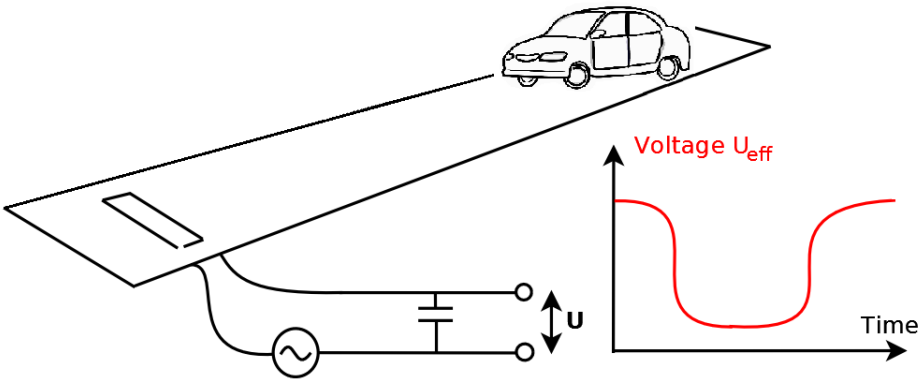


Figure 1.2: The induction loop is part of an inductor-capacitor circuit (complemented by an external capacitor and an alternating current voltage source) tuned to be in resonance if the loop is "unoccupied", yielding a high voltage U_{eff}. The metallic parts of a vehicle will increase the inductance of the loop upon driving over it. This puts the circuit out of tune and decreases the voltage U_{eff}. Source: [Treiber et al., 2012]

The detectors transmit aggregated data to the traffic control center. The traffic flow is defined as the number of vehicles ΔN_i passing a location i within a time interval Δh:

$$Q_{i,h} = \frac{\Delta N_i}{\Delta h}. \tag{1.3}$$

$Q_{i,h}$ is expressed in veh h^{-1}. It can be directly measured by loop detectors. The time intervals of detectors can be down to every 20 s and $\Delta h = \frac{20}{3600}$ h. The output of LADTA model is the cumulated flow at link resolution. It is non-decreasing and measurable on $\mathbb{R}$. The cumulated traffic during $\mathcal{H}_{\Delta h} = [h, h + \Delta h]$ for the link a is $X_a([h, h + \Delta h])$. Assuming that the traffic flow on link a can be measured at location i, then $\frac{X_a([h,h+\Delta h])}{\Delta h}$ can be compared with $Q_{i,h}$ measured by loop detectors.

Performance criteria

Let $(r_i)_i$ be the sequence of reference and $(s_i)_i$ be the sequence to be evaluate. Statistical criteria in Table 1.1 are used to compare these two sequences.

Table 1.1: Statistical criteria to evaluate a model by comparing the computed values $(s_i)_i$ with the reference sequence $(r_i)_i$. They have the same dimension denoted as N.

Criteria	Notations	Formula
Mean value	Mean	$\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r_i$, $\bar{s} = \frac{1}{N}\sum_{i=1}^{N} s_i$
Root mean square error	RMSE	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - r_i)^2}$
Bias	Bias	$\frac{1}{N}\sum_{i=1}^{N}(s_i - r_i)$
Correlation	Corr	$\frac{\sum_{i=1}^{N}(s_i-\bar{s})(r_i-\bar{r})}{\sqrt{\sum_{i=1}^{N}(s_i-\bar{s})^2}\sqrt{\sum_{i=1}^{N}(r_i-\bar{r})^2}}$
Normalized root mean square error	NRMSE	$\frac{RMSE}{\bar{r}}$
Mean normalized bias error	MNBE	$\frac{1}{N}\sum_{i=1}^{N}\frac{s_i-r_i}{r_i}$

With the additional time dimension, the output of DTA simulation has therefore two dimensions (i, h): (i) the spatial dimension represented by the index of location i and (ii) the temporal dimension represented by the time h. Let $r_{i,h}$ denote the reference value at (i, h). It can be the output of a reference model or be an observation at (i, h). Let $s_{i,h}$ be the computed value at (i, h). Let N_{link} be the number of links in the simulation network and T be the dimension of time for the simulation period. Then we use *total* criteria when we compare the two sequences at all locations and all times. N in Table 1.1 equals to $N_{link} \times T$. We can also evaluate the *temporal* performance and *spatial* performance of a model. The *temporal* performance is given by comparing *spatially-averaged* values of $s_{i,h}$ with those of $r_{i,h}$ at all times h. The *spatial* performance is given by comparing *temporally-averaged* values of $s_{i,h}$ with those of $r_{i,h}$ at all locations i. Table 1.2 presents the formulas for evaluating *total, temporal* and *spatial* performance of a DTA model.

Table 1.2: Total, temporal and spatial statistical criteria to evaluate a model. $s_{i,h}$ and $r_{i,h}$ respectively denote the computed and reference value at location i at time h. N_{link} is the spatial dimension and T is the temporal dimension. We reuse some notations and formula presented in Table 1.1.

Criteria	Formula
Spatially-averaged sequence (dimension T)	$\bar{r}_h = \frac{1}{N_{link}} \sum_{i=1}^{N_{link}} r_{i,h}$, $\bar{s}_h = \frac{1}{N_{link}} \sum_{i=1}^{N_{link}} s_{i,h}$
Temporally-averaged sequence (dimension N_{link})	$\bar{r}_i = \frac{1}{T} \sum_{i=1}^{T} r_{i,h}$, $\bar{s}_i = \frac{1}{T} \sum_{i=1}^{T} s_{i,h}$
Temporal bias	Bias $(\bar{s}_h, \bar{r}_h)$
Spatial bias	Bias $(\bar{s}_i, \bar{r}_i)$
Temporal RMSE	RMSE $(\bar{s}_h, \bar{r}_h)$
Spatial RMSE	RMSE $(\bar{s}_i, \bar{r}_i)$
Temporal NRMSE	NRMSE $(\bar{s}_h, \bar{r}_h)$
Spatial NRMSE	NRMSE $(\bar{s}_i, \bar{r}_i)$
Temporal correlation	Corr $(\bar{s}_h, \bar{r}_h)$
Spatial correlation	Corr $(\bar{s}_i, \bar{r}_i)$

1.3 Metamodeling

DTA models can compute traffic flows and speeds at link resolution for a whole metropolitan area. The additional time dimension can model time-varying traffic features. However, as reviewed in Section 1.1, dynamic models are more complex than static models in both analytical and numerical aspects. To speed up the DTA simulations on large-size networks is one of the main challenges for researchers and engineers in traffic modeling. One of the solutions to deal with computer intensive models is using metamodels (i.e., surrogate models) in order to approximate complex models.

A metamodel is an approximation of the input/output function that is implied by the underlying simulation model [Kleijnen, 2009]. The built metamodel preserves the main response of the original model to input, but with much lower computational time. Once a class of surrogate models has been chosen according to some prior knowledge, they can be built upon available samples [Roustant et al., 2012]. There are three main steps for constructing a metamodel based on an original model: (i) sampling selection or experimental design in order to build a set of training points, and the corresponding training values; (ii) construction of the metamodel in order to find an surrogate model based on the training points and training values; (iii) evaluation and optimization of the metamodel. There are various types of metamodeling methods for computer experiments include linear regression (polynomial regression) [Madu and Kuei, 1994; Kleijnen,

2009], splines, neural networks, Gaussian emulators [Rasmussen and Williams, 2006; Sacks et al., 1989], Kriging surrogates [Kleijnen, 2009], and radial basis function (RBF) [Dyn et al., 1986; Broomhead and Lowe, 1988], etc. For low-dimensional problems, linear regression may be straightforward and efficient. However, it has less robustness when applied to high-dimensional models, compared with Kriging and RBF methods [Jin et al., 2001].

Kriging method is a spatial interpolation method [Roustant et al., 2012]. The Kriging method was firstly proposed by D.G. Krige [Krige, 1951] and mathematically formulated in Matheron [1963] for geostatistics. It is applied to interpolate spatial data for constructing surrogate models [Rasmussen, 2004; Kleijnen, 2009; Stein, 2012]. Numerical Kriging approximation packages are developed such as DACE Matlab toolbox [Lophaven et al., 2002] and DiceKriging [Roustant et al., 2012]. Let us denote a model $y = f(\boldsymbol{p})$ with $\boldsymbol{p} \in \mathbb{R}^K$ and f is computationally costly. Basically, the Kriging metamodel aims at predicting the model's outputs on a target (unknown) input point $\boldsymbol{p}^{(0)} \in \mathbb{R}^K$, based on a linear combination of known training values $(\{f(\boldsymbol{p}^{(1)}), \ldots, f(\boldsymbol{p}^{(i)}), \ldots, f(\boldsymbol{p}^{(n)})\}$), computed at training points (denoted as $\{\boldsymbol{p}^{(1)}, \ldots, \boldsymbol{p}^{(i)}, \ldots, \boldsymbol{p}^{(n)}\}$ with $n \in \mathbb{N}$). The objective of Kriging interpolation is to find $\hat{f}$ so that for any unknown $\boldsymbol{p}^{(0)} \in \mathbb{R}^K$, $\hat{f}(\boldsymbol{p}^{(0)}) = \sum_1^n \omega_i(\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(1)}, \ldots, \boldsymbol{p}^{(n)}) f(\boldsymbol{p}^{(i)}) \simeq f(\boldsymbol{p}^{(0)})$. ω_i is the weight affected to $f(\boldsymbol{p}^{(i)})$ and it depends only on the relative position of $\boldsymbol{p}^{(0)}$ with respect to training points. They are chosen to minimize the variance of the prediction error. The training process of the Kriging interpolation consists in determining the covariance of the Gaussian process [Roustant et al., 2012; Mallet et al., 2013]. The interpolation process is unbiased so that the predictions at all training points coincide with the corresponding training values. Thorough reviews regarding Kriging method and Gaussian process can be found in Rasmussen and Williams [2006]; Kleijnen [2009]; Stein [2012]; Roustant et al. [2012].

The Kriging interpolation is unbiased. However, it is very time-consuming especially for interpolating large-dimension training points because it involves a K-dimension optimization process [Jin et al., 2001; Mallet et al., 2013]. Another similar interpolation method is with radial basis functions (RBFs) [Dyn et al., 1986; Broomhead and Lowe, 1988]. The RBF-based interpolation of the training points is more computationally efficient than the Kriging method. RBFs were developed initially in Duchon [1977]; Oeuvray and Bierlaire [2009] and their applications are in various fields in sciences and mathematics [Buhmann, 2003]. The concept of RBF-based interpolation is that the influence of a training point $\boldsymbol{p}^{(i)}$ to the approximation of $f(\boldsymbol{p})$ depends only on the distance between $\boldsymbol{p}^{(i)}$ and the target point $\boldsymbol{p}^{(0)}$: the bigger the distance, the less $\boldsymbol{p}^{(i)}$ should be influential. Interpolation based on RBFs can be briefly summarized as finding $\hat{f}$ so

that for any unknown $\boldsymbol{p}^{(0)} \in \mathbb{R}^K$, $\hat{f}(\boldsymbol{p}^{(0)}) = \sum_1^n \lambda_i \phi(d(\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(i)})) \simeq f(\boldsymbol{p}^{(0)})$, where ϕ is the chosen RBF, λ_i is the weight and $d(\cdot, \cdot)$ defines the distance between $\boldsymbol{p}^{(0)}$ and $\boldsymbol{p}^{(i)}$. λ_i is computed so that $\hat{f}(\boldsymbol{p}^{(i)}) = f(\boldsymbol{p}^{(i)})$ for $i = 1, \ldots, n$. The computation of the weights is in fact to solve a linear system. The *distance* here can be defined by users in order to take into account the sensitivity of the model f with respect to each input $\boldsymbol{p}^{(i)}$. Without specification, it is often defined as Euclidean norm. The choice of different radial basis functions depends on applications. A method employing multiquadric radial functions is proposed in Hardy [1990] and it can be used for a large variety of problems. Most of these applications are one, two or three dimensional [Oeuvray and Bierlaire, 2009].

The three main methods mentioned above are widely used in various domains, and the choice among different methods depends on the application. In recent years, metamodeling methods have also been used in traffic modeling and management, in order to solve the computational burden in algorithm, sensitivity analysis, safety management, etc. For example, surrogate of the microscopic Aimsun model is built with Kriging method for global sensitivity analysis (GSA) and model calibration [Ciuffo et al., 2013; Azevedo et al., 2015]. Some other studies of metamodeling applied to microscopic traffic simulators are reviewed in Song et al. [2017] for different applications. Macroscopic traffic assignment problems at urban scale are also addressed by embedding surrogates of microscopic traffic simulators in an optimization algorithm. For example, Osorio and Bierlaire [2013] propose an efficient simulation-based optimization framework embedding a polynomial-based metamodel that integrates information from a microscopic simulator with an analytical queuing network model. Based on this framework, different algorithms are proposed to carry out (i) static traffic assignment [Osorio and Chong, 2015; Osorio and Nanduri, 2015] and (ii) dynamic traffic assignment and calibration [Chong and Osorio, 2017; Zhang et al., 2017], at urban scale under tight computational budget. Metamodels based on microscopic simulators are also used in other optimization problems in transportation management such as highway charges and dynamic pricing [Chen et al., 2014; He et al., 2016]. Besides the initiatives in accelerating the optimization algorithm, there are also applications that make use of metamodeling methods in order to build surrogate for dynamic network loading (DNL) models, mapping the set of path departure rates to the set of path travel time. For example, Song et al. [2017] propose a framework of building metamodel of dynamic flowing models based on Kriging method. Their case studies on a SiouxFall network (76 links and 24 nodes) show the potential of surrogates of DNL models for speeding up DUE-based model.

However, the mentioned surrogate-based DTA models might be adequate for transportation planning, management and optimization problems in an operational context.

Applying these models to large-scale DUE-based model for a whole city is still computationally costly, in cases where large numbers of model evaluations are required at metropolitan scale. These cases include carrying out global sensitivity analysis of model inputs, long-term DTA simulations to be combined with traffic emission models for air quality simulations, and probabilistic simulations based on Monte Carlo simulation of the deterministic DTA model. In these cases, metamodels that replace DTA models applied to large-scale networks are needed. Based on the objectives of this PhD work, the on-road traffic emission estimation, GSA analysis and uncertainty quantification should be carried out at metropolitan scale. It is necessary to reduce the computational cost of the DTA model applied to a large-scale network of a whole city. The most important outputs computed from the DTA model for the emission estimation is time-depending street-level traffic flow and travel speed. This motivates us to build a statistical metamodel that directly emulates the relation between the DTA model inputs and the computed street-level traffic flow/speed. The final metamodel is efficient because it ignores the computationally expensive dynamic optimization problem lying behind the DTA model. This metamodeling approach has not yet been applied in traffic modeling, but the idea of applying the metamodeling method directly to a given large scale model has proved to be successful in various fields such as tsunami simulation [Sarri et al., 2012], as well as in air quality simulations with a global aerosol model [Lee et al., 2011], an urban model [Mallet et al., 2013] and an atmospheric dispersion model for radionuclides [Girard et al., 2016]. Kriging, RBF interpolations or alternatives are limited by the dimension of the input space [Saltelli et al., 2008; Mallet et al., 2013]. For large dimension models (which is the case for LADTA model applied to a large-scale network), a dimension reduction should be firstly carried out. The outputs are projected onto a reduced subspace, which can be obtained based on principal component analysis of the training values. Then the final metamodel combines an emulator of the reduced model and a reconstruction of the model outputs. Detailed methods and applications are presented in Chapter 4.

1.4 Modeling of atmospheric pollutant emissions

The on-road traffic emissions are often modeled as the product of (i) the traffic volume and (ii) the emission factors (EF, in $\mathrm{g\,km^{-1}}$), for different pollutants, vehicle types and technologies [Boulter and McCrae, 2007; EEA, 2016]. In metropolitan areas, the former can be estimated by traffic assignment models, and the latter are often modeled by empirical functional relations between pollutant emissions and characteristics

of on-road traffic. Various models are built for emission factors (EFs), and five main types of EF models are categorized by Smit et al. [2010] according to the required input variables: (i) 'average-speed' models (e.g., COPERT), where EFs are a function of the mean traveling speed; (ii) 'traffic-situation' models (e.g., HBEFA) where EFs are determined by descriptions of a particular traffic situation (e.g., 'stop-and-go-driving', 'free flow driving'); (iii) 'cycle-variable' models (e.g., MEASURE, VERSIT+), and (iv) 'modal' models (e.g., PHEM) where EFs (in $g\,s^{-1}$) are produced via engine of vehicle operating models at the highest resolution (down to seconds), etc. Different types should be applied for different contexts to adapt the level of detail required by researchers, city planners or policy makers. Detailed review concerning different emission models can be found in Shorshani [2014]; Shorshani et al. [2015a]. Table 1.3 adapted from [Shorshani, 2014; Shorshani et al., 2015a; Smit et al., 2010] summarizes the most commonly used traffic models.

Table 1.3: Summary of emission factor models

Categories	Application domain	Required traffic inputs	Example
Averaged-speed based	Urban area	Average speed	COPERT [EEA, 2016] [Ntziachristos et al., 2009]
Traffic-situation based	Neighborhood	Average speed corresponding to different traffic condition	HBEFA [INFRAS, 2017]
Model based on chronological speed	Neighborhood intersection area	Driving cycles Vehicle characteristics	PHEM [Zallinger et al., 2008]
Driving cycle variable based	Neighborhood	Driving cycles	MEASURE [Fomunung et al., 2001] VERSIT+ [Smit et al., 2007]

"Model-based" and "circle-based" emission models require detailed information of vehicle speed. They can be coupled with microscopic dynamic traffic simulators and compute on-road traffic emissions at neighborhood area [Boulter et al., 2007; da Rocha et al., 2015; Shorshani et al., 2015b; Thouron et al., 2018]. At urban scale for a whole city, the microscopic simulators are too computationally costly at street resolution for a large-scale network. Macroscopic traffic assignment models replace microscopic simulators. They compute average speed at street resolution instead of detailed vehicle moving dynamics. Therefore, average-speed based emission model is a straight-forward choice for building the emission modeling chain coupling at large urban scale. The methodology of COPERT IV model is one of the most used 'average-speed' based emission factor models.

The COPERT IV formulations are presented as the *Tier 3* method in the air pollutant emission inventory guide book published by the European Environment Agency (EEA) [EEA, 2016]. The core principal of COPERT IV method is that the emission factors are modeled as empirical functions of vehicle average speed. The parameterization in the functions depends on different pollutant type (main pollutants such as NOx, CO, VOC, and particulate matter PM), vehicle type (passenger car, light commercial car, heavy duty vehicles, motorcycles), motorization type (diesel, petrol, hybrid, electric), vehicle technology (Euro standard), etc. This method was initially proposed for estimating emissions at national or regional level. Recently, the COPERT emission factor model has been applied to street resolution at urban scale by coupling with static traffic assignment model [Soulhac et al., 2012; Shorshani et al., 2015a; Pu et al., 2015; Tang et al., 2017], in order to estimate street resolution emissions at urban scale. As already mentioned in Section 1.1, the static traffic assignment model computes only average traffic flows during a fix time period, ignoring the temporal-spatial variation of traffic flows over the whole urban area. With the modeling chain which couples static traffic assignment model and COPERT model, the computed emissions are stable during the simulation period. The time-varying dynamics of the traffic flow and emission are not captured more precisely than at this level [Shorshani et al., 2015b]. That is one of the main reasons for which we chose the DUE-based dynamic traffic assignment (DTA) model in order to compute the emission at a higher time resolution, yet preserving the efficiency of the traffic model for being able to simulation street-level traffic flows and speeds under tight computational budget in operational context. In addition, with the metamodeling method presented in Section 1.3, the computation of street resolution traffic flows and speeds can be much more efficient. This allows us to estimate emissions during long-term period (e.g., a whole year) with a more accurate time resolution than the static TA model, at street resolution at metropolitan scale. It also allows us to carry out a large number of evaluations of the modeling chain, in order to carry out global sensitivity analysis and uncertainty quantification for emission estimation, at urban scale.

1.5 Sensitivity analysis and uncertainty analysis in modeling traffic and its emissions

The uncertainty analysis aims at determining the distribution of the model output. To analyze the uncertainty of model output due to its inputs, the sensitivity analysis of the model can help us to identify to which input(s) the model output is *sensitive*.

The uncertainty analysis and sensitivity analysis can be carried out in tandem [Saltelli et al., 2008]. The *model* mentioned here corresponds to a mathematical or computational interpretation of a physical real-world system. Let us denote f as the model and X as the model output. U_i is the input and the model to be analyzed is $\boldsymbol{X} = f(U_1, \ldots, U_i, \ldots, U_k)$ where $\boldsymbol{X} \in \mathbb{R}^D$ is the model output, U_i is the i^{th} input and k is the number of inputs of the model.

1.5.1 Sensitivity analysis

The objective of sensitivity analysis (SA) is to *study how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input* [Saltelli et al., 2004]. From SA, we want to know what input of the model can cause large variation in the model output, or to know whether there are some factors that are not influential on the output. The SA is a useful tool for various applications such as testing the robustness of the model outputs regarding to inputs, calibrating model parameterization, identifying influential inputs for decision making, etc. There are both local sensitivity analysis (LSA) and global sensitivity analysis (GSA). Detailed reviews of different LSA methods can be found in Helton [1993]; Saltelli et al. [2004]; Borgonovo and Plischke [2016]. In LSA, the variability of the model output is explored around one point of interest in the model input space [Borgonovo and Plischke, 2016]. There are many widely used LSA methods such as the *one at a time method* (OAT) method, deviation-based method, differentiation-based methods. OAT is straightforward. One can vary one model input at a time and analyze the variation of the model output. Deviation-based method is also a widely used in LSA. The derivative $\frac{\partial X_j}{\partial U_i}$ of an output X_j with respect to the input U_i is in fact considered as the mathematical definition of the sensitivity of U_j versus Ui [Saltelli et al., 2008]. The LSA was carried out on mathematical models of User-Equilibrium (UE) traffic assignment in order to explore the variation of the optimization solution with respect to perturbation of traffic demand and parameters in route cost functions (e.g., in Tobin and Friesz [1988]; Yang [1997]; Cho et al. [2000]; Patriksson [2004]; Lu [2008]; Lu and Nie [2010]).

However, LSA is carried out around a point in the input space with small perturbations. It does not explore the entire input space and ignores the effect of input interactions when perturbations are given to more than one input simultaneously. Since the main objective of the PhD work is to quantify the uncertainty in traffic emission estimation due to model inputs, we are interested the relative importance of the inputs over their whole variation domains, with consideration of input interactions. In addition, increasing *computational* models are used in order to interpret physical system that

firstly expressed mathematically and then implemented in the form of computer program [Morris, 1991]. These models are computationally expensive and the closed form is not available. The differentiation-based methods cannot be applied. For example, the macroscopic traffic assignment model at urban scale can be considered as a model with (i) O-D matrix and network parameters as inputs, and (ii) street resolution traffic flows/speeds as outputs. The underlying input-output relation is not straightforward. At urban scale, the dimension of the outputs is usually high. The model evaluation relies on numerical simulation (e.g., for searching the minimum cost routes and reaching *User Equilibrium* condition) and is often computationally expensive. Emission models and air quality models have the similar features, too. Therefore, we are more interested in choosing a SA method that can cope with these kinds of models in order to study the relative influence of each input on outputs. Elementary effect method (e.g., Morris model [Morris, 1991]), variance-based method (e.g., FAST-Fourier Amplitude Sensitivity Test [Cukier et al., 1978] and Sobol' method [Sobol, 1993]). Iooss [2011] reviews various global sensitivity analysis methods of model output, in a relatively complete methodological framework. A comparison of different methods for local and global sensitivity analysis is synthesized in Figure 1.3, depending on (i) the model behavior and (ii) the (approximate) required number of model evaluations [Iooss, 2011].

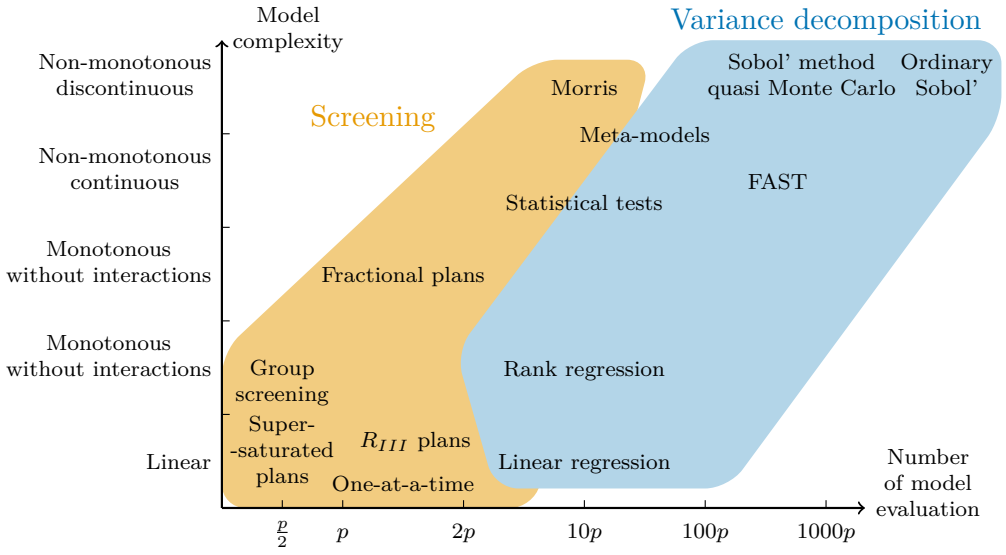


Figure 1.3: Methods for sensitivity analysis. p is the dimension of the model input. Source: [Iooss, 2011]. Figure regenerated by Sylvain Girard.

Among various GSA methods, both Morris method and VBSA have been used in traffic assignment models and emission models. Morris model is a qualitative GSA

method. It involves individually randomized OAT experiments. It is relatively efficient and can give qualitative information to rank the model inputs according to their influence on the output Campolongo and Saltelli [1997]. It has been successfully used in analyzing global sensitivity of COPERT III model for estimating transport emission in Italy [Kioutsioukis et al., 2004]. Thorough reviews regarding Morris method and its extension can be found in Campolongo and Saltelli [1997]; Saltelli et al. [2000]; Campolongo et al. [2007]; Saltelli et al. [2008]; Campolongo et al. [2011].

However, Morris method is carried out in a deterministic setting and does not take into account the uncertainty of inputs. For a quantitative GSA with consideration of input uncertainties, variance-based sensitivity analysis (VBSA) has been widely used in recent decades. VBSA methods were firstly employed in 1970s [Cukier et al., 1973]. Instead of being carried out in the deterministic setting as in LSA or Morris method, the VBSA is carried out in the probabilistic setting. The model inputs are considered as random variables with known probability distributions. This allows us to take into account the uncertainty of inputs throughout their full variation domains. The outputs of the model are also random variables, whose probability density functions (*pdf*s) are unknown. To determine the *pdf*s of the outputs is the objective of uncertainty quantification (see later in Section 1.5.2). The variance (or standard deviation) can be a measure of a model's uncertainty Saltelli et al. [2008]. Therefore, the reduction of the output variance when *fixing* an input or a group of inputs can be a measure of *importance* of the inputs. VBSA is based on the decomposition of the output variance firstly proposed by Sobol [1993]. Sobol' method makes use of the *first-order* index and *total effect* index in order to quantify (i) the main effect of the input U_i on the model output and (ii) the effect of all possible input interactions with U_i. The combination of Sobol' method with Monte Carlo simulation framework is applied to various domains for carrying out GSA, e.g., in microscopic traffic models [Ciuffo et al., 2013; Punzo and Ciuffo, 2011; Ge et al., 2015], transportation emission estimations [Kioutsioukis et al., 2004], hydrological models [Pianosi et al., 2016; Sarrazin et al., 2016] and atmospheric dispersion model [Girard et al., 2016]. Various computational tools are developed for implementing Sobol' method such as SIMLAB [Tarantola and Becker, 2016] software, or SAFE Matlab toolbox [Pianosi et al., 2015].

Sobol' method gives quantified sensitivity measures with consideration of uncertain inputs throughout the whole input space. It gives precise sensitivity estimates of input interactions, too. One of the main challenges of applying Sobol' method is that it requires a large number of model evaluations in order to reach the convergence of the sensitivity indices. Many researches are addressed to propose relatively efficient

strategies for computing Sobol' indices, for instance in Saltelli [2002]. Metamodeling method is increasingly used for better estimating model sensitivity with Sobol' method in modeling on-road traffic and its corresponding emissions [Ciuffo et al., 2013; Punzo and Ciuffo, 2011; Ge et al., 2015; da Rocha et al., 2015]. These studies are focus on microscopic scale. To the authors' best knowledge, the global sensitivity analysis for traffic assignment model and emission estimations have not been carried out at urban scale and street resolution. The computational burden is one of the main difficulties and that is the main reason why we make use of a metamodeling method at urban scale to address this challenge. Similar GSA study combined with a metamodeling method was successfully carried out in other fields such as in atmospheric dispersion model applied to the Fukushima nuclear accident [Girard et al., 2016].

1.5.2 Uncertainty quantification and ensemble simulation

For a model $X = f(U_1, \ldots, U_i, \ldots, U_k)$, the output X is a random variable of which the *pdf* is to be determined via uncertainty quantification. In practice, the distribution of X can be obtained via *ensemble simulation*. The concept of the latter method is firstly developed in the early 1960s in meteorology in order to evaluate the certainty of weather forecasts [Lorenz, 1963; Epstein, 1969b]. The main idea of Lorenz [1963] is to convert deterministic equations into stochastic dynamic equations by adding perturbations in the initial conditions, in order to approximate stochastic solution. To analyze the uncertainty in model output due to inputs, N model evaluations of f are carried out by sampling N times U. Each U_i is attributed a known *pdf* in its variation domain. The N-member ensemble can then give a discrete probability distribution of X. It is important to evaluate the quality of uncertainty quantification [Anderson, 1996]. Although the uncertainty estimation is in a probabilistic setting, we are able to use the observations to evaluate whether or not the built ensemble can well represent the uncertainty of the model. This involves statistical comparison between the forecast and observed values for the output variable. There are three desirable properties for a good ensemble system: reliability, sharpness and resolution. The reliability indicates the accuracy of prediction of an event. An ensemble with high reliability is so that the probabilistic forecast by the ensemble is (approximately) the same as the observed frequency of this event. The sharpness indicates the variability of the ensemble forecasts. The resolution indicates the ability of the ensemble to make distinct predictions for different events. Various criteria are used to evaluate the performance of an ensemble within observation data: Brier score [Brier, 1950], rank histogram [Anderson, 1996; Hamill, 2001], reliability diagram [Wilks, 2011], and RPS (Ranked Probability Score) [Epstein, 1969a;

Murphy, 1971], etc. A detailed review can be found in Candille [2003].

With increasing population and automobiles in urban area, the main objective of traffic management is how to make best use of the limited infrastructure resources in order to maintain the mobility, accessibility and reliability of our everyday transportation system [Hoogendoorn et al., 2012]. Traffic models play an important role for decision making and policy instruction, and increasing attention has been paying to reliability and uncertainty analysis in traffic models. In fact, adding stochastic aspects in inputs of traffic models is not a new topic, but early studies mainly focus on analytical models [Gendreau et al., 1996]. Based on LSA on the analytical models, main input uncertainties of macroscopic UE-based traffic assignment models lie in (i) traffic demand (O-D matrix) [Yang et al., 1992], (ii) network constraints (capacities and free flow travel times) [Chen et al., 1999]. Efforts have been done to reduce input uncertainty by better estimating the O-D matrix with the help of various type of traffic data [Yang et al., 1992; Ashok and Ben-Akiva, 2000; Zhou and Mahmassani, 2006; Shao et al., 2014; Carrese et al., 2017]. Uncertainty quantification of traffic demand estimation has also been carried out [Hugosson, 2005; Jones et al., 2017; Ma and Qian, 2018]. However, these studies only focus on the uncertainty of the traffic demand. There is no quantitative conclusion of how these uncertainties in traffic demand can propagate through the traffic models and affect the estimated traffic flows, travel times or travel speeds at urban scale. Recently, probabilistic approach with repetitive simulations of a deterministic model is proposed. Calvert et al. [2012] point out the inefficiency of overcomplicated analytical stochastic approaches, and make use of the Monte Carlo simulation of a macroscopic traffic model in order to better taking into account the input uncertainty in the model. Later in Calvert et al. [2018], the same authors applies the probabilistic approach to a DTA simulation on a network of about 8,200 links and 285 zones, with a macroscopic dynamic traffic model. They show the importance of taking into account both the traffic demand and road capacity uncertainties at the same time, in order to better evaluate the effect of the traffic management measures on a road of interest. They also show the uncertainty of network delay time (discrepancy between computed travel time and free flow travel time), and the distribution of travel times on that road. This study is interesting and shows the possibility of applying probabilistic approach to macroscopic traffic models at urban scale. However, the uncertainty given in inputs are focus on global demand of the network and ignores the spatial uncertainty between different O-D pairs. In addition, since the approach in Calvert et al. [2018] was used to test a traffic management measure on a specific road, only the capacity uncertainty on that road is taken into account in their probabilistic approach. Spatial uncertainty of different O-D

traffic demand and of different road capacities might also affect the uncertainty of traffic times at urban scale.

Based on the literature reviews, it can be concluded that in the context of traffic management and transportation planning, uncertainty analysis in traffic models has become an important issue. However, most emphasis is given to traffic demand or link travel time estimation. During the latest two decades, increasing attention is paid to the environmental effects of road traffic. The uncertainty analysis in modeling road traffic emissions is becoming an essential topic. For estimating road traffic emissions however, uncertainty in transportation-related inputs mostly lie in traffic flows and vehicle travel speed. There is few analysis that deals with the uncertainty of street-resolution traffic flow based on macroscopic traffic models. There are some researches that focus on uncertainty analysis of microscopic traffic models for better taking into account the traffic related uncertainties for estimating CO_2 emissions Zhu and Ferreira [2013], fuel consumption and emissions of NOx, particulate matter [da Rocha et al., 2015]. These studies are carried out at microscopic-scale, for one motorway or several trajectories. There is no study that focuses on the uncertainty quantification for macroscopic traffic models or traffic emission estimations at full-city scale. In addition, in the few existing uncertainty analyses for traffic models or emission models, only the distributions of the outputs are given. These distributions are obtained by ensemble simulations using a probabilistic framework. There is no assessment of the generated simulation ensemble in order to evaluate whether the ensemble is capable to well represent the uncertainty or not. This PhD work tackles these challenges, and proposes a complete methodological framework to address the uncertainty quantification in modeling road traffic and its emissions at urban scale. It also presents a corresponding ensemble evaluation method based on real-world measurement data, with a case study for a whole agglomeration area.

1.5.3 Different types of errors

There are increasing available data to measure traffic volume, travel times, vehicle speeds and vehicle trajectories of the network, measured from loop detectors, cameras, Global Positioning System (GPS), Vehicle Identification Systems, and floating car data, etc. Estimating the *error* of the models is an important issue for evaluating model performance. We tend to use the discrepancy between the simulated value and observed value to measure the *error*. However, when we compare the simulated results with measurements, the *error* can be in fact decomposed into three parts: measurement error, representativeness error and modeling error [Garaud and Mallet, 2012].

Here we take traffic flow as an example to illustrate different types of errors when we compare simulated results with measurements at a single time. Let $X \in \mathbb{R}^D$ be the vector of computed traffic flows of the modeled network $\mathcal{G}$, with D the number of links in the network. Let $X^{true} \in \mathbb{R}^D$ be the *true* state of the traffic flow vector. Let $O \in \mathbb{R}^d$ be the vector of induction-loop detector measurements on the real-world network, with d the number of detectors on the network. Let $O^{true} \in \mathbb{R}^d$ be the real number of vehicles passing through the positions of the detectors during a unit of time. When comparing a simulated value with an observed one, we use an operator H that maps the simulation space into the observation space, so that HX can be compared with O. The discrepancy between the computed vector and observed vector is

$$e = O - HX, \tag{1.4}$$

and it can be decomposed as

$$e = \underbrace{O - O^{true}}_{\text{measurement error}} + \underbrace{O^{true} - HX^{true}}_{\text{representativeness error}} + H(\underbrace{X^{true} - X}_{\text{modeling error}}). \tag{1.5}$$

Measurement error

With our example here, the measurement error can be denoted as $e_o = O - O^{true}$. The measurement error is mainly due to the limitations of the observation instruments, malfunctions, errors in the calibration of the instruments, and errors in the postprocessing of the raw measurements. Taking the induction-loop detectors for example, the aggregated data is sent to the data center during every time step of Δh. There might be errors in the data postprocessing during Δh. Different choices of Δh can lead to different *aggregated* traffic flow [Jacobson et al., 1990]. Other measurement errors in traffic modeling are reviewed in de Dios Ortúzar and Willumsen [2011] (page 65).

Representativeness error

In fact, when we compare the simulation results with measurements, we are comparing two vectors of two spaces. The discrepancy between the *true* value of the observation and the *true* value of the link-level traffic flow refers to the *representativeness error*. A loop detector measures road traffic flow at lane resolution at a certain point of a road, while the computed traffic flow on a link is the average flow on the whole segment. In addition, for DUE-based traffic models, the computed traffic flows are at link resolution. For roads where the number of detectors is less than the number of lanes, the operator

H aims to convert X so that HX is comparable with O, and this conversion can lead to *representativeness error* as well.

Modeling error

The modeling error is mainly due to the assumptions, simplifications and parameterizations of a mathematical model, when we use the latter to interpret a complex physical system. With increasing development of numerical technologies, a computational approach is often used to approximate the solution of complex analytical mathematical models. This computational model can bring numerical error, which is also a part of modeling error. When quantifying the *a priori* uncertainty of a model, we are actually trying to determine the distribution of the modeling error. Ensemble simulations give a discrete probability distribution of X. We can use the standard deviation of X to measure the uncertainty, or use the relative standard deviation $\sigma_X/\bar{X}$, with $\bar{X}$ the mean value of X.

Conclusions

This chapter reviews the literature related to the PhD work. It gives an introduction to models and methods used in this PhD work, and highlights the originality of this PhD work compared with existing studies. Firstly, a review of macroscopic static and dynamic user-equilibrium based traffic assignment models is given. We compare static traffic assignment models with dynamic models (DTA models), and also compare dynamic user equilibrium based (DUE-based) macroscopic DTA models with hydrological models. We conclude that for carrying out DTA simulation on a large-scale network for a whole city, DUE-based models are more suitable in an operational context. In addition, they are more suitable for a coupling with an emission model in order to estimate traffic emissions at metropolitan scale with high spatial and temporal resolutions. Then a DUE-based DTA model named LADTA is introduced with an application example on a two-link network.

Secondly, a review of metamodeling is given, focused on three important methods: polynomial regression, Kriging and radial basis functions (RBF). Their applications to traffic modeling are also reviewed. In the recent decades, there have been increasing studies on applying metamodeling methods to traffic models in order to deal with the computational burden of DTA models. Some metamodeling methods are used for speeding up optimization algorithms and to improve model calibration methods. Some others are used in order to propose a surrogate model based on complex traffic flowing

models, in order to speed up DUE-based traffic models. However, in the cases where large numbers of model evaluations are need, it is necessary to build a faster metamodel, which directly maps the input of a DTA model applied to traffic flows and travel time on a large-scale network. This allows the application of Monte Carlo simulations so that quantitative global sensitivity analysis and uncertainty quantification be possible. Based on these needs, a method for building such a metamodel combing RBF method and a dimension reduction is proposed in this PhD work. Then a review of state-of-the-art emission models is given. Among different emission models, an average-speed-based emission model, COPERT method, is adopted for coupling with LADTA model. This choice is made based on the objective of estimating the road traffic emissions throughout a whole city scale.

Finally, a review regarding important methods for sensitivity analysis and uncertainty quantification are given, especially their applications in traffic modeling and on-road emission estimations. Concerning sensitivity analysis, there are many research works related to traffic models. However, most of them focus on local sensitivity analysis based on analytical models. Some other studies focus on global sensitivity analysis of microscopic traffic simulators. The application of quantitative global sensitivity analysis to DTA simulation on a full city size network needs to be explored. This PhD work deals with these challenges, and presents global sensitivity analyses of traffic models and emission estimations at urban scale with high spatial and temporal resolutions. Concerning uncertainty quantification, taking into account the uncertainty in traffic modeling is not a new topic. However, the existing uncertainty analyses of traffic models are carried out either for traffic demand forecasts, or for travel time estimations. These might be adequate in the operational context for transportation management and planning, but are insufficient for studying the uncertainty propagation from the traffic models to traffic emission estimations, especially at urban scale with high spatial and temporal resolutions. Uncertainty quantification in estimating on-road traffic and its emissions is becoming increasingly important in the context of better assessing the environmental impact of road transportation, with consideration of input uncertainty lying in both traffic models and emission models. This requires applying probabilistic approaches in order to generate two ensembles of both DTA simulations and emission estimations at urban scale. Furthermore, the ensemble evaluation using measurements is also important for uncertainty quantification. It has not yet been addressed in the existing uncertainty analyses in dynamic traffic assignment models or traffic emission models, but similar studies have already been successfully carried out for air quality models at urban scale. Therefore, compared with existing studies, the uncertainty quantification

in modeling on-road traffic and its emissions at urban scale is on the one hand original yet with a clear theoretical background, and on the other hand tractable with a thorough case study of a full agglomeration. This PhD work also aims to show how the uncertainties in the inputs of traffic models are propagated to the final traffic emission estimations, through a complete modeling chain combining a traffic model and an emission model at urban scale.

Chapter 2

A dynamic traffic assignment model: LADTA and its implementation LTK

Summary

This chapter presents the most important definitions, assumptions and formulations of the dynamic user-equilibrium (DUE) traffic assignment problem. Then a DUE-based traffic assignment model named LADTA is presented in detail, with its inputs, outputs and assumptions. The framework of the implementation of LADTA model, LTK (for ***L****ADTA* ***T****ool* ***K****it*), is also presented. Finally, a DTA simulation using LADTA is presented with an example of a two-link network.

Résumé

Ce chapitre présente les définitions, les hypothèses et les formulations les plus importantes pour l'affectation dynamique du trafic basée sur le principe de l'équilibre dynamique de Wardrop (*dynamic user-equilibrium*, i.e., DUE). Un modèle de ce type d'affectation, LADTA, est ensuite présenté, avec ses entrées, sorties, les hypothèses principales associées, et son logiciel LTK (pour ***L****ADTA* ***T****ool* ***K****it*). Pour finir, un exemple est présenté pour illustrer le modèle LADTA, sur d'un cas d'école avec un réseau composé de deux arcs et deux noeuds.

Contents

2.1 Dynamic user-equilibrium assignment problem

In dynamic traffic assignment (DTA) models, the traffic demand, travel costs, link parameters (capacities and speed limits) and route choice strategies are time-dependent [Chiu et al., 2011; Peeta and Ziliaskopoulos, 2001]. For the DTA problem, the Wardrop equilibrium principle can be extended to the *dynamic user equilibrium (DUE)* principle as follows.

Under equilibrium conditions in networks where congestion varies over time, traffic arranges itself so that at each instant the costs incurred by drivers on those routes that are used are equal and no greater than those on any unused route [de Dios Ortúzar and Willumsen, 2011]. We present some general notations and definitions of DUE traffic assignment problem in this section. They are mainly summarized from Leurent [2003]; Aguiléra and Leurent [2009]; Meunier and Wagner [2010]; Wagner [2012]; Aguiléra [2014].

- $\mathcal{G} = (\mathcal{N}, \mathcal{A})$: an oriented graph to model the road network. $\mathcal{N}$ is the set of nodes, and $\mathcal{A}$ is the set of links. The total number of links is $D = card(\mathcal{A})$.
- $q_{(o,d),o,d\in\mathcal{N}} \in \mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$: the traffic demand (in veh h^{-1}) from node o to node d. Note that $q_{(o,d)}$ and $q_{(d,o)}$ represent traffic demands of opposite directions.
- $\mathcal{H}$: bounded time interval for carrying out DTA simulation.
- $\mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$: set of positive measurable functions on $\mathcal{H}$.
- $\mathcal{C}(\mathbb{R}_+)$: set of all continuous maps from $\mathbb{R}_+$ to $\mathbb{R}$.
- $\mathcal{M}(\mathbb{R})$: the set of measures on the set $\mathbb{R}$. Given a measure $\mathcal{M}$ on $\mathbb{R}$ defined by $\mathcal{M}(]-\infty, h]) = \int_{-\infty}^{h} m(\tilde{h})d\tilde{h}$, then the map m in $\mathcal{L}^1(\mathbb{R}, \mathbb{R}_+)$ is the *density* of $\mathcal{M}$.
- $(y_a(h))_{a\in\mathcal{A}}$: traffic flow at h, $y(h) \in \mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$.
- Accumulated traffic flow $(Y_a)_{a\in\mathcal{A}}$: $Y_a([-\infty, h]) = \int_{-\infty}^{h} y_a(\tilde{h})d\tilde{h}$. $Y_a(h)$ is the number of users having entered link a until h. $(y_a(h))_{a\in\mathcal{A}}$ is the *density* of $(Y_a)_{a\in\mathcal{A}}$.
- $(t_a)_{a\in\mathcal{A}}(Y_a)(h)$: link travel time functions for the network $\mathcal{G}(\mathcal{N}, \mathcal{A})$. They express the time needed to travel along link a when travel begins at h. They are defined from $\mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$ to $\mathcal{C}(\mathbb{R}_+)$. They can be obtained by congestion models mentioned in Section 1.1.2
- Link exit time function: $H_a(Y_a)(h) := h + t_a(Y_a)(h)$ for $Y \in \mathcal{M}(\mathbb{R})$ and $h \in \mathcal{H}$. Given a cumulated flow Y_a, $H_a(Y_a)(h)$ is the instant at which the link a is left if it has been entered at h. Given that $\mathcal{H} \subset \mathbb{R}$, the subset $H_a(Y_a)^{-1}(\mathcal{H})$ is all the instants at which link a can be entered in order to leave it at some instant in $\mathcal{H}$.
- $\mathcal{S} := \mathcal{H} \times R$: the *strategy* set, where R is the set of acyclic directed paths (the set of routes). Spatially, $R_{(o,d)}$ is the route from the original node o to the destination

node d. Each network users chooses his strategy from $\mathcal{S}$: (i) a departure time $h \in \mathcal{H}$ and (ii) a route $r = a_1, a_2, \ldots, a_n$, with $a_1, a_2, \ldots, a_n \in \mathcal{A}$. User goes through links in the order that they appear in the route: entering a_{i+1} after leaving the link a_i.

– $X \in \mathcal{M}(\mathcal{S})$: a distribution of users strategy.

2.1.1 Definition of dynamic flowing model

We define the dynamic traffic flowing model by the *cumulated flow function* to express the interaction between t_a and Y_a. The link cumulated flow $(Y_a)_{a\in\mathcal{A}}$ induced by a route choice strategy distribution $X \in \mathcal{M}(\mathcal{S})$ is a collection of measures on $\mathbb{R}$ such that there exists $(Y_a^r)_{a\in\mathcal{A},r\in R}$ that satisfies the system

$$Y_a = \sum_{r\in R: a\in r} Y_a^r, \tag{2.1}$$

and for all $r = a_1, \ldots, a_i, \ldots, a_n \in R$,

$$Y_{a_1}^r = X^r, \tag{2.2a}$$

$$Y_{a_i}^r = Y_{a_{i-1}}^r \circ (H_{a_{i-1}}(Y_{a_{i-1}}))^{-1}, i = 2, \ldots, n, \text{and} \tag{2.2b}$$

$$Y_a^r = 0 \text{ if } a \notin r, \tag{2.2c}$$

where X^r is the cumulated number of users entering into the route r for some instant in $\mathcal{H}$. Y_a^r is the cumulated flow on link a with respect to route r during $\mathcal{H}$. In other words, Y_a^r is the number of users whose chosen route is r and who enter the link a at some instant in $\mathcal{H}$. Equation 2.2 states that the users going through link a can be decomposed by routes. Equation 2.2a means that the number of users entering the first link of route r is the number of users entering the route r during $\mathcal{H}$. Equation 2.2b is the conservation of number of users from link a_{i-1} to link a_i during $\mathcal{H}$. Equation 2.2c means that if link a does not belong to a route r, there is no user having chosen r that has traveled or will travel along a. With some assumptions for link travel time function t_a, Meunier and Wagner [2010] proved that the existence and uniqueness are guaranteed for solutions of the system 2.1-2.2 for all $X \in \mathcal{S}$. Link-travel time functions used in LADTA satisfy these assumptions and they are presented later in Section 2.3. Then the dynamic flowing model associated to each link a is defined as

$$\Phi_a(X) := Y_a, \text{ for all } a \in \mathcal{A} \text{ and } X \in \mathcal{S}. \tag{2.3}$$

With the above definitions, the *route exit time function* for all $r = a_1, \ldots, a_i, \ldots, a_n \in R$, is defined as

$$H^r(X) := H_{a_n}(Y_{a_n}) \circ H_{a_{n-1}}(Y_{a_{n-1}}) \circ \cdots \circ H_{a_1}(Y_{a_1}). \tag{2.4}$$

H^r depends on the whole measure on the strategies while H_a only depends on the cumulated flow on the link a. From Equation 2.3 and Equation 2.4, we can get the route exit time as

$$H^{a_1,\ldots,a_n}(X) = H_{a_n}(\Phi_{a_n}(X)) \circ H_{a_{n-1}}(\Phi_{a_{n-1}}(X)) \circ \cdots \circ H_{a_1}(\Phi_{a_1}(X)). \tag{2.5}$$

2.1.2 Dynamic User Equilibrium (DUE) assignment of traffic

We define the *dynamic assignment of traffic* $x = (x)^r_{r \in R_{(o,d)}} \in \mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$ such that $\sum_{r \in R_{(o,d)}} x^r(h) = q_{(o,d)}(h)$ for all $(o,d) \in \mathcal{N} \times \mathcal{N}$ and $h \in \mathcal{H}$. $R_{(o,d)}, o \in \mathcal{N}, d \in \mathcal{N}$ is the set of routes connecting o to d. Then we define the *route travel time function* t_r:

$$t_r(X)(h) := H^r(X)(h) - h, \tag{2.6}$$

where X is the measure whose density is x. $t_r(X)(h)$ is then the time needed to travel through the route r when leaving at instant h for an assignment x. $H^r(X)$ is the *route exit time function* defined in Definition 2.5 with $r = a_1, \ldots, a_i, \ldots, a_n \in R$. In fact, the users route choice strategy can be represented by X^r: the *cumulated flow* of vehicles (users). $X^r[-\infty; h]$ counts the number of users that have already entered route r. Then we can define the *DUE assignment of traffic* as follows.

Definition 2.1.1. (DUE assignment problem) Find an assignment $x \in \mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)^R_{(o,d)}$ such that whenever $r, r' \in R_{(o,d)}$,

$$x^r(h) > 0 \Rightarrow t_r(X)(h) \leq t_{r'}(X)(h) \text{ for almost every } h \in \mathcal{H}, \tag{2.7}$$

where $t_r(X)(h)$ is defined in Definition 2.6. The assignment is called *dynamic* because the equilibrium is established at every instant $h \in \mathcal{H}$.

Figure 2.1 shows the structure of DUE-based assignment.

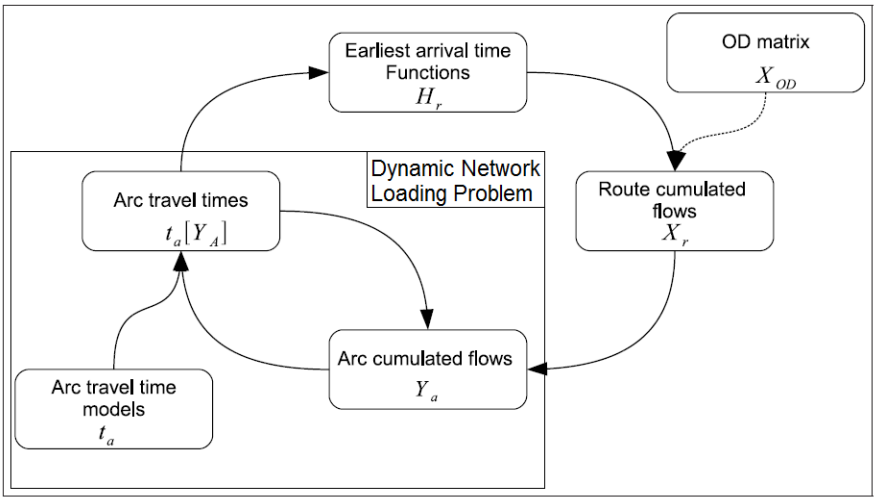


Figure 2.1: Structure of DUE-based assignment. Source: [Leurent, 2003; Wagner, 2012]

2.2 Inputs and outputs of LADTA model

LADTA model is a DUE-based assignment model. Inputs of LADTA model are defined here for carrying out a DTA simulation with LADTA model during the period $\mathcal{H}$.

- **Network**: oriented graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$. $\mathcal{N}$ is the set of nodes, and $\mathcal{A}$ is the set of links. The total number of links is $D = card(\mathcal{A})$.
- **Zones**: $\mathcal{Z}_O \subset \mathcal{N}$ and $\mathcal{Z}_D \subset \mathcal{N}$ representing the set of nodes for Origin-Destination zones. $z = card(\mathcal{Z}_O) = card(\mathcal{Z}_D)$. z is the number of zones in the simulation area. The dimension of the O-D matrix is $z \times z$.
- **Dynamic O-D matrix**: $\mathcal{Q}(h) = (Q_{(o,d)}(h))_{(o,d),o\in\mathcal{Z}_O,d\in\mathcal{Z}_D}$ representing traffic demand $Q_{(o,d)} \in \mathcal{M}(\mathbb{R})$ between origin-destination zones. $Q_{(o,d)}$ represents the *accumulated* demand from o to d.
- **Density of traffic demand during $\mathcal{H}$**: $q^{\mathcal{H}}_{(o,d)}(h) \in \mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$.
- **Vector of capacity, free flow travel time and speed limit**: $\boldsymbol{K} = [K_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$, $\boldsymbol{T}_0 = [T_{0a}(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$, $\boldsymbol{V}_0 = [V_{0a}(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$. The dimension of $\boldsymbol{K}$, $\boldsymbol{T}_0$ and $\boldsymbol{V}_0$ equals to the number of links. K is in veh h^{-1}, T_0 is in h and V_0 is in km h^{-1}.
- **Vector of generalized travel cost**: $\boldsymbol{U}(h) = [U_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$: a vector of generalized travel costs on each link a for time h. In particular, if we consider that users minimize their travel time, $\boldsymbol{U}(h) = \boldsymbol{T}(h) = [T_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$.

The outputs of LADTA model are link travel time and *cumulated traffic flow* for all links. The latter corresponds to the DUE assignment solution satisfying Equation 2.7.

- **Link travel time**: $\boldsymbol{T}(h) = [T_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$.
- **Minimum cost routes**: $\boldsymbol{R}(h) = [R_{(o,d)}(h)]_{(o,d)\in\mathcal{N}\times\mathcal{N},h\in\mathcal{H}}$. In the software LTK (for ***L****ADTA* ***T****ool* ***K****it*), $R_{(o,d)}(h)$ is defined as piecewise linear function of h to give indices of links to take at instant h, where h is the departing instant (clock time) from o.
- **Cumulated flow for route**: $\boldsymbol{X}^r, r \in R_{(o,d)}$. Its density $x^r(h)$ satisfies the definition of DUE in 2.1.1 in Section 2.1.2. In practice, the *density* is the traffic flow expressed in veh h^{-1}.
- **Link cumulated flow**: $\boldsymbol{Y} = [Y_a]_{a\in\mathcal{A}}$. $Y_a \in \mathcal{M}(\mathbb{R})$. Y_a satisfies the equation system 2.1-2.2 in Section 2.1.1.

The outputs of LADTA model are link travel time and *cumulated traffic flow* for all links. The latter corresponds to the DUE assignment solution satisfying Equation 2.7.

- **Link travel time**: $\boldsymbol{T}(h) = [T_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$.
- **Minimum cost routes**: $\boldsymbol{R}(h) = [R_{(o,d)}(h)]_{(o,d)\in\mathcal{N}\times\mathcal{N},h\in\mathcal{H}}$. In the software LTK (for ***L****ADTA* ***T****ool* ***K****it*), $R_{(o,d)}(h)$ is defined as piecewise linear function of h to give indices of links to take at instant h, where h is the departing instant (clock time) from o.
- **Cumulated flow for route**: $\boldsymbol{X}^r, r \in R_{(o,d)}$. Its density $x^r(h)$ satisfies the definition of DUE in 2.1.1 in Section 2.1.2. In practice, the *density* is the traffic flow expressed in veh h^{-1}.
- **Link cumulated flow**: $\boldsymbol{Y} = [Y_a]_{a\in\mathcal{A}}$. $Y_a \in \mathcal{M}(\mathbb{R})$. Y_a satisfies the equation system 2.1-2.2 in Section 2.1.1.

In the DTA simulation for a metropolitan area, we are interested in $\boldsymbol{T}$ and $\boldsymbol{Y}$ because they can give traffic information at each instant h. We can deduce average travel speed at link resolution from $\boldsymbol{T}$ so that the emission estimation at link resolution can be possible. The *density* of $\boldsymbol{Y}$ is traffic flow at link resolution. It can be compared with loop detector measurements to evaluate the model performance. In this PhD work, the *density* of Y_a is compared with loop detector measurements at high time resolution for a one-day DTA simulation and a one-month DTA simulation, with LADTA applied to a real-world network (Chapter 3 and Chapter 4).

2.3 Assumptions in LADTA model

2.3.1 Assumption of traffic demand input

In LADTA model, users are supposed to be *rational* and choose their routes by minimizing their generalized travel cost. They are also assumed to have full travel

information during $\mathcal{H}$, so that they can adapt their route choice strategies according to the time-dependent link travel time. In practice, for transportation planning, the O-D matrix is often given (or modeled) during a bounded period. The traffic demand is constant in the given O-D matrix and the latter is called *static* O-D matrix. It is difficult to get *cumulated* traffic demand $Q_{(o,d)} \in \mathcal{M}(\mathbb{R})$ since the measure $\mathcal{M}$ is defined on the set of $\mathbb{R}$ in the definition of DUE problem, and in practice, the simulation period is always bounded. Therefore, we state the following Assumption 1 so that we can define $q_{(o,d)} \in \mathcal{L}^1(\mathbb{R}, \mathbb{R}_+)$ as the density of Q. Then $Q_{(o,d)} \in \mathcal{M}(\mathbb{R})$ can be deduced from $q^{\mathcal{H}}_{(o,d)}(h) \in \mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$ by imposing its boundary condition (see in Assumption 2).

Assumption 1. *We denote the simulation period as $\mathcal{H} = [h_0, h_1]$ and $h_0 < h_1$. The density of $Q_{(o,d)}$ is zero everywhere on $\mathbb{R}$ except on $\mathcal{H}$, $q_{(o,d)}(h) = q^{\mathcal{H}}_{(o,d)}(h), h \in \mathcal{H}$.*

$$q_{(o,d)}(h) = \begin{cases} q^{\mathcal{H}}_{(o,d)}(h) & \text{if } h \in \mathcal{H} \\ 0.0 & \text{if not} \end{cases}, \forall (o,d) \in \mathcal{Z} \times \mathcal{Z}. \tag{2.8}$$

Therefore, $Q_{(o,d)}$ can be considered as the *cumulated traffic volume* from o to d. We also state the limit condition at the instant h_0 in Assumption 2.

Assumption 2. *The network is empty before the beginning of the simulation period. There is no cumulated traffic demand before h_0 so that $Q_{(o,d)}(]-\infty, h_0]) = 0$. Then $Q_{(o,d)}$ can be obtained based on its density as*

$$Q_{(o,d)}(h) = \int_{-\infty}^{h} q_{(o,d)}(\tilde{h})d\tilde{h} = \begin{cases} 0.0 & \text{if } h \in]-\infty, h_0] \\ \int_{h_0}^{h} q^{\mathcal{H}}_{(o,d)}(\tilde{h})d\tilde{h} & \text{if } h \in [h_0, h_1] \\ \int_{h_0}^{h_1} q^{\mathcal{H}}_{(o,d)}(\tilde{h})d\tilde{h} = Q_{(o,d)}(h_1) & \text{if } h \in]h_1, +\infty[\end{cases}. \tag{2.9}$$

In practice for a DTA simulation, the *density* ($q_{(o,d)}$) can be considered as the temporal variation of traffic demand on the network, which is always non-negative. If we know the total cumulated traffic demand during $\mathcal{H}$ for each O-D pair, i.e. $Q_{(o,d)}(h_1)$, with the Assumption 1 and Assumption 2, the input $\boldsymbol{Q}_{o,d}(h)$ can be obtained from $(q_{(o,d)}(h))_{o \in \mathcal{Z}_O, d \in \mathcal{Z}_D}$. This helps us to establish the *dynamic* O-D matrix based on a static one. Detailed approach for converting a static O-D matrix to a dynamic one is presented in Section 3.3.2 with the case study for the agglomeration of Clermont-Ferrand.

2.3.2 Assumption of link travel time model

In LADTA model, a first-in-first-out (FIFO) bottleneck queuing model is used to model the link travel time [Vickrey, 1969; Kuwahara and Akamatsu, 1993; Mounce, 2006, 2007]. It is proved to satisfy the assumptions that can ensure the existence of DUE [Mounce, 2007; Meunier and Wagner, 2010]. If the income flow at the bottleneck exceeds the capacity, a queue began to form and users have to wait according to a FIFO rule before leaving the bottleneck. The waiting time is called *bottleneck delay*. The link travel time T_a on the link a for an entrance time h is then the sum of (i) a constant travel time and (ii) a *bottleneck delay*. Let $L(h)$ denote the number of users queuing at h in the bottleneck. $x_a(h)$ is the inflow of the link a (in veh h^{-1}).

$$T_a(h) = T_{0a} + \frac{L(h)}{K_a(h)}, \tag{2.10}$$

where L can be obtained by the following differential equation:

$$\frac{dL}{dh}(h) = \begin{cases} x_a(h) - K_a(h) + \frac{L(h)}{K_a(h)} & \text{if } L(h) \neq 0 \text{ or } x_a(h) - K_a(h) > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.11}$$

The point queue indeed ignores the spillback of queues. However, it can better model the travel time in congestion when compared with volume/delay functions in static models. If $\mathcal{H}$ is long enough to cover the congestion period, it is verified that the cumulated traffic demand of all O-D pairs equals to the sum of computed cumulated link-level $\boldsymbol{X}$ [Aguiléra and Leurent, 2009]. In other words, $\sum_{(o,d)\in\mathcal{Z}\times\mathcal{Z}} Q_{(o,d)}(h_1) = \sum_{a\in\mathcal{A}} X_a(h_1)$, where h_1 is the end of the simulation period $\mathcal{H}$. For a DTA simulation at metropolitan scale, the use of point queue allows more efficient algorithm in order to carry out DTA simulation for a real-world network [Aguiléra and Leurent, 2009; Chen et al., 2017].

2.4 Structure of LADTA model and main algorithm framework of LTK

There are in total 5 main steps in the main loop of LTK framework. LATDA model and LTK can also take highway toll into account for modeling generalized travel cost at link resolution. In the case where there are toll highways, it is necessary to initialize toll prices and add toll price into the model for calculating generalized cost of each link.

In this case, the generalized cost can be modeled as a function of travel time and the *value of time* coefficient β of users at instant h:

$$C_{a,general}(h) = T_a(h) + v_t \cdot P_a(h), \tag{2.12}$$

where P_a is the toll price of link a and v_t is the *value of time* in *monetary unit per unit of time* (e.g., €/h). Then the generalized travel cost can be expressed in time unit. The following steps present the main simulation loop of LTK, summarized from [Aguiléra and Leurent, 2009].

- **Step 0: Initialization**
  Initialization of the modeled network $\mathcal{G}$ and cumulated demand volume to all the destination nodes d: $(\boldsymbol{Q}_d)_{d\in\mathcal{Z}_D}$. The initial link travel time is free-flow travel time $\boldsymbol{T_0}$.
- **Step 1: Formation of costs**
  The travel cost is computed as a function of the link travel time computed from the last iteration for link a and the toll price: $C_{a,k}(h) \leftarrow T_{a,k-1}(h) + \beta \cdot P_a(h)$.
- **Step 2: Route choice (RC)** $R_{\mathcal{I},k}(h) \leftarrow RC(C_{a,k}(h), T_{a,k-1}(h), Q_d)$
  This step computes the dynamic least-cost routes for every (o, d) pair for each destination $d \in \mathcal{Z}_D$, departing at instant h from every node $o \in \mathcal{N}$. At iteration k, this step saves the indices $\mathcal{I}_d(h)$ of least-cost path as a function of h for each destination d, composed of links from $\mathcal{A}$.
- **Step 3: Volume loading (VL)**
  Routes computed in **step 2** are used here with the help of $\mathcal{I}_d(h)$. This step computes cumulated traffic flow $X_{a,k}$ with $a \in \mathcal{I}_d$ when entering the link a at h: $X_{a,k}(h) \leftarrow VL(T_{a,k-1}(h), Q_d, R_{\mathcal{I},k}(h))$. This step composes of a complex dynamic network loading problem (DNLP). In the implementation of LTK, all-or-nothing assignment with Djkastra's algorithm [Leurent and Aguiléra, 2009] is used to get auxiliary traffic flow $Y_{a,k}(h)$ as a function of $Q_d(h)$. Then a convex combination algorithm [Leurent, 2003] is used to compute $X_{a,k}(h)$ as a function of $Y_{a,k}(h)$ and $X_{a,k-1}(h)$: $X_{a,k}(h) = w_k \cdot Y_{a,k}(h) + (1 - w_k) \cdot X_{a,k-1}(h)$.
- **Step 4: Traffic flowing (TF) model**
  This step computes link travel time at iteration k as a function of loaded traffic flow on link a from the **step 3**. The flowing model used in LADTA is the bottleneck model and the link travel time is a function of loaded cumulated flow, link capacity and free flow travel time. $T_{a,k}(h) \leftarrow TF(X_{a,k}(h), K_a(h), T_{0,a})$.

A stopping criterion is defined to decide whether to exit the main loop or to do one more iteration. The final outputs are cumulated flows and travel times on each link

$a \in \mathcal{A}$: $\boldsymbol{X}_{a,k}(h)$ and $\boldsymbol{T}_{a,k}(h)$. Figure 2.2 shows the algorithm framework of LTK model.

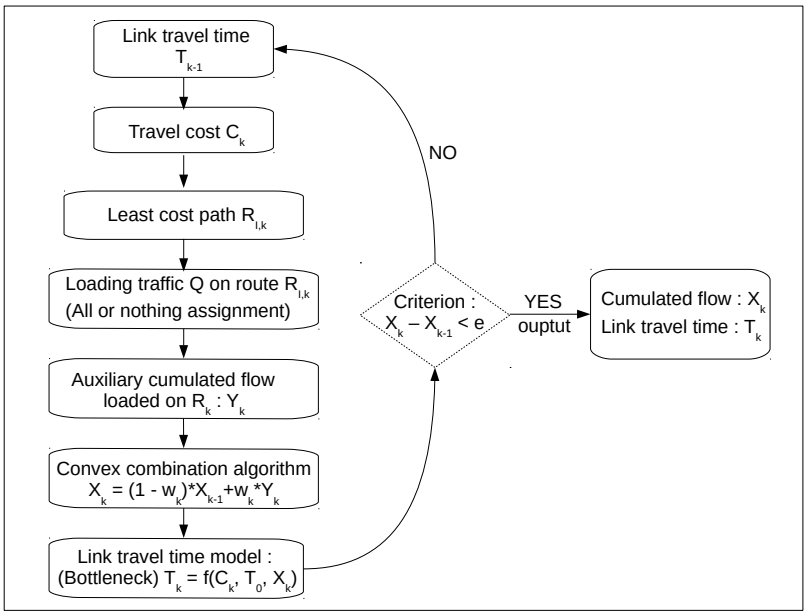
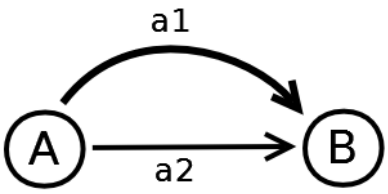


Figure 2.2: The algorithm framework of LTK software

2.5 Example of two-link DTA assignment with LADTA

A network with two links (a_1 and a_2) and two nodes (A and B) is presented in Figure 2.3 (left). The cumulated traffic demand is presented in Figure 2.3 (right). The simulation period $\mathcal{H} = [0.0, 2.0]$. The *density* of traffic demand is

$$q_{(A,B)}(h) = \begin{cases} 0.0 & \text{if } h < 0.0 \\ 500.0 & \text{if } h \in [0.0, 1.0] \\ 1500.0 & \text{if } h \in [1.0, 2.0] \\ 0.0 & \text{if } h > 2.0 \end{cases} \tag{2.13}$$



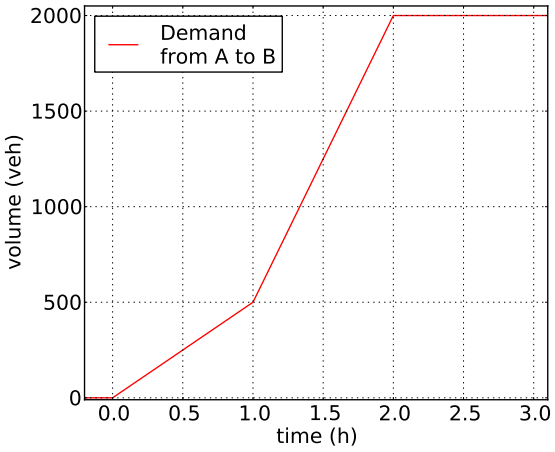


Figure 2.3: Network and cumulated traffic demand (in veh) of two links and two nodes. The link a_1 is longer but it has a higher capacity (illustrated by its width). The simulation period is $[0.0, 2.0]$ hour. Before 0.0 h and after 2.0 h, the traffic demand is 0.0. During [0.0, 1.0] hour, the cumulated traffic demand increases with a density of 500 veh h^{-1}. During [0.0, 1.0] hour, the cumulated traffic demand increases with a density of 1500 veh h^{-1}.

Link capacity and free flow travel time is presented in Table 2.1.

Table 2.1: Capacity and free flow travel time of the two-link network

Link	Origin	Destination	Free flow travel time (h)	Capacity (in veh h^{-1})
a_1	A	B	0.5	1000.
a_2	A	B	0.4	500.

According to the given link capacity and free flow travel time, $T_{0a_1} > T_{0a_2}$ and $K_{0a_1} > K_{0a_2}$. Therefore, when users departing from node A during $[0.0, 1.0]$, the *density* of traffic demand is 500.0 veh h^{-1}. Users choose the route that minimizes their travel time: a_2. For users leaving after 1.0 h, the traffic demand density is 1500.0 veh h^{-1} and it is bigger than the link capacity of a_2. The queuing time at the end of a_2 is $\frac{x_{a_2}-K_{a_2}}{K_{a_2}} = \frac{1500-500}{500} = 2$ h per unit of time (per hour). The difference of link free-flow travel time between a_1 and a_2 is $\Delta T = T_{a_1} - T_{a_2} = 0.1$ h. With the DUE condition, the final assignment will lead to $T_{a_1} = T_{a_2}$. Let h_c define the entering instant into the network so that the travel times on a_1 and a_2 are the same. Therefore, h_c can be calculated from the system 2.14:

$$\begin{cases} T_{a_1}(h_c) = T_{a_2}(h_c) \\ T_{a_1}(h_c) = T_{0a_1} \\ T_{a_2}(h_c) = T_{0a_2} + \frac{x_{a_2}(h_c) - K_{a_2}}{K_{a_2}}(h_c - 1.0\ hour) \quad h_c \in [1.0, 2.0]\ hour \\ T_{0a_1} = 0.5\ hour \\ T_{0a_2} = 0.4\ hour. \end{cases} \tag{2.14}$$

The resulting $h_c = 1.0 + \frac{\Delta T}{2.0} = 1.0 + \frac{0.1}{2.0} = 1.05$ hour. Therefore, users entering node A from the instant 1.05 h will choose the link a_1 and during the period [1.05, 2.0], the link travel time of the two links are the same. The assignment on the network is then $x_{a_1}(h) = 1000\,\text{veh}\,\text{h}^{-1}$ and $x_{a_2}(h) = 500\,\text{veh}\,\text{h}^{-1}$ with $h \in [1.05, 2.0]$ hour, and the incoming demand density $q_{(A,B)} = 1500\,\text{veh}\,\text{h}^{-1}$. When $h > 2.0$ h, the *density* of demand departing from 2.0 h decreases to $0.0\,\text{veh}\,\text{h}^{-1}$ and the travel time on link a_2 decreases back to its free-flow travel time (0.4 h). Figure 2.4 shows the cumulated traffic flow on two links and the corresponding link travel time against the instant (h) entering to the network from node A.

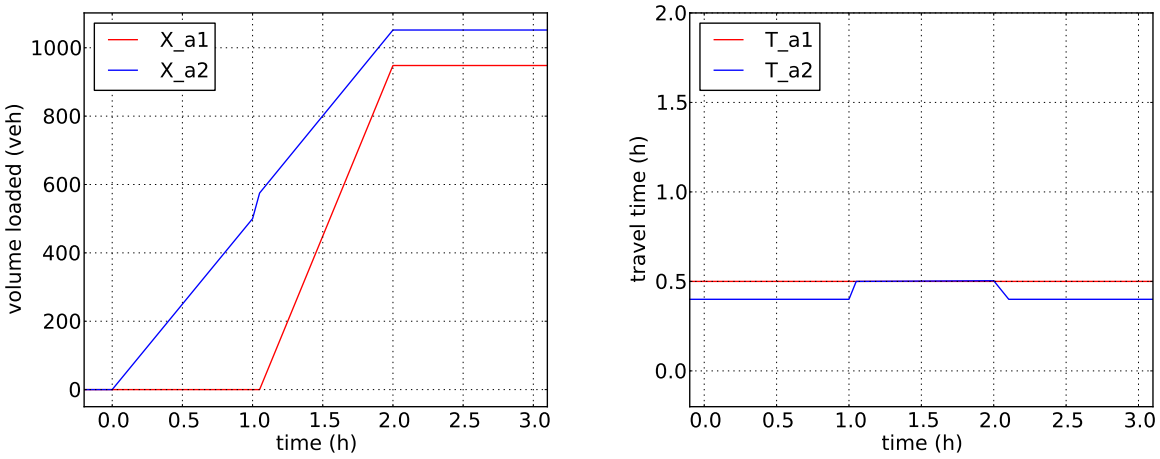


Figure 2.4: Network and cumulated traffic demand (in veh) of two links and two nodes

This is an example of a DTA simulation on an elementary network with LADTA model and its software. The model is then applied to the network of the agglomeration of Clermont-Ferrand. Detailed model setting and model evaluation results are presented in Chapter 3.

Conclusions

This chapter introduces the basic definitions and formulas for dynamic traffic assignment (DTA) problem based on dynamic user equilibrium (DUE) principle. The latter interprets an equilibrium at each instant between traffic demand and network supply. Under DUE condition, traffic arranges itself so that at each *instant* travel costs of network users on those routes that are used are equal and no greater than those on any unused route. The LADTA model used in this PhD work is a DUE-based model. It can be applied to a road network for carrying out DTA simulation in continuous time during a given simulation period. The traffic demand inputs in LADTA are *cumulated* traffic flow between origin-destination zones, which vary continuously as a function of the *clock time* when users leave from the original zone. The outputs of LADTA are (i) *cumulated* traffic volume that are passing through each of the links and (ii) travel time on each of the links in the network, as a function of the *clock time*. These outputs can give traffic information of a network with high spatial and temporal resolutions. These two outputs can then be coupled with emission models at link-resolution, in order to compute atmospheric pollutant emissions due to road traffic for a large-scale network. The DTA simulation of road traffic for a metropolitan-scale network is presented in Chapter 3, and the estimation of on-road traffic emissions is presented in Chapter 6.

Chapter 3

Dynamic traffic assignment for the agglomeration of Clermont-Ferrand

Summary

Two dynamic traffic assignments (DTA) with LADTA model are carried out in the agglomeration of Clermont-Ferrand (France): (i) a whole-day simulation on the 20^{th} November 2014, and (ii) a whole-day simulation during a working Tuesday. The use of dynamic LADTA model allows us to visualize the temporal variation of traffic flow in the agglomeration with high temporal resolution down to 1 minute. It is observed that during morning peak period, traffic on the network can change significantly and bring congestions even within a quarter of an hour. A reference static traffic assignment using VISUM model is also carried out for the same agglomeration. The traffic flows at street resolution computed by LADTA model are compared with those of VISUM model during the evening peak hour. The results of both models are compared to loop detector measurements. Statistical scores show that traffic flows computed by both models have similar performance when compared with observations. The whole-day simulation during a working Tuesday is also compared with loop detector measurements of all the working Tuesdays from September 2014 to July 2015, at lane resolution at all time intervals of 15 minutes. Results show that LADTA model preserves well the temporal variation of traffic flows on the network. It has however some limitations when predicting route choices of users, due to uncertainties in the O-D matrix and the assumptions in the model. A sensitivity analysis is also carried out to the DTA simulation with LADTA. We analyzed throughout the network during a whole day the influences on the computed vehicle travel time (in veh·h) of (i) the total demand in the O-D matrix and (ii) the speed limits on the network Results show that the total vehicle travel time is very sensitive to the total demand. It is less influenced by the speed limits.

Résumé

Deux affectations dynamiques du trafic sont réalisées avec le modèle LADTA pour l'agglomeration de Clermont-Ferrand: (i) l'une pour le 20 novembre 2014, and (ii) l'autre pour une journée entière d'un mardi *normal* hors vacances scolaires ou jours fériés. L'utilisation du modèle dynamique LADTA nous permet de visualiser la variation temporelle du débit de trafic avec une résolution temporelle fine jusqu'à 1 minute. Les résultats montrent que pendant la période de pointe du matin, l'augmentation du trafic trafic peut être importante et peut rendre le réseau congestionné pendant un quart d'heure. Une affectation de trafic statique de référence est également effectuée avec le modèle VISUM pour la même agglomération. Les débits simulés par LADTA sont comparés à ceux du modèle VISUM pendant l'heure de pointe du soir à la résolution de la rue. Les résultats des deux modèles sont aussi comparés aux observations obtenues par les boucles de comptages sur le réseau routier de Clermont-Ferrand. Les scores statistiques montrent que les débits simulés par les deux modèles ont des performances similaires par rapport aux observations. La simulation d'un mardi *normal* est également comparée avec observations de tous les mardis ouvrés de septembre 2014 à juillet 2015, à la même résolution que les comptages pendant toutes les 15 minutes. Les résultats montrent que le modèle LADTA peut bien prédire la variation temporelle des débits sur l'ensemble du réseau. Il est toutefois constaté certaines limites lors de la simulation des choix d'itinéraire par les usagers. Ces limites spatiaux sont dues aux incertitudes dans la matrice O-D, aux hypothèses du modèle, etc. Une analyse de sensibilité qualitative est ensuite réalisée pour la simulation avec LADTA. Nous avons analysé la sensibilité du véhicule-heure total (en veh·h) de tous les usagers sur le réseau par rapport aux variations des deux entrées principales: (i) la demande totale dans la matrice O-D et (ii) les vitesses maximales autorisées sur le réseau. Les résultats montrent que le véhicule-heure total est très sensible à la variation de la demande totale en entrée. Il n'est néanmoins pas sensible de manière significative aux vitesses maximales autorisées sur le réseau.

Contents

Some results of this chapter are published in the article Chen, R., Aguiléra, V., Mallet, V., Cohn, F., Poulet, D., and Brocheton, F. (2017). A sensitivity study of road transportation emissions at metropolitan scale. *Journal of Earth Science & Geotechnical Engineering*, 7(1):151-173.

3.1 Introduction

Traffic assignment aims at determining the network traffic flows according to network users' route choices when they travel from their origins to their destinations. It can also be considered as an economical equilibrium between the demand and the supply. The Origin-Destination matrix (O-D matrix) describes the total fluxes from origin zones to destination zones. The road network shows limited capacities to absorb all the demand. One of the main hypotheses for traffic assignment problems is that every network user makes their route choice by minimizing their own travel cost, such as travel time, toll, etc. The cost is often converted to one general criterion, such as time, which is called generalized cost of route choice. At equilibrium when every travel succeeds in finding such a route, all used routes associated with the same O–D pair should have the same minimum generalized cost, so that there is no possibility for users to shift to another route. This is the user equilibrium condition [Wardrop, 1952], and it was proved to be adequate for both static and dynamic traffic assignment models, with a certain number of assumptions [Leurent, 2003; Aguiléra and Leurent, 2009; Meunier and Wagner, 2010; Wagner, 2012].

In both static and dynamic traffic assignment models, the travelers are assumed to have complete information about each link in a network (the supply): physical capacity, length, speed limit and toll price (if any), in order to estimate the generalized travel cost of each available route. In a static model, the effect of road capacity on travel time is modeled by a volume-delay function (VDF): the travel cost (or time) is a strictly increasing function of the traffic volume [Bureau, 1964] (see Figure 1.1). In static models, the volume may increase indefinitely and exceed the link capacity. The travel time calculated by VDF in static models does not depend on physical features of congestion (such as travel speed, density, or queue) [de Dios Ortúzar and Willumsen, 2011]. In order to better represent the congestion phenomenon and the temporal and spatial evolution of traffic flow, we use a dynamic traffic assignment (DTA) model.

In DTA simulation, the travel time depends not only on the traffic flow on the link, but also depends on the vehicles' entrance (clock) time on the link. In fact, the traffic already existing on the link can decrease the link capacity. This may result in an increase of link travel time if the actual inflow volume exceeds the actual capacity of the link. There are different methods to calculate the link travel time in DTA, such as the bottleneck modeling where queue appears in upstream junction when the volume assigned to the link exceeds the capacity [Vickrey, 1969]. In this model, the time spent in queue is taken into consideration to compute the link travel time.

The LADTA model is one of the DTA models which can be applied to a large-scale

real network. Together with its implementation, **L**ADTA **T**ool**K**it (LTK), it can handle the DTA problem of a metropolitan network frame within a reasonable time [Aguiléra and Leurent, 2009]. LADTA model has already been applied to the network of Paris region (*Île-de-France*) in Aguiléra and Leurent [2009]. The model is evaluated by comparing the computed average speed on main roads of Paris region with measured average speed. For this chapter, we applied the LADTA model to a network at metropolitan scale for the agglomeration of Clermont-Ferrand. The DTA simulation with LADTA model is evaluated by comparing the computed traffic flows with the results computed by the static model, and with traffic flows measured by loop detectors at lane resolution. In addition, a qualitative sensitivity analysis is also carried out to study the sensitivity of LADTA outputs to the inputs. It studies the sensitivity of the computed vehicle travel time with respect to the variation of the total traffic demand in input and the speed limits of the network.

Section 3.2 presents network user behaviors and data analysis from measurements of loop detectors. Section 3.3 presents inputs for carrying out dynamic traffic assignment (DTA) simulation with LADTA model. Then in Section 3.4, the results of DTA simulation are compared with those of VISUM model, which is a commonly used static assignment model in transportation planning problem. The performance of LADTA model is also evaluated by comparing the computed traffic flows with loop detector measurements. At last in Section 3.5, qualitative sensitivity analysis is carried out for the DTA simulation with LADTA model.

3.2 Traffic measurements and users travel behavior

Clermont-Ferrand has in total 535 inductive loop traffic detectors on the road network, which are generally located in main city boulevards and/or crossroads (see in Figure 3.1). We have filtered out (i) the detectors for which the data are always zero and (ii) the data that are bigger than 4000 veh h^{-1} detector^{-1}. At last, there are in total about 400 detectors in use for analyzing travelers behavior and evaluating LADTA model. The detectors collect the traffic density (in veh km^{-1}) and traffic flow (in veh h^{-1}) every minute. These precious and rich traffic data help us to analyze the network users' behavior and to set up the LADTA simulation. The City of Clermont-Ferrand (*Ville de Clermont-Ferrand*) has provided two years of observational data of every detector with a resolution of 15 minutes, from September 2013 to September 2015. The data from September 2013 to August 2014 have been analyzed in details in order to set up and calibrate LADTA model, and the data from September 2014 to September 2015 were

then used to evaluate the traffic flow results predicted by LADTA.

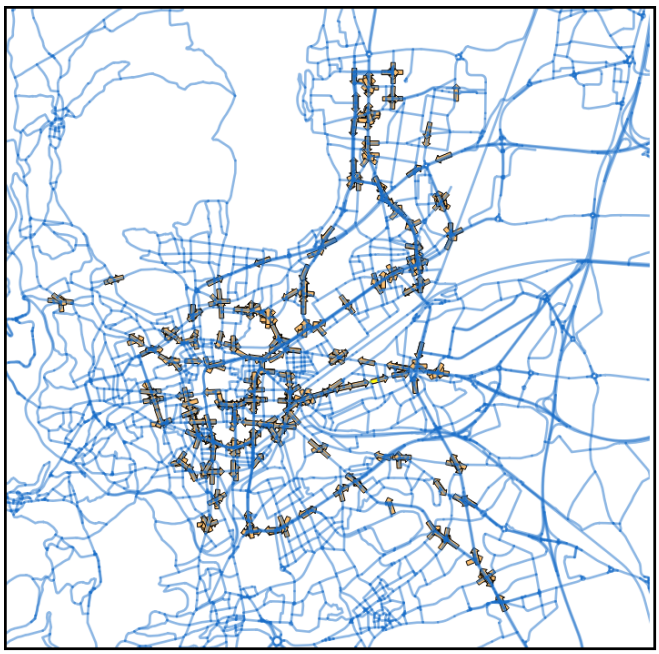


Figure 3.1: Locations and directions of loop detectors in the City of Clermont-Ferrand. The detectors are at lane resolution. The blue lines represent the roads of Clermont-Ferrand. The yellow arrows represent the positions and directions of traffic detectors.

The observational data of the real road traffic help us to better understand the behaviors of road network users. It is observed that the network users in Clermont-Ferrand almost always behave the same for the same type of day. Figure 3.2 shows temporal variation of spatially-averaged flow for all Tuesdays from September 2013 to September 2014. For each interval of 15 minutes, we calculate the spatially-averaged traffic flow over all the 469 detectors, and then plot this average against the clock time of a day (Clermont-Ferrand local time, GMT+1).

The left profile in Figure 3.2 presents the temporal variation of spatially-averaged traffic flow during a normal period. The right profile is the traffic during school vacation period. It is clear that for all working Tuesdays, the morning peak appears almost at the same moment of the day around 08:00. The evening peak hour also appears at the same moment around 17:00. The differences among traffic flows during peak hours of different days are not significant. Moreover, the total volume of traffic during a day remains almost unchanged from one Tuesday to another. The same feature is observed for other working weekdays, and the temporal variation of the same weekday during a year can be easily represented by an average temporal profile (yellow line). Figure 3.3 shows that the temporal variation of a same weekday remains almost the

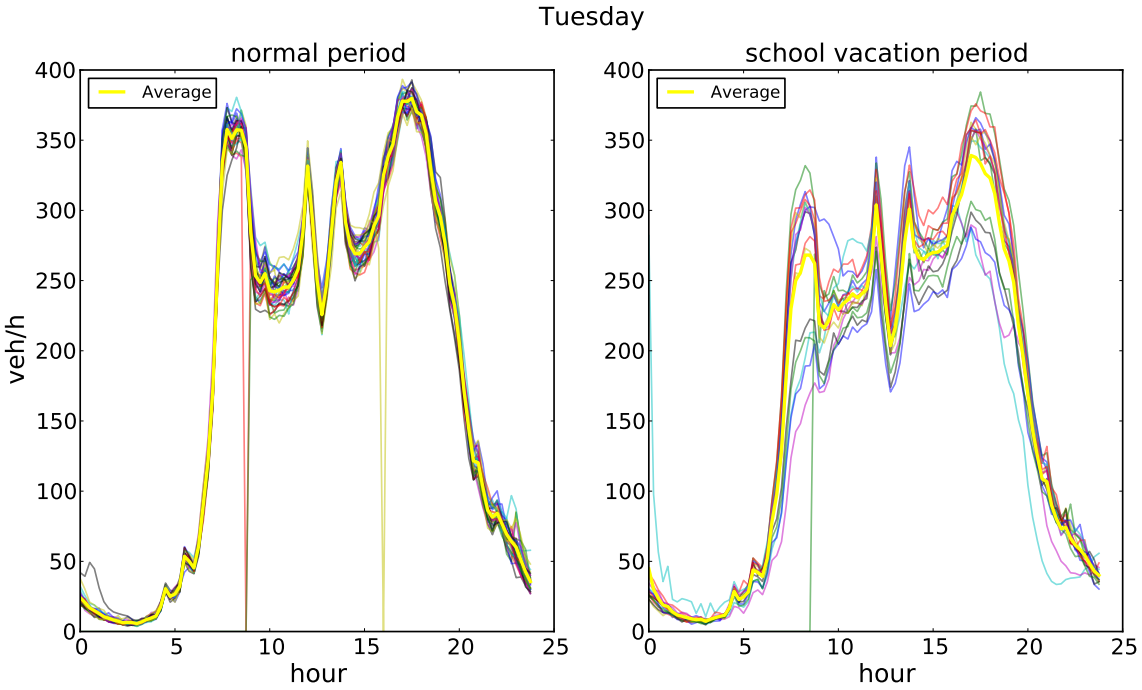


Figure 3.2: Temporal profiles of spatially-averaged traffic flow for all Tuesdays. Each thin line represents a temporal profile of traffic on each Tuesday during normal periods (left) and scholar vacation periods (right), during 2013.9 – 2014.9, and the yellow lines are the average temporal profiles of traffic flow over all Tuesdays in normal periods (left) and school vacation periods (right). Several points reach zero when no data was collected. This rarely happens (less than 5 % of the time), and the data in these cases have not been taken into account for calculating the average temporal profile.

same during normal periods. Figure 3.4 shows that the total daily volume of the traffic measured from all the detectors slightly changes from Monday to Friday, but remains quite stable from month to month during non-vacation periods. However, temporal variations in Figure 3.2 (right) and daily volume difference in Figure 3.4 show that during the vacation period, the temporal variation of traffic flow and total daily traffic volume are less regular than during working periods, and the traffic flow during peak hours are not the same neither. This might lead to DTA simulation errors if we use only the temporal variation of an average working day to represent the temporal variation of traffic demand during vacation periods, due to the uncertainties of traffic demand from day to day.

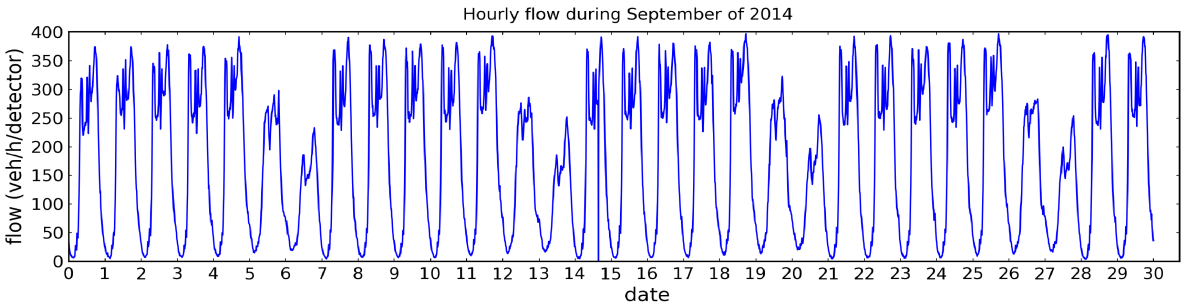


Figure 3.3: Spatially-averaged traffic flow measured by loop detectors, during September 2014. The value reaches zero when no data was collected.

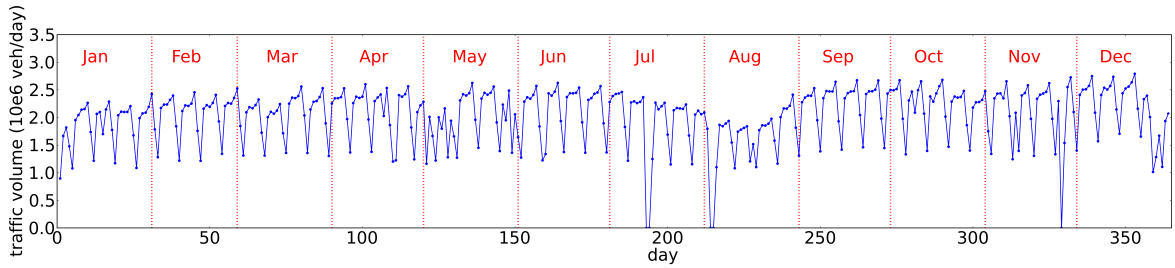


Figure 3.4: Daily traffic volume measured during the year 2014 (in 10^6 veh day^{-1}). The value of each point shows the sum of observational data over all detectors during each day. The value reaches zero when no data was collected.

3.3 Inputs and outputs for dynamic traffic assignment in the agglomeration of Clermont-Ferrand

3.3.1 Network

The road network of the agglomeration of Clermont-Ferrand is modeled as an oriented graph with nodes and links. Two modeled networks are used in this work. They are provided by the City of Clermont-Ferrand (*Ville de Clermont-Ferrand*). The first network is modeled based on the network of the agglomeration in 2003, presented in Figure 3.5. There are 739 nodes and 2194 links. Data are available for capacity, speed limit, length and number of lanes for each link. The free flow travel time of a link is calculated from its length and speed limit. This network is only used for evaluating the LADTA model when compared with the static model VISUM applied to the agglomeration. In LADTA model, the network speed limits and capacities can be time-dependent. In our case study, we assume that they are constant over the simulation period.

The second network is the main modeled network we used in the PhD work. This

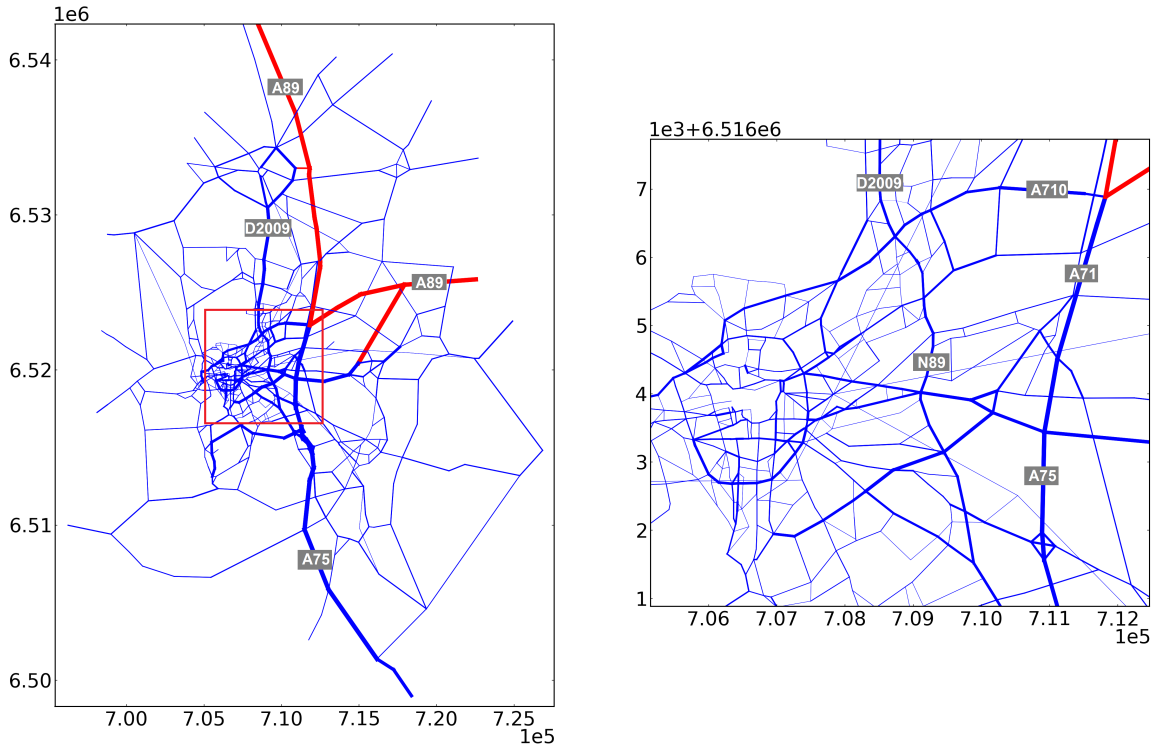


Figure 3.5: Modeled network of the agglomeration of Clermont-Ferrand based on the road network in 2003. The red rectangle delimits the city of Clermont-Ferrand. The width of lines represents link capacity. Blue links represent normal freeways and roads. Red links represent the highway A89 with toll. Other highways (A75, A71, A710), national road (N89) and departmental road (D2009) are free of charge. The projection system is Lambert-93 (RGF93).

network is modeled based on the road network of the agglomeration in 2012. It is presented in Figure 3.6. There are in total 19628 links and 8844 nodes. Detailed information is available for each link, including its beginning and ending nodes, length, capacity, speed limit and number of lanes. The free flow travel time of a link is calculated from its length and speed limit. Except the comparison of simulation results from VISUM and LADTA, all the results and simulations in this PhD work are computed based on the network of 2012 with 19628 modeled links.

In order to compare the loop detector measurements with computed traffic flows from VISUM model or LADTA model, we should link the positions of detectors to the modeled network. To this end, we use the QGIS software to display the detectors, links and nodes in a same map with different layers. Then we noted manually the directions of detectors by their origin nodes and destination nodes. Figure 3.7 illustrates a part of

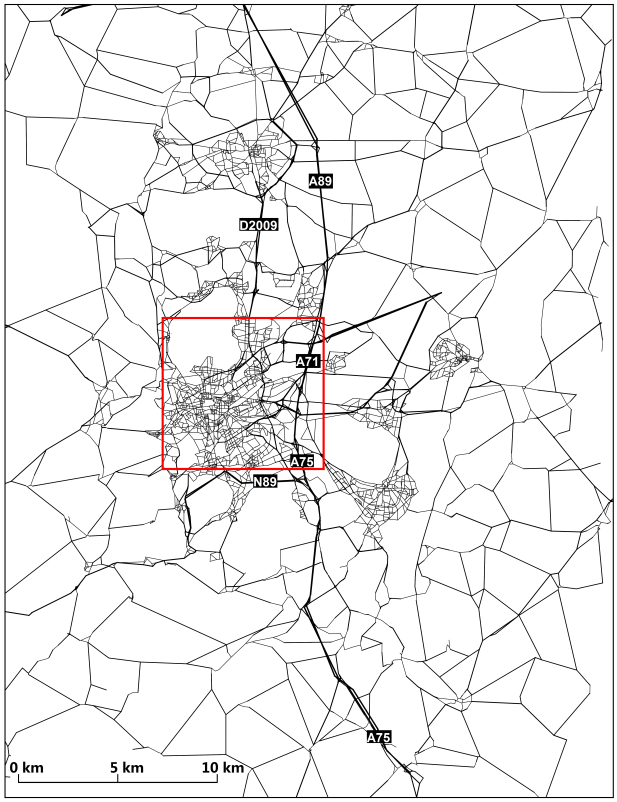


Figure 3.6: Modeled network for the agglomeration of Clermont-Ferrand based on the road network in 2012. $D = 19,628$ links, and 8,844 nodes, with $z = 124$ zones. It has about 200,000 residents in an area of about $300\,\mathrm{km}^2$. The red rectangle delimits the *city of Clermont-Ferrand*. A89, A71 and A75 are three highways passing through the agglomeration and Clermont-Ferrand city. D2009 and N89 are departmental road and national road, connecting the city and the suburbs.

QGIS map with different layers of detectors of the real-world network, and the nodes of the modeled network for the agglomeration of Clermont-Ferrand. The position and direction of a detector can then be associated to a link in the modeled network, by identifying the origin-destination nodes of the link.

3.3.2 Traffic demand

The whole simulation domain is divided into 124 zones. They are both origin zones and destination zones. The traffic demand is represented by an O-D matrix. The latter is obtained by summarizing trip-makers' origin zones and destination zones during a certain period. This information is often modeled based on household survey questionnaires. In our study, a static O-D matrix during the evening peak hour is given for 2008.

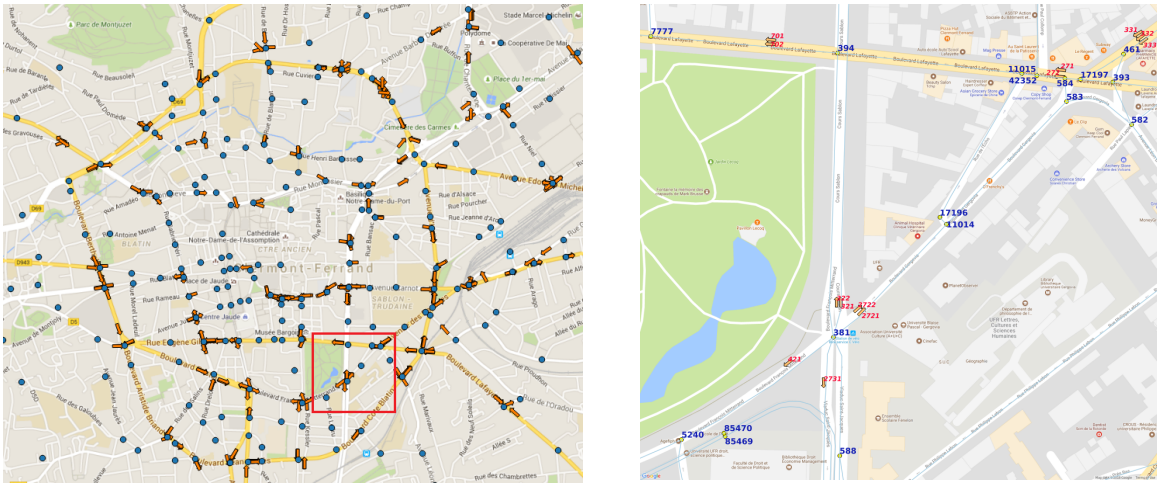


Figure 3.7: Illustration of the layers for determining the detectors origin and destination nodes of the modeled network of Clermont-Ferrand in QGIS software. The figure on the right side is the zoom of the region delimited by the red rectangular in the left-side figure.

It summarizes the averaged vehicle users' trips in the agglomeration of Clermont-Ferrand from 17:00 to 18:00. The unit of traffic demand is veh h^{-1}. The static O-D matrix during an hour (or during a fixed time period) is sufficient for carrying out static traffic assignment study. However, for DTA simulation with LADTA model, time-dependent traffic demand is required. Therefore, it is necessary to convert the static O-D matrix into a dynamic one. To this end, we use information from loop detector measurements in the network. We present here how to convert the static O-D matrix of evening peak hour into a dynamic one with traffic demand continuously varying as a function of time, using an example on the 20^{th} November, for the old network presented in Figure 3.5.

Notation

h_0: the beginning time of the simulation.

h: the time when users enter the road network. It is expressed in local time of Clermont-Ferrand (GMT +1).

$q_{(o,d)}(h)$: the traffic demand (in veh h^{-1}) from the Origin zone o to the Destination d from at instant h. (o,d) is called as an *O-D pair*. $q_{(o,d)}(h)$ is integrable in $\mathbb{R}$. Note that $q_{(d,o)}(h)$ and $q_{(o,d)}(h)$ represent traffic demands of opposite directions. The static O-D matrix gives hourly average traffic flow during the evening peak hour for each O-D pair. We denote it as $q_{(o,d)}^{17-18}$.

$Q_{(o,d)}(h)$: the cumulated traffic demand $Q_{(o,d)}(h)$ from o to d, from h_0 until the instant h. It is measurable on $[h_0, h]$ and in our DTA case study, it is expressed in veh. Its *density* is $q_{(o,d)}(h)$ and $Q_{(o,d)}(h)$ can be obtained by integrating $q_{(o,d)}(h)$ from h_0 until h: $Q_{(o,d)}(h) = \int_{h_0}^{h} q_{(d,o)}(\tilde{h})d\tilde{h}$. $Q_{(o,d)}(h)$ is non-negative and non-decreasing. It is defined as the *temporal profile* of traffic demand from o to d. The objective for converting a static O-D matrix into a dynamic one is to determine the *temporal profile* for each *O-D pair* during the simulation period, based on the static O-D matrix.

N_{det}: total number of loop detectors whose positions have been associated with nodes in the modeled network presented in Figure 3.5. The detectors whose measured values are always zero should be filtered out.

T : total number of time steps during the simulation period. $T = 1440$ for LADTA simulation on the 20^{th} November 2014 with temporal resolution down to every minute for LADTA simulation on the network of 2003 (Figure 3.5).

$o_{k,h}$: average traffic flow measured by the detector k during $[h, h + \Delta h]$ ($o_{k,h}$ in veh h^{-1} $\mathrm{detector}^{-1}$). $[h, h + \Delta h]$ is the time interval in LADTA during which we considered $q_{(o,d)}(h)$ is constant. Δh is the same as the temporal resolution of data observed by detectors: $\Delta h = \frac{1}{60}$ *hour*.

$\bar{o}_h = \frac{1}{N_{det}} \sum_{k=1}^{N_{det}} o_{k,h}$: the spatially-averaged traffic flow measured by loop detectors during $[h, h + \Delta h]$.

$O_{k,(h_1,h)}$: number of cumulated vehicles passing through the detector k from h_1 to h (in veh): $O_{k,(h_1,h)} = \sum_{h=h_1}^{h}(o_{k,h} \times \Delta h)$. Therefore, $O_{k,(h_0,h)}$ denotes the number of vehicles passing through detector k from the beginning time h_0 until the instant h.

$O^{total}_{(h_1,h)} = \sum_{k=1}^{N_{det}} O_{k,(h_1,h)}$: the total observed traffic (in veh) of all detectors on the network, during $[h_1, h]$. In particular, the spatially-averaged measured traffic during the evening peak hour is denoted as $O^{total}_{(17,18)}$. The spatially-averaged measured traffic from h_0 till h is denoted as $O^{total}_{(h_0,h)}$. For a one-day simulation, the total cumulated traffic is denoted as $O^{total}_{(h_0,h_0+24)}$.

$\rho(h) = \frac{O^{total}_{(h_0,h)}}{O^{total}_{(h_0,h_0+24)}}$: the *temporal variation ratio* between (i) the total cumulated traffic on the network from h_0 till h, and (ii) the total cumulated traffic during the whole simulation period of one day. For our case study with temporal resolution of 1 minute, we can also denote this ratio as $\boldsymbol{\rho} = (\rho_t)_{t=1,2,\ldots,T}$ with $T = 1440$. In particular, $\rho_0 = 0.0$ and $\rho_T = 1.0$.

From static O-D matrix to dynamic O-D matrix

In order to convert a static O-D matrix into a dynamic one, we make the following assumptions.

Assumption 3. *The temporal variation of the measured traffic on the network can reflect the temporal variation of the traffic demand in the O-D matrix.*

Assumption 4. *The traffic demands of each O-D pair in the O-D matrix share the same temporal variation.*

Therefore, if we know the total traffic demand during one day for each *O-D pair*, as well as $\boldsymbol{\rho}$ obtained from loop detector measurements, we can build the temporal profile of traffic demand for LADTA simulation during a whole day. Since the static O-D matrix only gives traffic demand during evening peak hour, we use an *evening peak coefficient* α to represent the ratio between (i) the total traffic demand during the whole day and (ii) the traffic demand during evening peak hour on the same day. Besides the loop detector measurements with time resolution of 15 minutes mentioned in Section 3.2, measured traffic flows with temporal resolution down to 1 minute are also available during a week from the 17^{th} to the 23^{th} November 2014. Based on Assumption 3 and Assumption 4, α can be obtained as $\alpha = \frac{O^{total}_{(h_0,h_0+24)}}{O^{total}_{(17,18)}}$ for a one-day simulation of 24 hours. Therefore, all the parameters that we need to convert the static O-D matrix into a dynamic one can be obtained based on loop detector measurements.

Nevertheless, we have only one static O-D matrix but that does not model a specific day. The total number of vehicles measured by the detectors is not the same from one day to another. All the parameters obtained from loop detector measurements depend on the which *day* the simulation is carried out. For example, in our case study on the 20^{th} November 2014, *temporal variation ratio* is denoted as $\rho_{20141120}$ and the *evening peak coefficient* is $\alpha_{20141120}$. However, we are not sure whether the traffic demand given in the static O-D matrix can well represent the traffic demand during evening peak hour on the 20^{th} November 2014. Therefore, we introduce *total demand coefficient* ξ in order to adjust the total demand so that at the end of the simulation when $h = h_0 + 24\,\text{h}$ and the time step $t = T = 1440$ with $\Delta h = 1\,\text{min}$, the total computed traffic volume on the network is the same as the measured total traffic for the specific simulation day. ξ depends only on which *day* the DTA simulation with LADTA is carried out. It can be obtained by calibrating computed traffic flows with measured ones. The calibration of the total demand is presented in the following Section 3.4. Therefore, the time-dependent cumulated traffic demand for the *day* at simulation time h can be represented as

$$Q_{day,(o,d)}(h) = \rho_{day}(h) \times (\xi_{day}\alpha_{day} \times q^{17-18}_{(o,d)}), \tag{3.1}$$

where $\rho_{day}(h)$ and α_{day} are directly obtained from loop detector measurements on the

particular *day* for LADTA simulation. ξ_{day} is obtained from the calibration based on the comparison between the simulation results and observations.

3.4 Calibration and evaluation of dynamic traffic assignment results and comparison with static model

In this section, we compare the computed traffic flows from LADTA model to those computed from the static model VISUM. The results from both models are also compared to the loop detector measurements. The comparison between LADTA and VISUM is carried out on the network of 2003 (see Figure3.5) for the 20^{th} November 2014. Then with the new modeled network of 2012 (see Figure3.6), we evaluate the LADTA model by comparing LADTA outputs during a Tuesday with loop detector measurements of all Tuesdays during a whole year from September 2014 to September 2015.

Before presenting the results, we present here the formulas for calculating comparison criteria. Let N denote the total number of modeled links or the total number of detectors on the network, and let T_{size} denote the total time steps of the simulation period. Let $s_{i,t}$ be the simulation results of traffic assignment model at link i (or at detector i) at time step t, and we denote $r_{i,t}$ as the reference value at the same location i and time step t. The reference sequence $r_{i,t}$ can be either (i) the traffic flows computed by VISUM at link resolution or (ii) the loop detector measurements at detector resolution (lane resolution), depending on different comparison cases. Table 3.1 shows statistical criteria for model evaluation.

3.4.1 Calibration of the total traffic demand

The LADTA simulation is carried out from 03:15 on the 20^{th} November 2014 till 03:15 on the next day. After filtering the detectors whose values are always zero on that day, $N_{det} = 387$. For calibrating the total demand, the used reference sequence is observed traffic flows $(o_{k,h})_{k=1,...,N_{det},h=3.25 \text{ hour},...,(3.25+24.0) \text{ hour}}$, where k represents detector location and h represents the clock time (local time). Figure 3.8 shows spatially-averaged values of $\bar{o}_h = \frac{1}{N_{det}} \sum_{k=1}^{N_{det}} o_{k,h}$ during each time interval $[h, h + \Delta h]$ with $\Delta h = \frac{1}{60}$ (hour). We choose the beginning time of the simulation is 03:15 of the day: $h_0 = 3.25$ hour. In fact, this is the clock time when the traffic on the network is the minimum on the 20^{th} November. For carrying out the DTA simulation with LADTA on the target day,

Table 3.1: Statistical criteria and formula for model evaluation

Criteria	Formula
Total mean value	$\bar{r}_{total} = \frac{1}{N \times T_{size}} (\sum_{i=1}^{N} \sum_{t=1}^{T_{size}} r_{i,t})$ $\bar{s}_{total} = \frac{1}{N \times T_{size}} (\sum_{i=1}^{N} \sum_{t=1}^{T_{size}} s_{i,t})$
Spatially-averaged values	$\bar{r}_t = \frac{1}{N} \sum_{i=1}^{N} r_{i,t}$; $\bar{s}_t = \frac{1}{N} \sum_{i=1}^{N} s_{i,t}$
Temporally-averaged values	$\bar{r}_i = \frac{1}{T_{size}} \sum_{t=1}^{T_{size}} r_{i,t}$; $\bar{s}_i = \frac{1}{T_{size}} \sum_{t=1}^{T_{size}} s_{i,t}$
Spatial mean	$\bar{r}_{spatial} = \frac{1}{N} \sum_{i=1}^{N} \bar{r}_i$; $\bar{s}_{spatial} = \frac{1}{N} \sum_{i=1}^{N} \bar{s}_i$
Temporal mean	$\bar{r}_{temporal} = \frac{1}{T_{size}} \sum_{t=1}^{T_{size}} \bar{r}_t$; $\bar{s}_{temporal} = \frac{1}{T_{size}} \sum_{t=1}^{T_{size}} \bar{s}_t$
Mean bias error	$e_{total} = \frac{1}{N \times T_{size}} \sum_{i=1}^{N} \sum_{t=1}^{T_{size}} (s_{i,t} - r_{i,t})$ $e_{spatial} = \frac{1}{N} \sum_{i=1}^{N} (\bar{s}_i - \bar{r}_i)$ $e_{temporal} = \frac{1}{T_{size}} \sum_{t=1}^{T_{size}} (\bar{s}_t - \bar{r}_t)$
RMSE	$RMSE_{total} = \frac{1}{N \times T_{size}} \sum_{i=1}^{N} \sum_{t=1}^{T_{size}} (s_{i,t} - r_{i,t})^2$ $RMSE_{spatial} = \frac{1}{N} \sum_{i=1}^{N} (\bar{s}_i - \bar{r}_i)^2$ $RMSE_{temporal} = \frac{1}{T_{size}} \sum_{t=1}^{T_{size}} (\bar{s}_t - \bar{r}_t)^2$
NRMSE	$NRMSE_{total} = \frac{RMSE_{total}}{\bar{r}_{total}}$ $NRMSE_{spatial} = \frac{RMSE_{spatial}}{\bar{r}_{spatial}}$ $NRMSE_{temporal} = \frac{RMSE_{temporal}}{\bar{r}_{temporal}}$
Correlation	$R_{total} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_{size}} (r_{i,t} - \bar{r}_{total})(s_{i,t} - \bar{s}_{total})}{\sqrt{\sum_{i=1}^{N} \sum_{t=1}^{T_{size}} (r_{i,t} - \bar{r}_{total})^2} \sqrt{\sum_{i=1}^{N} \sum_{t=1}^{T_{size}} (s_{i,t} - \bar{s}_{total})^2}}$ $R_{spatial} = \frac{\sum_{i=1}^{N} (\bar{r}_i - \bar{r}_{spatial})(\bar{s}_i - \bar{s}_{spatial})}{\sqrt{\sum_{i=1}^{N} (\bar{r}_t - \bar{r}_{spatial})^2} \sqrt{\sum_{i=1}^{N} (\bar{s}_i - \bar{s}_{spatial})^2}}$ $R_{temporal} = \frac{\sum_{t=1}^{T_{size}} (\bar{r}_t - \bar{r}_{temporal})(\bar{s}_t - \bar{s}_{temporal})}{\sqrt{\sum_{t=1}^{T_{size}} (\bar{r}_t - \bar{r}_{temporal})^2} \sqrt{\sum_{t=1}^{T_{size}} (\bar{s}_t - \bar{s}_{temporal})^2}}$

the network can be considered as nearly *empty* before this moment so that we can approximately satisfy the Assumption 1 and Assumption 2.

$O^{total}_{(h_0,h)}$, $O^{total}_{(h_0,h_0+24)}$ and $\rho(h)$ can be obtained from $(o_{k,h})$. The *evening peak coefficient* is obtained by $\alpha_{20141120} = \sum_{h=3.15}^{h=27.15} \bar{o}_h / \sum_{h=17}^{h=18} \bar{o}_h = 11.849$. Note that the traffic simulated in LADTA model is the cumulated traffic at link resolution while the detectors are at lane resolution. For comparing computed results with measurements, let $s_{k,h}$ denote the computed cumulated traffic from h_0 till h at detector k. $s_{k,h}$ for a detector can be obtained by dividing the computed results of LADTA by number of lanes. The total simulated cumulated traffic over all detectors is $S^{total}_{h_0,h_0+24} = \sum_{k=1}^{N_{det}} \sum_{h_0}^{h_0+24} s_{k,h}$. Then we get the error of computed total traffic over all detectors compared with measured total traffic on theses detectors: $e = S^{total}_{h_0,h_0+24} - O^{total}_{(h_0,h_0+24)}$. Then we calibrate the temporal

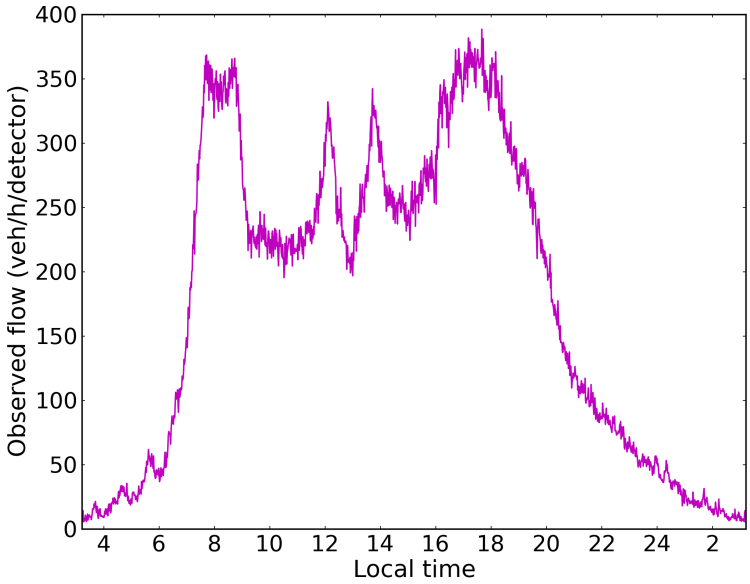


Figure 3.8: Spatially-averaged traffic flow measured by loop detectors ($\bar{o}_h$) on the 20^{th} November 2014. The time resolution is 1 minute.

profile of demand for the dynamic O-D matrix by adjusting ξ mentioned in Equation 3.1 so that $e = 0.0$. The final *optimal* total demand coefficient for the 20^{th} November 2014 is $\xi_{20141120} = 0.704$.

3.4.2 LADTA simulation versus VISUM simulation

Case study of LADTA without taking highway toll into account

Since the total demand of LADTA is calibrated by the $\xi_{20141120}$ on the 20^{th} November 2014, the same coefficient is also multiplied to the static O-D matrix for carrying out VISUM simulation on the same day during evening peak hour (17:00 to 18:00). Both LADTA and VISUM compute traffic flows at link resolution of the agglomeration. Since the outputs of VISUM model are not time-dependent, we only use spatial criteria in Table 3.1 to compare the performance of the two models during evening peak hour. The $\bar{r}_i$ in Table 3.1 is the link-resolution traffic flow computed by VISUM during evening peak. For the LADTA model, we take the temporal average of traffic flow computed during 17:00 to 18:00 as $\bar{s}_i$. The scatter plot of traffic flow computed by LADTA model against those computed by VISUM during evening peak hour is presented in Figure 3.9.

Results in Table 3.2 and Figure 3.9 show that there are less traffic assigned by LADTA model than VISUM. This might due to the fact that the static model averagely assigns traffic on the network without taking into account the temporal variation within

Table 3.2: Comparison of flows computed by VISUM and LADTA model during 17:00 to 18:00 of a working day. The comparison criteria are calculated based on formula presented in Table 3.1.

Criteria	Values
Spatial mean bias error (in veh h^{-1})	-37
Spatial RMSE (in veh h^{-1})	250
Spatial NRMSE (in %)	70.56
Spatial correlation	0.87

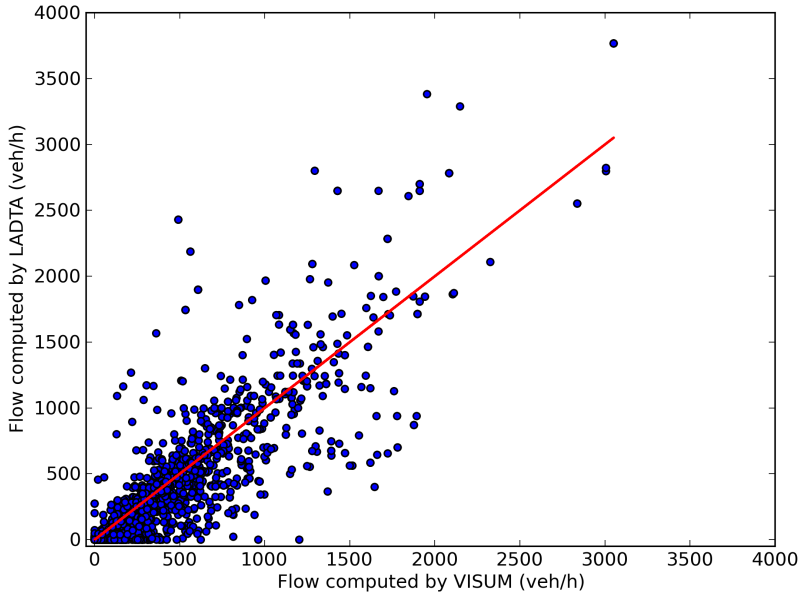


Figure 3.9: Scatter plot of hourly-averaged traffic flow computed by LADTA model against VISUM model, during 17:00 to 18:00 on the 20^{th} November 2014.

an hour. Another reason might be the ignorance of monetary cost in modeling users travel cost in the DTA simulation with LADTA. Figure 3.10 shows the assignment maps of traffic flows computed by VISUM model and LADTA model during evening peak hour from 17:00 to 18:00. We can see that bigger traffic flow is computed by LADTA on the highways in the east side of the Clermont-Ferrand city.

The traffic flows computed by both models are also compared with loop detector measurements. Table 3.3 shows statistical scores of VISUM and LADTA when compared with loop detector measurements. In this comparison, the reference sequence $r_{i,t}$ in Table 3.1 is composed by measured values, while $s_{i,t}$ are traffic flows computed by

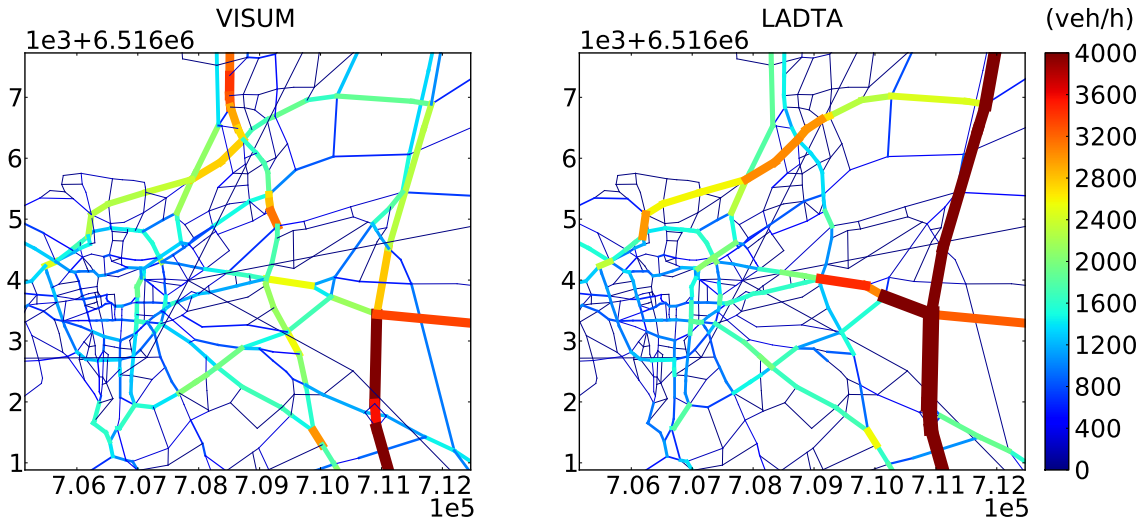


Figure 3.10: Maps of computed traffic flow on the network of the city of Clermont-Ferrand. In these two maps, both colors and widths of lines represent values of computed traffic flows. For the links that share the same nodes but are oriented in opposite directions, the sum of the computed flows are calculated and displayed in these two maps.

VISUM or LADTA, at location i at time step t.

Table 3.3: Comparison of flows computed by VISUM and LADTA model from 17:00 to 18:00 of a working day at link resolution. The comparison criteria are calculated based on formulas presented in Table 3.1.

Comparison	VISUM v.s. measurements	LADTA v.s measurements
Spatial mean bias error (in veh h^{-1})	22	-14
Spatial RMSE (in veh h^{-1})	290	302
Spatial NRMSE (in %)	79.07	82.54
Spatial correlation	0.36	0.38

Results in Table 3.3 show that the traffic flows computed by both models during the evening peak hour have big spatial errors when compared with loop detector measurements. This might due to the following reasons. (i) The O-D matrix is modeled for 2008 but the comparison is carried out based on observed data in 2014. In addition, the O-D matrix only contains trips between working places and homes. Trips of other purposes were not taken into account. (ii) The modeled network is based on the real-world network in 2003 while the observations are measured in 2014. Traffic on the network may change due to the infrastructural changes from 2003 to 2014. For the

agglomeration of Clermont-Ferrand, the tramway system has officially opened in 2006. This might influence the structure of the network and spatial distribution of traffic on the network. (iii) Clermont-Ferrand is a city in the center part of France. There must be road transit traffic through the city but the O-D matrix does not take it into account. (iv) Assumptions in both models that travelers choose their routes by minimizing travel costs may not be true in reality. In VISUM model, the generalized travel costs are composed of travel time and toll prices. In the case study here for DTA simulation with LADTA model presented here, only travel time is taken into consideration as the travel cost to be minimized. There are also other criteria for travelers to choose their route from one zone to another in reality. (v) The uncertainty in modeling the network. In fact, the Origin-Destination zones (O-D zones) are represented by centroids of zones in the modeled network. The departing and arriving traffic comes from or go to centroids of O-D zones. The centroids are connected to the nodes of the modeled network through *connectors* with unlimited capacity and negligible free-flow travel time. In the modeled network, not all the nodes are connected with *connectors* to centroids. This might generate an unbalance of computed traffic flows around centroids of zones.

However, we can see that the LADTA model has approximately the same performance as the VISUM model when compared with loop detector measurements during evening peak hour. Moreover, LADTA can give detailed temporal variation of traffic flow within an hour, during a whole-day simulation or even longer. For example, Figure 3.11 shows the spatially-averaged traffic flow computed by LADTA during the whole day on the 20^{th} November 2014, from 03:15 to 03:15 of the next day. Figure 3.11 shows that LADTA model can well represent the temporal variation of spatially-averaged traffic flow. This is because we gave a temporal variation to inputs in O-D matrix that reflects the temporal variation of measured traffic flows on the network.

Case study of LADTA with highway toll

Then we added monetary cost in modeling generalized travel cost in order to influence the users route choice decision. Let $\mathcal{A}_{toll}$ denote the set of links that model the tolled highway in the network. Then at link resolution, the generalized travel cost is modeled as a linear combination of the travel time and toll cost

$$C_a(h) = T_a(h) + v_t \times p_a(h) \times l_a, \tag{3.2a}$$

$$p_a(h) = \begin{cases} p_{unit}(h) & a \in \mathcal{A}_{toll} \\ 0.0 & \text{if not} \end{cases} \tag{3.2b}$$

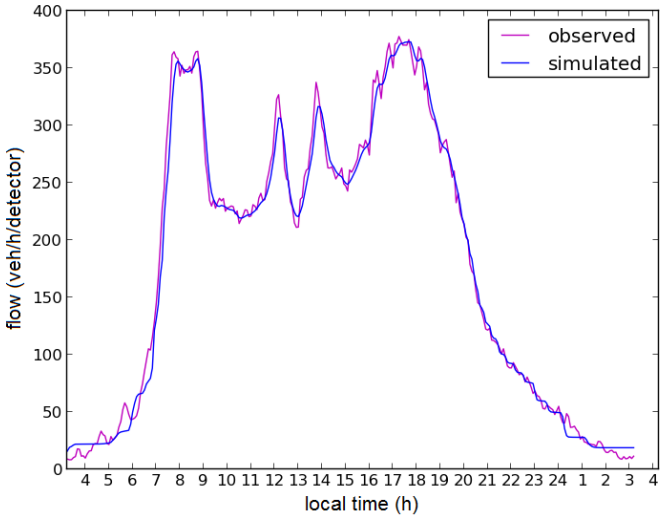


Figure 3.11: Temporal variation of spatially-averaged traffic flow computed by LADTA and measured by loop detectors, on the 20^{th} November 2014. The simulation is carried out from 03:15 to 03:15 of the next day.

where v_t is the *value of time* of travelers in €/h, $T_a(h)$ is the travel time at time h, $p_a(h)$ is the monetary price that network users should pay on link a and $p_{unit}(h)$ is the price of the tolled highway (in €/km). We set $p_{unit}(h) = 0.5$ €/km for our case study. In general, the *value of time* varies for different categories of users and for difference purpose of traveling. Here we assume that there is only one category of users on the network and v_t is the same for all users. In the agglomeration of Clermont-Ferrand, A89 is a north-south tolled highway. It connects Riom in the northern part of the agglomeration and Clermont-Ferrand city. There is also a free departmental road (D2009) connecting these two zones. Since we do not have the observations of these two roads, we used the results of VISUM model to approximate the ratio between the traffic volume passing through these two roads during evening peak hour: $\beta_{D2009/A89}$. Then we fixed other parameters such as ξ, α for the demand, and link parameters including p_{unit}, capacity and speed limit. We varied v_t for carrying out DTA simulation with LADTA in order to find an optimal v_t^* resulting in the same ratio $\beta_{D2009/A89}$. Therefore, with a unit toll price of $p_{unit} = 0.5$ €/km, we found $v_t^* = 66.9$. Figure 3.12 shows the DTA simulation results compared with traffic assigned by VISUM for the whole agglomeration of Clermont-Ferrand, with or without adding the highway toll price into the routing choice criteria.

Results in Figure 3.12 show that the adding of monetary cost in the route choice modeling can significantly change the spatial distribution of the traffic on the roads

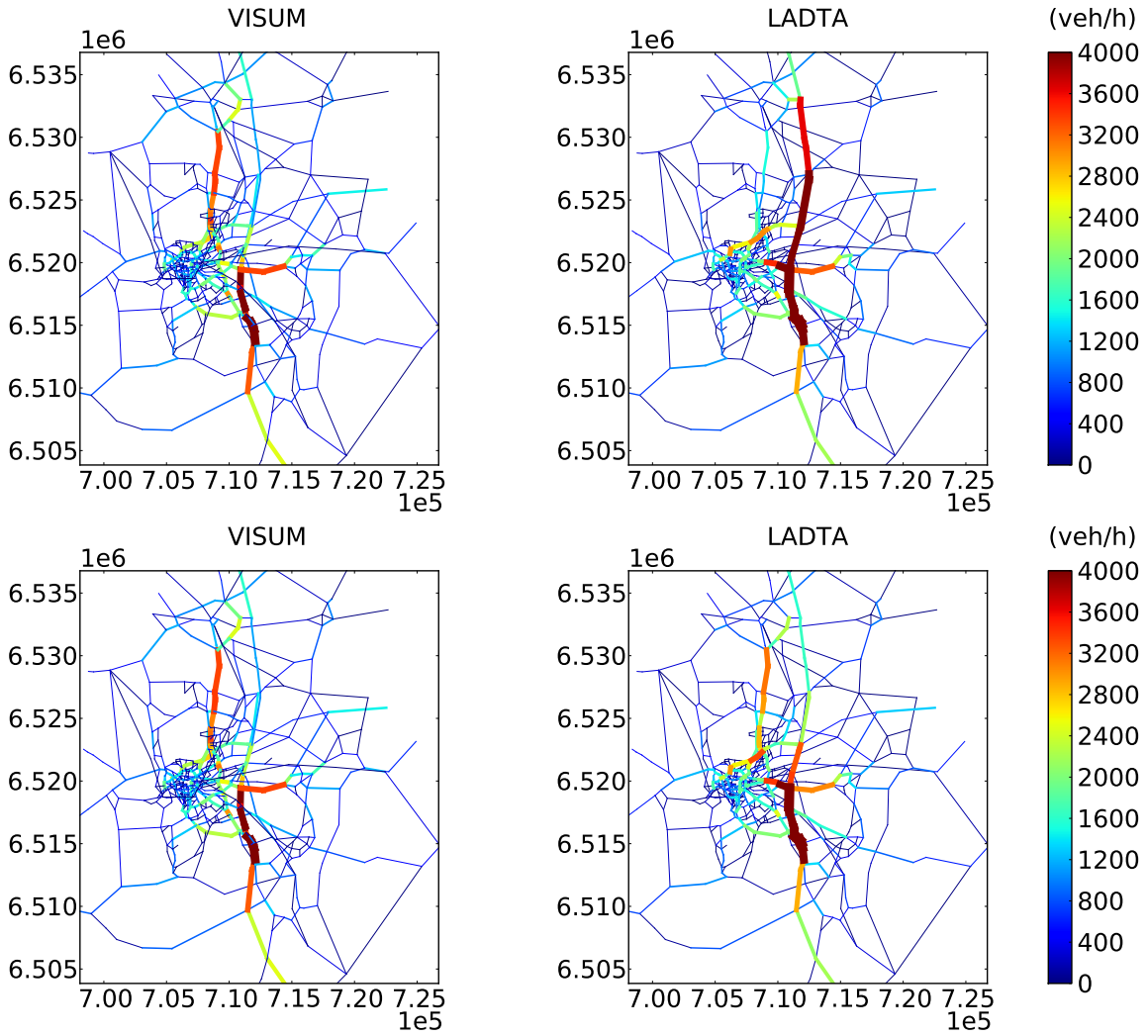


Figure 3.12: Maps of computed traffic flow on the network of the agglomeration of Clermont-Ferrand from 17:00 to 18:00 on the 20^{th} November. In these two maps, both colors and widths of lines represent values of computed traffic flows. For the links that share the same nodes but are oriented in opposite directions, the sum of the computed flows are calculated and displayed. The highway toll is taken into account in the static simulation and the assignment results are shown by the two figures on the left side). For LADTA simulation results, the upper figure on the right side presents the assignment map in the case study where the highway toll is not taken into account in the route choice criteria. The lower figure on the right side presents the results where the highway toll is taken into account in modeling users travel cost.

around the tolled highway in the network. In addition, we calculated the difference between the traffic flows computed by the two models: $e_i^{17-18} = \bar{s}_i^{17-18} - \bar{r}_i^{17-18}$, where

$\bar{s}_i^{17-18}$ and $\bar{r}_i^{17-18}$ are respectively the temporally-averaged traffic flow during 17:00 to 18:00 computed by LADTA and VISUM. The spatial distributions of (e_i^{17-18}) for the two case studies with or without taking into account the toll in route choice modeling are presented in Figure 3.13. Results show that adding toll into route choice criteria can influence not only the traffic flows computed on the highway in question, but also can affect traffic flows on links of other parts of the network.

Table 3.4 presents the statistical results for the comparison between LADTA model and VISUM model, during evening peak hour on the 20^{th} November 2014. The computed flows from LADTA model are also compared with loop detector measurements during the same period. If we compare the results in Table 3.4 with those in Table 3.2, we can see that the difference of computed flows between VISUM and LADTA is decreased after adding toll cost into users route choice criteria.

Table 3.4: Comparison of flows computed by LADTA with (i) traffic flows computed by VISUM at link resolution (ii) loop detector measurements, in the case where the highway toll has been taken into account in modeling users route choice, for the simulation from 17:00 to 18:00 on the 20^{th} November 2014. The comparison criteria are calculated based on formula presented in Table 3.1.

Comparison	LADTA v.s VISUM at link resolution	LADTA v.s measurments
Spatial mean bias error (in veh h^{-1})	-34	2
Spatial RMSE (in veh h^{-1})	199	308
Spatial NRMSE (in %)	55.92	84.31
Spatial correlation	0.92	0.38

If we compare the flows computed by LADTA with loop detector measurements, the statistical scores do not change significantly when compared with the results in Table 3.4 and those in Table 3.3. This might due to the lack of observations on links where the computed flows are sensitive to the addition of toll costs. However, we adjusted the value of v_t in order to get the same $\beta_{D2009/A89}$ in LADTA and in VISUM simulation. The final $v_t = 67$ is much higher than the average *value of time* (17.5 €/h) in France for professional traveling purpose [Roquigny, 2013]. In addition, even with big value of $v_t = 67$, the resulting spatial variation cannot significantly decrease the big spatial errors of LADTA when comparing its results with loop detector measurements. This suggests that the addition of toll cost in the model of generalized cost is not sufficient for improving the performance of LADTA in predicting spatial distribution of traffic on the network, when comparing computed traffic flows with measured ones. Therefore,

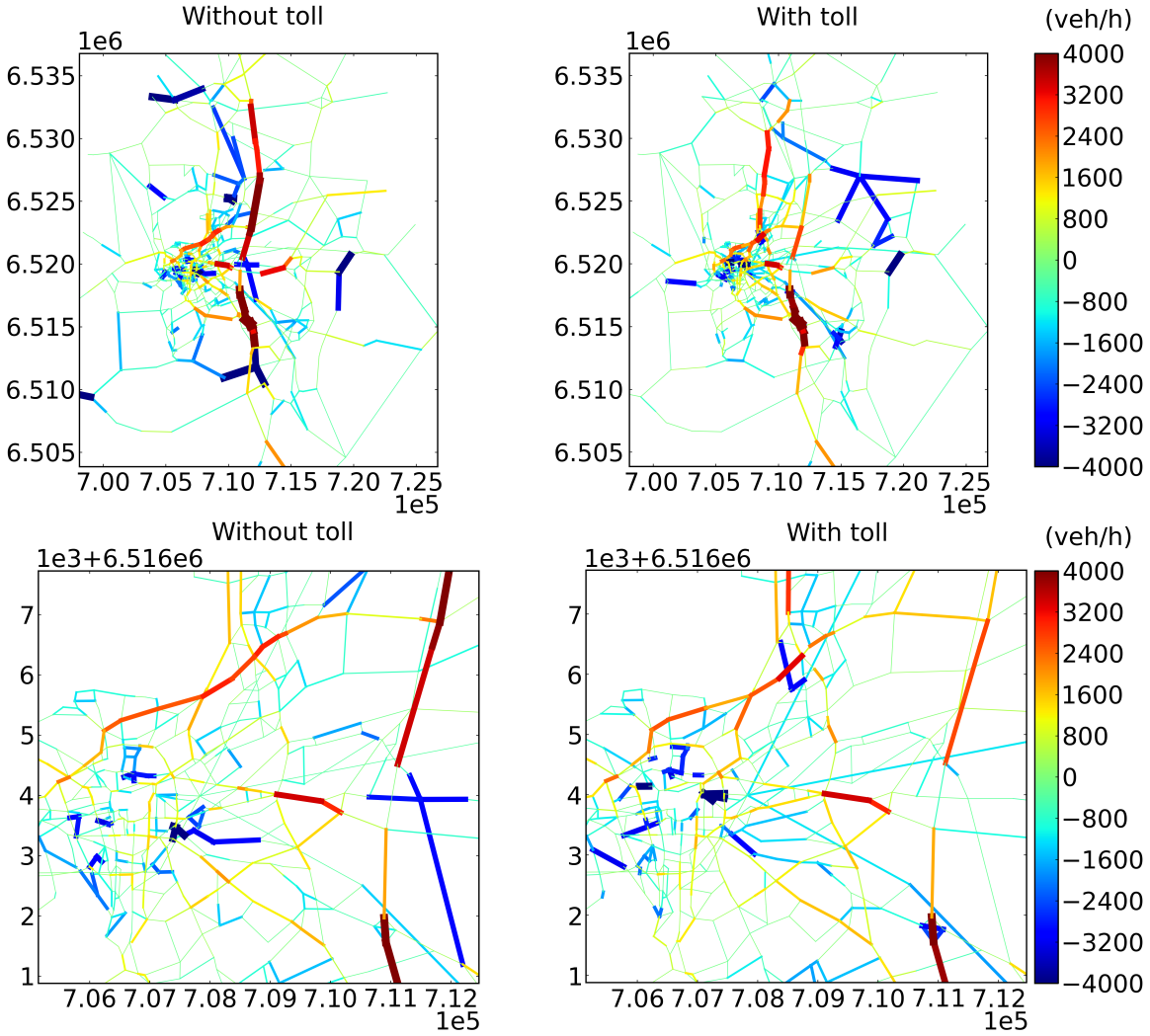


Figure 3.13: Maps of differences between hourly-averaged loaded flows computed by LADTA and flows computed by VISUM: LADTA results - VISUM results. We calculate the difference between the loaded traffic by two models during evening peak hour on the 20^{th} November 2014. In these two maps, both colors and widths of lines represent the differences. For the links that share the same nodes but are oriented in opposite directions, the sum of traffic flows on both directions is firstly calculated, and then the difference between the two resulting sums are calculated and displayed in the maps.

we carried out LADTA simulation on the latest modeled network with more detailed links, presented in Figure 3.6. The simulation results are presented in the following Section 3.4.3.

3.4.3 LADTA simulation results on the network of the agglomeration of Clermont-Ferrand of the year 2012

In this simulation on the larger network, loop detector measurements data from September 2013 to September 2015 are used for carrying out the LADTA simulation and for model evaluation. The time resolution of measured traffic flows is every 15 minutes: $\Delta h = 0.25$ hour. A one-day DTA simulation with LADTA is carried out for a working Tuesday during non-vacation period. The static O-D matrix used here is the same one that used in Section 3.4.2. Then we converted the static O-D matrix into a dynamic one by applying a temporal variation to each O-D pair of the static O-D matrix, and calibrated the total demand using the same methods mentioned in Section 3.4.1. The temporal profile of the working Tuesday is obtained by averaging temporal profiles of all the Tuesdays from September 2013 to July 2014. The temporal variation of the working Tuesday is represented by the yellow line in Figure 3.2. The *evening peak coefficient* for the average temporal profile of Tuesdays is $\alpha_{Tuesday} = 11.70$. The *total demand coefficient* is $\xi_{Tuesday} = 0.9823$ for calibrating the total traffic demand compared with loop detector measurements of all Tuesdays from September 2013 to July 2014. Then LADTA simulation is carried out, and the results are compared with loop detector measurements of all Tuesdays from September 2014 to July 2015 in order to evaluate the performance of LADTA simulation. The toll of the highway A89 has not been taken into account.

Temporal variation of spatially-averaged traffic flows

Figure 3.14 shows the temporal variation of spatially-averaged traffic flows computed by LADTA and measured by loop detectors, during a working Tuesday. For plotting the measurement curve in Figure 3.14, we firstly calculate the spatially-averaged measured traffic flows of each Tuesday from September 2014 to July 2015 to obtain a temporal profile of each Tuesday. Then we calculate an average temporal profile of all these Tuesdays to obtain the measurement curve and compare it with LADTA simulation results. This result shows that LADTA can well simulate the temporal variation of spatially-averaged traffic flows.

Note that the comparison results presented in this subsection is not obtained by the same way as those presented in Section 3.4.2. In fact, in the simulation in Section 3.4.2, the input temporal variation for converting the static O-D matrix into the dynamic one is obtained from the loop detector measurements on the 20^{th} November 2014. The comparison results presented in Figure 3.11 are also obtained from the same day. It

is normal (and reassuring) to find that the two temporal profiles fit with each other. However, in this subsection, the input temporal profile is learned from observed data of all Tuesdays from September 2013 to June 2014, while the comparison is carried out for the spatially-averaged traffic flow of all the Tuesdays in another period from September 2014 to June 2015. The fit between the two temporal profiles refers to a good prediction of temporal variation of on-road traffic flows during working Tuesdays.

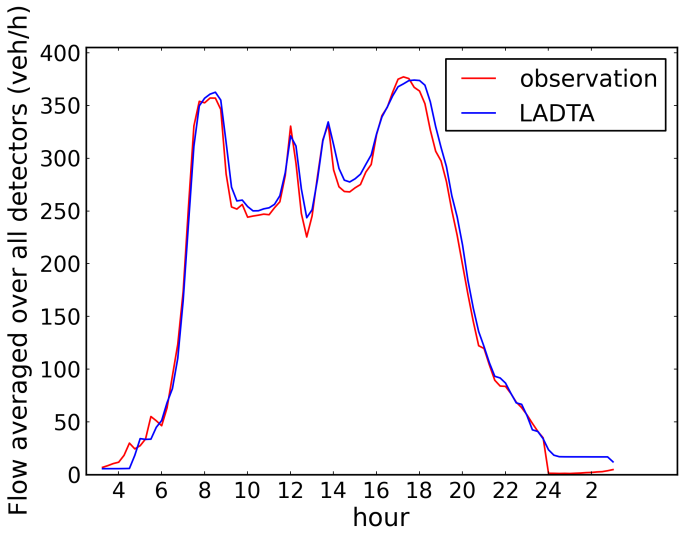


Figure 3.14: Temporal variation of spatially-averaged flow on a normal Tuesday during non vacation periods from the 1^{st} September 2014 to the 30^{th} June 2015.

Spatial distribution of traffic flows and congestions

Figure 3.15 and Figure 3.16 show the traffic assignment and the congestion distribution of Clermont-Ferrand from 07:00 a.m. to 08:30 a.m., during each time interval of 15 minutes. In this section, a link is considered to be congested when the average travel time on it (denoted as t) is 50 % more than its free flow travel time (denoted as t_0). The results show that the links with higher capacities have been assigned more traffic volume, and the congestions appear firstly and more frequently on crossroads. These fit well to the reality of urban traffic situation. The spatial distributions at different time periods show that the transition to the morning peak is almost immediate, and some links become congested within a quarter of an hour. The dynamic traffic assignment model gives more detailed information of traffic temporal evolutions than the static assignment model. Moreover, if we compare the computed traffic flow results with congestion results from 07:45 to 08:30, we can see that even though it seems like

the traffic flows do not change significantly during this period (Figure 3.15), the number of congested links increase from 07:45 to 08:30 (Figure 3.16). In fact, in LADTA, the interaction between link travel time and link loaded traffic is modeled by *bottleneck model* [Vickrey, 1969]. The travel time of a link equals to the free flow travel time if the entering traffic flow does not exceed the link capacity, and the travel time equals to the sum of free flow travel time and queuing time if the assigned traffic flow is higher than link capacity. Therefore, the computed traffic flows on congested links do not increase any more if they exceed the link capacity, and the congestion is represented by the increase of travel time.

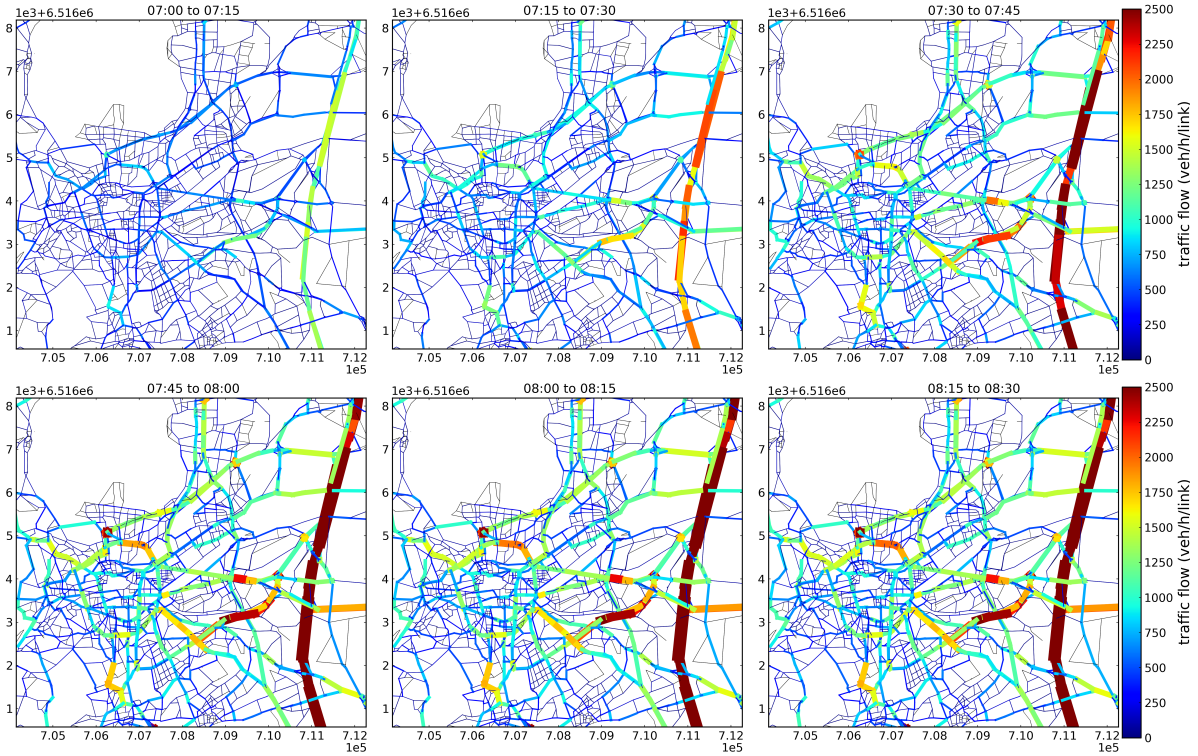


Figure 3.15: The traffic assignment results of LADTA from 07:00 a.m. to 08:30 a.m. for a working Tuesday. Black lines represent unused links. Bold lines with various colors represent traffic flow computed by LADTA model. For the links that share the same nodes but are oriented in opposite directions, the sum of the computed flows are calculated and displayed. The line width is proportional to the traffic flow on the link. The projection system of the map is Lambert-93 (RGF93).

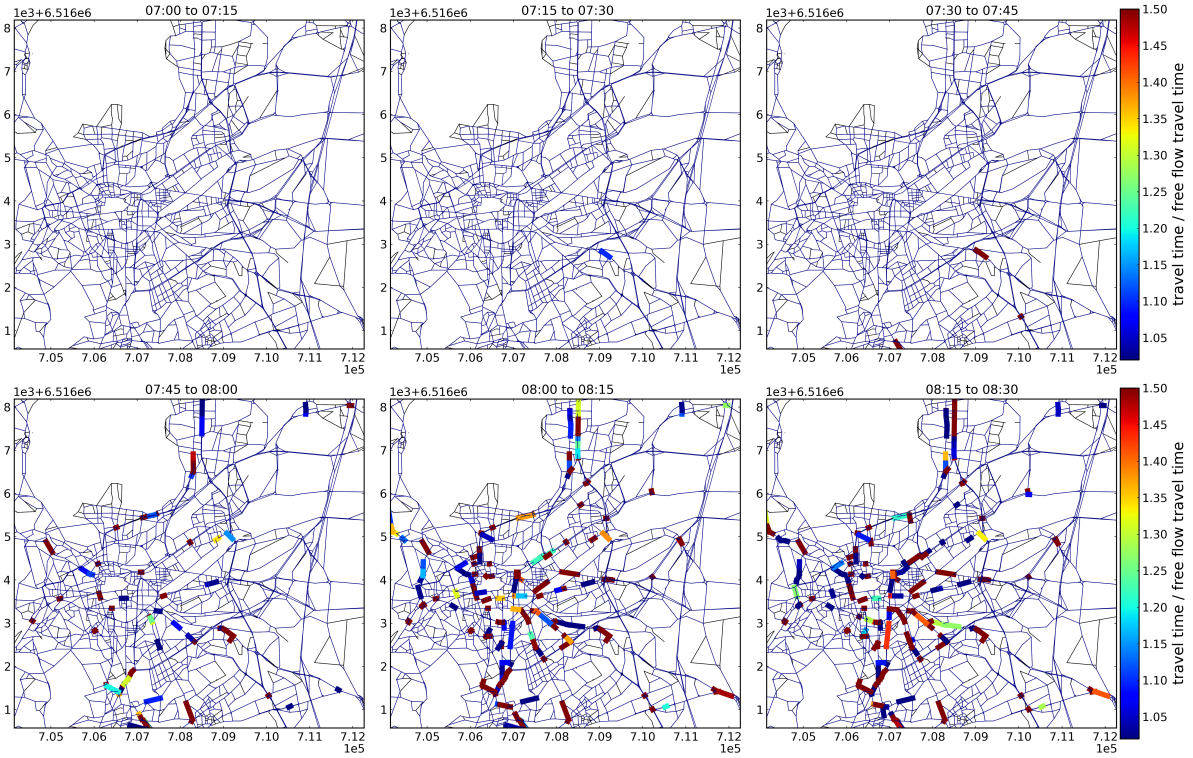


Figure 3.16: The spatial distribution of congestion from 07:00 a.m. to 08:30 a.m. for a working Tuesday. Black lines represent unused links. Thin blue lines represent links whose travel time equals to its free flow travel time. Bold lines with various colors represent links with travel time larger than its free flow travel time. For the links that share the same nodes but are oriented in opposite directions, the bigger travel time / free-flow travel time ratio is displayed in the maps. The projection system of the map is Lambert-93 (RGF93).

Statistical scores

We use the formula in Table 3.1 to evaluate the performance of LADTA model for the simulation on the larger network. The comparison is carried out for all the working Tuesdays (during non-vacation period) from September 2014 to July 2015. There are 36 Tuesdays during the simulation period. For each day, there are 96 time steps. Therefore, the total time step in Table 3.1 is $T_{size} = 36 \times 96 = 3456$. The total number of detectors associated with the larger network is 469 so that $N = 469$ in Table 3.1. For the loop detector measurements, the comparison sequence is composed by measured traffic flows during all Tuesdays from September 2014 to July 2015: $r_{i,t}$ with $i = 1, \ldots, N$ and $t = 1, \ldots, T_{size}$. Since we only carried out LADTA simulation for an average working Tuesday, we assumed that for all the Tuesdays, the computed traffic flows are the

same so that the dimension of simulation sequence $(s_{i,t})_{i=1,...,N;t=1,...,T_{size}}$ equals to that of $(r_{i,t})_{i=1,...,N,t=1,...,T_{size}}$. Therefore, the *total*, *temporal* and *spatial* criteria defined in Table 3.1 can be computed. The statistical scores are presented in Table 3.5.

Table 3.5: Comparison of traffic flow results between LADTA and measurements at all times and detectors during all Tuesdays from September 2014 to June 2015. The scores are calculated based on the criteria presented in Table 3.1.

Score	Total	Temporal	Spatial
Mean bias error (veh h^{-1} detector^{-1})	6	6	5
RMSE (veh h^{-1} detector^{-1})	149	12	112
Mean value of measurements (veh h^{-1} detector^{-1})	183	183	183
NRMSE (%)	81.4	6.6	61.2
Correlation	0.73	0.99	0.57

Results in Table 3.5 show that LADTA model can well predict the temporal evolution of the daily traffic. The model performed well with high correlation, low bias and low RMSE. However, the model still has limitations to predict the spatial distribution of the traffic in the network. This may be mainly due to the following three reasons. (i) The method to estimate the simulated results for each detector by dividing link volume with lane number brings *representativeness errors* (see definition and example in Section 1.5.3). This simplification ignores the fact that left side lanes and right side lanes might have different traffic flows. (ii) It is not always true in reality that all the travelers choose their route by minimizing their generalized travel cost. Different users may use several different criteria for their route choice strategies. (iii) The ignorance of monetary cost in modeling generalized travel cost. In this case, the toll of the highway A89 does not influence the traffic assignment of LADTA. This leads to a heavier traffic assignment on the highway since it has higher capacity and lower free flow travel time. This might then lead to an underestimation of traffic flow on the road in parallel of the highway, and an eventual underestimation of traffic flow passing through the city. (iv) The temporal variation of traffic demands might be different for an O-D pair from an industrial zone to a residence zone, and for a pair from a residence zone to an industrial zone. (v) The O-D matrix only contains travel purpose of users with passenger cars. Therefore, the computed traffic flows and travel times are only for this class of users. While the traffic flow measured by detectors contain all kinds of vehicles including light utility vehicles, heavy duty vehicles, buses, etc. Comparing computed traffic flows of passenger cars to measurements of various kinds of vehicles can also bring representativeness errors.

3.5 Qualitative sensitivity analysis

In this section, we are interested in how the inputs of LADTA can influence the computed traffic flows on the network of the agglomeration of Clermont-Ferrand. The simulation results in Section 3.4.3 are considered as references. A local sensitivity analysis is carried out with *one at a time* method. We vary each of two main inputs: the total demand during a day on the network, and the speed limits of the links (i.e. the free flow travel times on links). We take two outputs as indicators to evaluate the sensitivity of LADTA simulation: the total vehicle travel time (in veh·h) and the proportion of congested links on the whole network. For the latter indicator, a link is considered as congested using the same criterion in Section 3.4.3: travel time 50 % higher than the free flow travel time. The congestion proportion is then calculated as the ratio between the total length of congested roads on the network, and the total length of the whole network.

3.5.1 Total demand of traffic volume in the network

We vary the total demand volume from −50 % to 100 %, with the same temporal profile of a working Tuesday as in Section 3.4.3. The speed limits and capacities remain unchanged. Then the vehicle travel time and the proportion of congested roads in the whole network are calculated for each link during each 15 minutes. Figure 3.17 shows the spatial distribution of these two indicators during the period from 7:45 a.m. to 8:00 a.m., with a total traffic demand volume of −50 %, 0 %, and 50 %, compared with the referenced simulation case in Section 3.4.3. It can be observed that the traffic volume and the link travel time are strongly influenced by the evolution of the total demand volume entered to the LADTA model.

In order to analyze the sensitivity in detail, the total values of these indicators of the whole network during 24 h are calculated for each case of different total demand volume. The evaluations of these two global indicators in function of different total traffic demand volume are presented in Figure 3.18 and Figure 3.19.

From Figure 3.18 and Figure 3.19, we can see that the total vehicle travel time and the congestion rate of the network are very sensitive to the total demand volume, especially when congestion phenomenon appears. Figure 3.18 (left) and Figure 3.19 (left) show that the total vehicle travel time and the congestion proportion of the network becomes about 10 times higher than the reference case, when the total demand is doubled. In Figure 3.18 (right) and Figure 3.19 (right), below the reference total demand volume, the increase of vehicle travel time and the congestion rate is almost linear with

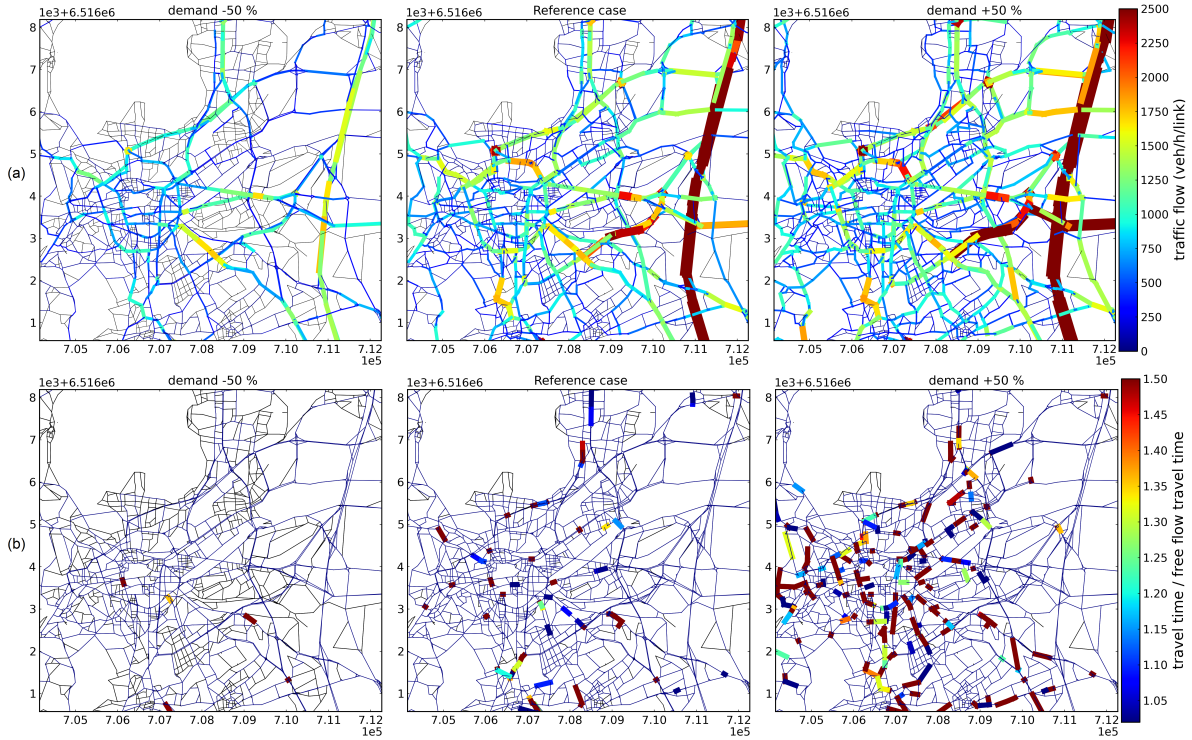


Figure 3.17: The spatial distribution of (a) traffic flow (in veh h^{-1}) and (b) congestion ratio during 07:45 a.m. to 08:00 a.m. of a working Tuesday, with change of total demand volume of $-50\,\%$, reference case, and $50\,\%$. Black lines represent unused links in LADTA model. The projection system of all the maps is Lambert-93 (RGF93). In (a), the line width is proportional to the traffic flow on the link. For the links that share the same nodes but are oriented in opposite directions, the sum of the computed flows are calculated and displayed.In (b), thin blue lines represent links whose travel time equals to its free flow travel time. Bold lines with various colors represent links with travel time larger than its free flow travel time. For the links that share the same nodes but are oriented in opposite directions, the bigger travel time / free-flow travel time ratio is displayed in the maps.

the increase of the demand. After that, the network becomes more and more congested and the growth of travel time of vehicles becomes increasingly fast. Furthermore, the road traffic during the evening peak period is more congested than during the morning peak. Travelers suffer a longer travel time during evening peak with congestions than during morning peak hour. In fact, the network before evening peak hour is no longer empty and the traffic appeared before the evening peak can affect the dynamic assignment traffic results. This is a phenomenon that we can only observe in DTA simulations. In fact, the existing traffic before evening peak hours can be considered as an

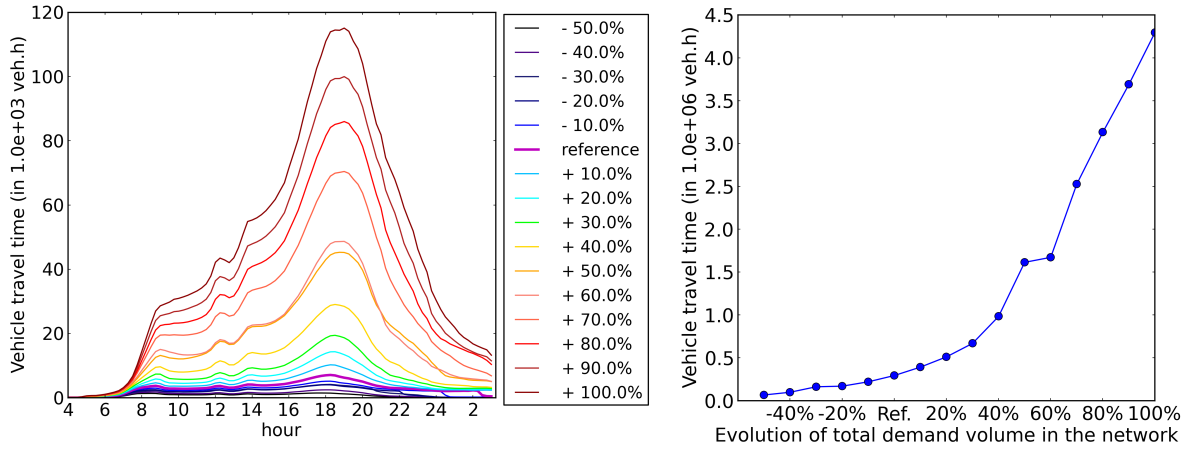


Figure 3.18: Sensitivity of vehicle travel time to the total demand volume. Left: evolution of vehicle travel time (in 10^3 veh·h) during the day at all time steps of 15 minutes. Right: evolution of the total vehicle travel time (in 10^6 veh·h) in the whole network.

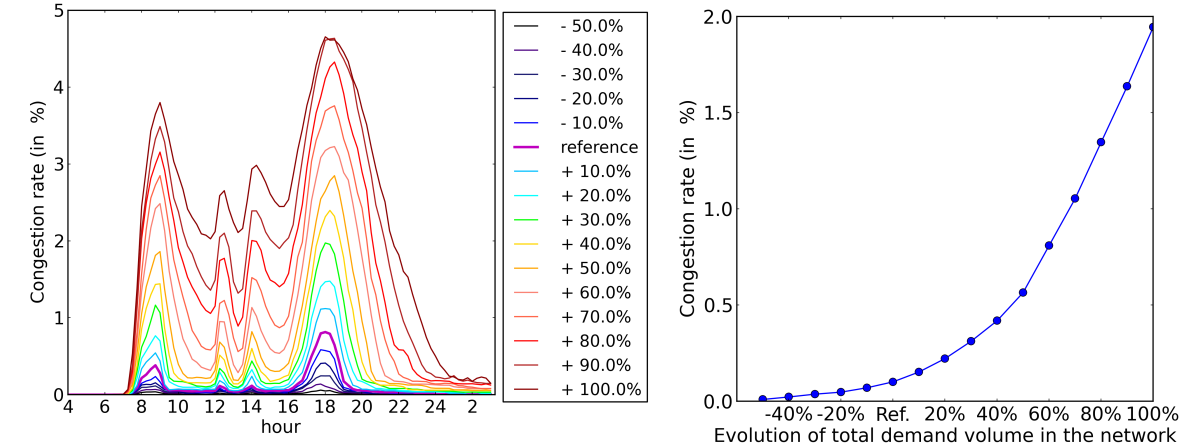


Figure 3.19: Sensitivity of the congestion rate (in %) to the total demand volume. Left: evolution during the day all time steps of 15 minutes. Right: evolution of the proportion of congested links in the whole network.

decrease of the link capacity. The increasing traffic demand during evening peak hour needs longer time to pass through the network than during the morning peak where the network is nearly empty. This phenomenon cannot be simulated by static traffic assignment models even with separate time periods, since the static models do not take into account the influence of the existing traffic on the network.

3.5.2 Speed limits

In this sub-section, we decrease the speed limit of all the roads on the network by 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 % and 50 %. LADTA model assumes that the users free flow speed is the same as speed limit of the network. The decrease of speed limits leads to an increase of free flow travel time. It is found that the total vehicle travel time increases almost linearly with the decrease of speed limits, but it is less sensitive to the speed limitation than to the total demand, as shown in Figure 3.20. This might due to the fact that the speed limits have been changed homogeneously for all roads on the network.

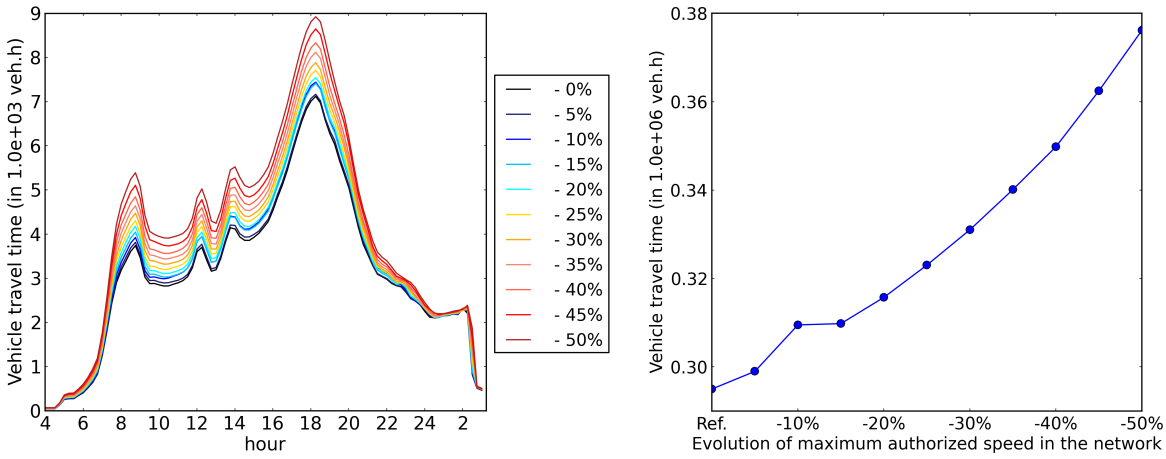


Figure 3.20: Sensitivity of vehicle travel time to the decrease of speed limits of links on the network. Left: evolution during the day (in 10^3 veh·h). Right: evolution of the total vehicle travel time in the whole network (in 10^6 veh·h) of the proportion of congested links in the whole network

Conclusions

The LADTA model was applied to the road network of the agglomeration of Clermont-Ferrand. A reference simulation with static VISUM model was also carried out. The results computed by LADTA and VISUM were compared at link resolution. Results show that the correlation between the traffic flow computed by the two models was 0.87. When we took the toll into account for modeling users travel cost, the correlation between traffic flows computed by the two models was 0.92. The traffic flows computed by two models were also compared with loop detector measurements. Results show that LADTA and VISUM model have similar performance for simulating traffic flow during

evening peak hour from 17:00 to 18:00. Both models have limitations for predicting spatial distribution of traffic flow on the network when compared with traffic flow measurements. These may due to (i) the uncertainties in O-D matrix, (ii) assumptions for route choice criteria, (iii) errors in the modeled network and (iv) the conversion from link-level flows to detector-level flows by dividing the link-resolution flow by the number of lanes. Then we added the monetary cost for modeling travel cost of users. With the help of VISUM simulation results, we adjusted the ratio of assigned traffic flow between the toll highway and the departmental road in parallel in LADTA simulation. The setting price of toll highway is 0.5 €/km and users *value of time* is 66.9 €/h^{-1}. This brings a decrease of NRMSE from 71 % to 56 % when comparing the traffic flows computed by LADTA model with those computed by VISUM model at link resolution (see Table 3.2 and Table 3.4). The correlation between the traffic flows computed by VISUM and LADTA increases from 0.87 to 0.92. Nevertheless, it does not improve significantly statistical scores of LADTA simulation when compared with loop detector measurements.

A DTA simulation with LADTA model for a working Tuesday was also carried out for the same agglomeration with the latest modeled network of 2012. The computed flows were compared with loop detectors measurements of all the working Tuesdays from September 2014 to July 2015. In this case study, users choose their route by minimizing only travel time, without taking into monetary cost into account. Results show that the spatial error still exists, but it is less than in the case with the old network of 2003. Results also show that LADTA model can well predict temporal variation of spatially-averaged traffic flow in the whole agglomeration, even for a long period of more than 30 days with temporal resolution of 15 minutes. The temporal NRMSE is about 6.6 % and the correlation is 0.99 when comparing the computed and measured spatially-averaged traffic flow for all time intervals. The DTA simulation with LADTA gives more detailed temporal evaluation of traffic flows and congestions on the network. For example, the LADTA results during morning peak hour from 07:00 to 08:30 show that the traffic flows on the network vary significantly within a quarter of an hour. In addition, the congestion on the network is represented by the increase of vehicle travel time. The use of DTA model can better reflect the temporal variation of traffic flow and congestions with a high time resolution than static traffic assignment models.

A qualitative sensitivity analysis was then carried out to analyze the sensitivity of computed flows with respect to total traffic demand and link speed limits. Results show that the total vehicle travel time (in veh·h) is very sensitive to the input traffic demand, especially during evening peak hour due to the existing traffic on the network before

the peak. The total vehicle travel time increases almost linearly with the decrease of speed limits on the network.

Even though the LADTA simulation is able to compute time-dependent traffic flow and travel time on the network, it is computationally costly for the large size network of more than 19,000 links. For a whole-day simulation, it requires several hours. This might already be quite a computational burden in an operational context for transportation planning, model calibration or qualitative sensitivity analysis. However, if we want to analyze the quantitative global sensitivity of computed flows with respect to all the inputs and then carry out uncertainty quantification using ensemble prediction, large numbers of model evaluations are required. A more efficient model is then required. This is the reason why we build a metamodel in the next chapter.

Chapter 4

Metamodeling for a dynamic traffic assignment model at metropolitan scale

---

Summary

Dynamic traffic assignment (DTA) allows for better modeling of traffic flows than static assignments because all the features of the demand and supply are time-varying. However, their high computational costs are one of the limitations for their operational application to large-size networks. In this chapter, we suggest a metamodeling method for a DTA model applied to a given network, i.e., a reasonable close approximation to the original model, but with a very low computational cost. It consists of a dimensionality reduction and a statistical emulation. A reduced base for the model outputs is firstly built with the help of principal component analysis (PCA). The outputs of the original model are projected onto this reduced subspace. Then, the relations between the projection coefficients and the inputs of the original model are reproduced by a statistical emulator, based on radial basic functions (RBF). In our work, we have applied this metamodeling method to a DTA model named LADTA. A case study is carried out for the agglomeration of Clermont-Ferrand (France), whose network has more than 19,000 links and 8,800 nodes. In comparison with traffic flow measurements, the performance of the metamodel is similar to that of the complete model during a one-month simulation, but the computational time for a one-month simulation decreases from 2 days on 110 cores to less than 1 minute on one core. The spatio-temporal correlation between the metamodel and the traffic flow measurements during one month is about 0.7. Note that the metamodeling treats the traffic assignment (TA) model as black box, and builds reasonable relationships between the TA inputs and outputs for a given road network. The independence of the method to the model type gives us new insights in TA problems on large scale

networks during long-term periods, and opens the way to further studies such as global sensitivity analysis and uncertainty quantification.

Résumé

L'affectation dynamique du trafic permet une meilleure modélisation des débits que les modèles statiques. En effet, les modèles dynamiques prennent en compte la dimension temporelle des entrées et des sorties: sur la demande, les paramètres du réseau, les débits et temps de parcours simulés, etc. Cependant, l'affectation dynamique du trafic reste coûteuse, notamment quant à l'application sur un réseau urbain ou bien métropolitain. Dans ce chapitre, nous présentons une méthode de méta-modélisation pour un modèle d'affectation dynamique configuré pour un réseau routier métropolitain. Les sorties du méta-modèle sont très proches de celles simulées par le modèle complet, mais avec un coût de calcul très faible. Plus précisément, il y a trois étapes principales dans la méthode de méta-modélisation présenté dans ce chapitre: (i) un échantillonnage par hypercube latin avec un grand nombre de simulations du modèle original pour obtenir les points d'apprentissage, (ii) une réduction de dimension à l'aide d'une analyse en composantes principales des sorties résultant des points d'apprentissage, et (iii) une émulation statistique entre les points d'apprentissage et les coefficients de projection correspondant à chaque composante principale. Dans notre travail, nous avons appliqué cette méthode de méta-modélisation au modèle dynamique LADTA. Un cas d'étude de cas est réalisé pour l'agglomération de Clermont-Ferrand (France), dont le réseau routier est modélisé par plus de 19 000 arcs et 8 800 nœuds. Les débits simulés par le modèle original et par le méta-modèle sont comparés avec les observations obtenues par les comptages sur le réseau, à la résolution de la voie pour tous les intervalles de 15 minutes pendant un mois complet. Les scores montrent que la performance du méta-modèle est similaire à celle du modèle complet, mais le temps de calcul pour une simulation d'un mois passe de 2 jours sur 110 cœurs à moins de 1 minute sur un seul cœur. La corrélation spatio-temporelle entre le méta-modèle et les débits mesurés pendant un mois est d'environ 0,7. Notons que le méta-modèle traite le modèle d'affectation de trafic comme une boîte noire et établit des relations raisonnables entre les entrées et sorties du modèle pour un réseau routier modélisé. L'indépendance de la méthode par rapport au type de modèle nous donne de nouvelles perspectives sur les problèmes d'affectation du trafic sur les grands réseaux pendant une longue période. L'efficacité du méta-modèle nous permet ensuite de réaliser d'autres études où il est demandé de réaliser un

grand nombre d'évaluations du modèle, comme par exemple l'analyse de sensibilité globale et la quantification d'incertitude.

Contents

4.1 Introduction

This chapter focuses on proposing a metamodeling method to reduce significantly the computational cost of dynamic traffic assignment (DTA) problems at metropolitan scale. In the field of transportation planning and modeling, static assignment models are being replaced by dynamic assignment models. The time dimension in dynamic assignment allows for more accurate modeling of various phenomena, including for instance congestion and time-dependent route choice. However, more detailed models lead to complex high-dimensional simulations, which lead to computational requirements.

Traffic assignment (TA) aims at determining traffic flows on a transportation network. TA is often modeled as a demand/supply equilibrium problem. The demand comprises two parts: (i) an Origin-Destination matrix (O-D matrix) and (ii) a model of users route choice behavior. The supply is modeled oriented graph with nodes and links. The interaction between supply and demand is expressed by a flowing model: to each link is associated a flowing function that maps a flow to a travel time [Wardrop, 1952; Beckmann et al., 1956; Smith, 1979; Sheffi, 1985], with constrains of link constraints. In static TA models, the O-D matrix expresses peak hour flows. Flowing functions are volume/delay functions [Bureau, 1964]. Users are assumed to choose routes of least travel cost. The equilibrium follows Wardrop's user equilibrium (UE) principle. The TA problem is often formulated as a classical optimization problem. By definition, static TA models ignore the time dimension: traffic conditions do not vary during the modeled time interval. This approximation is adequate for transportation planning studies. It is less adequate when fine grain time variations of traffic conditions have to be modeled.

Dynamic traffic assignment (DTA) models have been developed to address this issue [Chiu et al., 2011; Peeta and Ziliaskopoulos, 2001]. The demand in a DTA model is expressed as a dynamic O-D matrix. For each departure instant, a dynamic O-D matrix expresses the instantaneous flow between each O-D pair. The underlying flowing model allows to map the (time-varying) traffic load on each route to its (time-varying) travel time [Chen and Hsueh, 1998; Ran and Boyce, 1996]. Some DTA models use a microscopic traffic simulator [Mahut and Florian, 2010] as flowing model detailing traveler behavior which is quite computer-intensive. Some use a simpler macroscopic approach. For instance, a first-in-first-out (FIFO) point queue model [Kuwahara and Akamatsu, 1993] associated with each link essentially replaces the volume/delay functions used in static TA. A point queue model allows to map the time-varying load on a link to its time-varying travel time. The time-varying route travel times are then computed by composing the individual link travel time functions along each route.

The DTA problem can be considered as a dynamic optimization problem (DOP).

The mathematical programming, optimal control and variational inequality (VI) formulations are three main approaches for expressing the DTA problem [Friesz et al., 1989, 1992, 1993]. Ran et al. [1996] extend the VI formulations in [Friesz et al., 1993], proposing a link-based VI model and a *dynamic user optimum* (DUO) route choice condition. Based on the latter condition, a flowing model can then be associated to each link. With link-based VI formulations and point queue model of each link, it is possible to carry out within-day DTA for a large scale network. However, numerical simulations of DTA models and DOP raise challenges. They are far more computationally demanding than static ones, and may exceed in many cases the capability of current hardware.

One of the solutions to deal with computer intensive models is using metamodel (or surrogate models, emulators) in order to approximate complex models. A metamodel is an approximation of the input/output function that is implied by the underlying simulation model [Kleijnen, 2009]. The built metamodel preserves the main response of the original model to input, but with much lower computational time. Once a class of surrogate models has been chosen according to some prior knowledge, they can be built upon available samples [Roustant et al., 2012]. There are three main steps for constructing a metamodel based on an original model: (i) sampling selection or experimental design in order to build a set of training points, and the corresponding training values; (ii) construction of the metamodel in order to find an surrogate model based on the training points and training values; (iii) evaluation and optimization of the metamodel. There are various types of metamodeling methods for computer experiments include linear regression (polynomial regression) [Madu and Kuei, 1994; Kleijnen, 2009], splines, neural networks, Gaussian emulators [Rasmussen and Williams, 2006; Sacks et al., 1989], Kriging surrogates [Kleijnen, 2009], and radial basis function (RBF) [Dyn et al., 1986; Broomhead and Lowe, 1988], etc. For low-dimensional problems, linear regression may be straightforward and efficient. However, it has less robustness when applied to high-dimensional models, compared with Kriging and RBF methods [Jin et al., 2001].

Kriging method is a spatial interpolation method [Roustant et al., 2012]. The Kriging method was firstly proposed by D.G. Krige [Krige, 1951] and mathematically formulated in Matheron [1963] for geostatistics. It is applied to interpolate spatial data for constructing surrogate models [Rasmussen, 2004; Kleijnen, 2009; Stein, 2012]. Numerical Kriging approximation packages are developed such as DACE Matlab toolbox [Lophaven et al., 2002] and DiceKriging [Roustant et al., 2012]. Let us denote a model $y = f(\boldsymbol{p})$ with $\boldsymbol{p} \in \mathbb{R}^K$ and f is computationally costly. Basically, the Kriging metamodel aims at predicting the model's outputs on a target (unknown) input point $\boldsymbol{p}^{(0)} \in \mathbb{R}^K$, based on a linear combination of known training values $(\{f(\boldsymbol{p}^{(1)}), \ldots, f(\boldsymbol{p}^{(i)}), \ldots, f(\boldsymbol{p}^{(n)})\}$),

computed at training points (denoted as $\{\boldsymbol{p}^{(1)}, \ldots, \boldsymbol{p}^{(i)}, \ldots, \boldsymbol{p}^{(n)}\}$ with $n \in \mathbb{N}$). The objective of Kriging interpolation is to find $\hat{f}$ so that for any unknown $\boldsymbol{p}^{(0)} \in \mathbb{R}^K$, $\hat{f}(\boldsymbol{p}^{(0)}) = \sum_1^n \omega_i(\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(1)}, \ldots, \boldsymbol{p}^{(n)}) f(\boldsymbol{p}^{(i)}) \simeq f(\boldsymbol{p}^{(0)})$. ω_i is the weight affected to $f(\boldsymbol{p}^{(i)})$ and it depends only on the relative position of $\boldsymbol{p}^{(0)}$ with respect to training points. They are chosen to minimize the variance of the prediction error. The training process of the Kriging interpolation consists in determining the covariance of the Gaussian process [Roustant et al., 2012; Mallet et al., 2013]. The interpolation process is unbiased so that the predictions at all training points coincide with the corresponding training values. Thorough reviews regarding Kriging method and Gaussian process can be found in Rasmussen and Williams [2006]; Kleijnen [2009]; Stein [2012]; Roustant et al. [2012].

The Kriging interpolation is unbiased. However, it is very time-consuming especially for interpolating large-dimension training points because it involves a K-dimension optimization process [Jin et al., 2001; Mallet et al., 2013]. Another similar interpolation method is with radial basis functions (RBFs) [Dyn et al., 1986; Broomhead and Lowe, 1988]. The RBF-based interpolation of the training points is more computationally efficient than the Kriging method. RBFs were developed initially in Duchon [1977]; Oeuvray and Bierlaire [2009] and their applications are in various fields in sciences and mathematics [Buhmann, 2003]. The concept of RBF-based interpolation is that the influence of a training point $\boldsymbol{p}^{(i)}$ to the approximation of $f(\boldsymbol{p})$ depends only on the distance between $\boldsymbol{p}^{(i)}$ and the target point $\boldsymbol{p}^{(0)}$: the bigger the distance, the less $\boldsymbol{p}^{(i)}$ should be influential. Interpolation based on RBFs can be briefly summarized as finding $\hat{f}$ so that for any unknown $\boldsymbol{p}^{(0)} \in \mathbb{R}^K$, $\hat{f}(\boldsymbol{p}^{(0)}) = \sum_1^n \lambda_i \phi(d(\boldsymbol{p}^{(0)}, \boldsymbol{p}^{(i)})) \simeq f(\boldsymbol{p}^{(0)})$, where ϕ is the chosen RBF, λ_i is the weight and $d(\cdot, \cdot)$ defines the distance between $\boldsymbol{p}^{(0)}$ and $\boldsymbol{p}^{(i)}$. λ_i is computed so that $\hat{f}(\boldsymbol{p}^{(i)}) = f(\boldsymbol{p}^{(i)})$ for $i = 1, \ldots, n$. The computation of the weights is in fact to solve a linear system. The *distance* here can be defined by users in order to take into account the sensitivity of the model f with respect to each input $\boldsymbol{p}^{(i)}$. Without specification, it is often defined as Euclidean norm. The choice of different radial basis functions depends on applications. A method employing multiquadric radial functions is proposed in Hardy [1990] and it can be used for a large variety of problems. Most of these applications are one, two or three dimensional [Oeuvray and Bierlaire, 2009].

The three main methods mentioned above are widely used in various domains, and the choice among different methods depends on the application. In recent years, metamodeling methods have also been used in traffic modeling and management, in order to solve the computational burden in algorithm, sensitivity analysis, safety management, etc. For example, surrogate of the microscopic Aimsun model is built with Kriging method for global sensitivity analysis (GSA) and model calibration [Ciuffo et al., 2013;

Azevedo et al., 2015]. Some other studies of metamodeling applied to microscopic traffic simulators are reviewed in Song et al. [2017] for different applications. Macroscopic traffic assignment problems at urban scale are also addressed by embedding surrogates of microscopic traffic simulators in an optimization algorithm. For example, Osorio and Bierlaire [2013] propose an efficient simulation-based optimization framework embedding a polynomial-based metamodel that integrates information from a microscopic simulator with an analytical queuing network model. Based on this framework, different algorithms are proposed to carry out (i) static traffic assignment [Osorio and Chong, 2015; Osorio and Nanduri, 2015] and (ii) dynamic traffic assignment and calibration [Chong and Osorio, 2017; Zhang et al., 2017], at urban scale under tight computational budget. Metamodels based on microscopic simulators are also used in other optimization problems in transportation management such as highway charges and dynamic pricing [Chen et al., 2014; He et al., 2016]. Besides the initiatives in accelerating the optimization algorithm, there are also applications that make use of metamodeling methods in order to build surrogate for dynamic network loading (DNL) models, mapping the set of path departure rates to the set of path travel time. For example, Song et al. [2017] propose a framework of building metamodel of dynamic flowing models based on Kriging method. Their case studies on a SiouxFall network (76 links and 24 nodes) show the potential of surrogates of DNL models for speeding up DUE-based model.

However, the mentioned surrogate-based DTA models might be adequate for transportation planning, management and optimization problems in an operational context. Applying these models to large-scale DUE-based model for a whole city is still computationally costly, in cases where large numbers of model evaluations are required at metropolitan scale. These cases include carrying out global sensitivity analysis of model inputs, long-term DTA simulations to be combined with traffic emission models for air quality simulations, and probabilistic simulations based on Monte Carlo simulation of the deterministic DTA model. In these cases, metamodels that replace DTA models applied to large-scale networks are needed. Based on the objectives of this PhD work, the on-road traffic emission estimation, GSA analysis and uncertainty quantification should be carried out at metropolitan scale. It is necessary to reduce the computational cost of the DTA model applied to a large-scale network of a whole city. The most important outputs computed from the DTA model for the emission estimation is time-depending street-level traffic flow and travel speed. This motivates us to build a statistical metamodel that directly emulates the relation between the DTA model inputs and the computed street-level traffic flow/speed. The final metamodel is efficient because it ignores the computationally expensive dynamic optimization problem lying

behind the DTA model. This metamodeling approach has not yet been applied in traffic modeling, but the idea of applying the metamodeling method directly to a given large scale model has proved to be successful in various fields such as tsunami simulation [Sarri et al., 2012], as well as in air quality simulations with a global aerosol model [Lee et al., 2011], an urban model [Mallet et al., 2013] and an atmospheric dispersion model for radionuclides [Girard et al., 2016]. Kriging, RBF interpolations or alternatives are limited by the dimension of the input space [Saltelli et al., 2008; Mallet et al., 2013]. For large dimension models (which is the case for LADTA model applied to a large-scale network), a dimension reduction should be firstly carried out. The outputs are projected onto a reduced subspace, which can be obtained based on principal component analysis of the training values. Then the final metamodel combines an emulator of the reduced model and a reconstruction of the model outputs.

In the mentioned literature, the metamodels are built for either the users route choice behavior models to propose surrogates for the objective functions of SO problems, or for building approximations for flowing model. The DTA simulation still needs to solve expensive DOP for large-scale network: searching for minimum-cost route at each moment for each destination. They might be adequate for management and optimization problems for transportation planners, however, in cases where large numbers of model evaluations are required such as global sensitivity analysis of model inputs, or long-term DTA simulations to be combined with traffic emission models in air quality simulations, more efficient alternatives of DTA models are needed.

The work presented here applies the metamodeling method in a different way: to directly build a DTA metamodel at metropolitan scale. A DTA model applied in a given network is treated as a black-box. The relationship between the inputs and outputs of the given DTA model is approximated by a metamodel, which is significantly more efficient than the original one since it ignores the computationally expensive DOP lying behind the DTA model. The idea of applying the metamodeling method directory to a given large scale model has proved to be successful in various fields such as tsunami simulation [Sarri et al., 2012] and air quality simulations with a global aerosol model, an urban model and an atmospheric dispersion model for radionuclides [Lee et al., 2011; Mallet et al., 2013; Girard et al., 2016].

The proposed metamodeling approach comprises (i) dimension reduction of the inputs and outputs of the original modal, and (ii) statistical emulation. The DTA outputs considered here are traffic flows at link level. Other outputs such as travel time or travel speed at link-level can also be considered. Briefly, this chapter deals with the following challenges.

Metamodeling framework for TA models at metropolitan scale. We propose a model-free metamodeling framework in order to build surrogates of TA models applied to large-scale network. We take a DTA model called LADTA [Leurent, 2003; Aguiléra and Leurent, 2009] to illustrate the process, but the metamodeling method can be used to any other macroscopic TA model, whatever static or dynamic.

Efficiency of the built metamodel. A case study is carried out for the agglomeration of Clermont-Ferrand (France) during a whole month. The evaluation time of an *atomic* simulation of LADTA in Clermont-Ferrand decreases from 2 hours to about 0.02 seconds with the help of the metamodel. A one-month DTA simulation for this large-scale network decreases from 48 hours on 110 cores to less than 1 minute on only one core.

Good approximation performance when compared with original model as well as with observations. In our case study, the area is divided into 124 origin/destination zones, and the road network comprises more than 19,000 links and 8,800 nodes. The accuracy of the metamodel is close to the original DTA model. The performance of the metamodel is acceptable when comparing the traffic flows computed by the metamodel to the flows computed by the original model, and to traffic flows observed by loop detector measurements during one month, as shown in Figure 4.1. More detailed results are presented in Section 4.4.

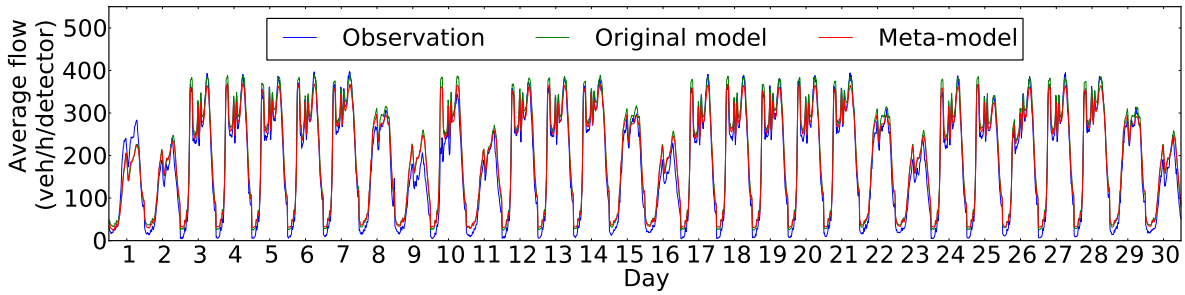


Figure 4.1: Temporal variation of the spatially-averaged flows (in veh h^{-1}) of observation (blue), flows computed by the original model (green) and flows computed by metamodel (red), with a time step of 15 min during the November 2014 for the agglomeration of Clermont-Ferrand.

The ability to carry out TA traffic simulations at a large scale and during long time intervals opens up new fields of application. Long-period traffic assignment results at street level can be coupled with emission models to simulation on-road emissions for long-term (several years) air quality simulations at metropolitan scale. Moreover, global sensitivity analysis, where large numbers of model evaluations are required ($\simeq 10^4$

times of the input dimension), can be carried out for TA models in order to identify the most influential inputs and help us understand the interaction between them. In addition, once a metamodel with low computational cost exists, a stochastic model can easily be built upon, so that inputs are taken from probability distributions to better represent their day to day variability in traffic conditions. Eventually, detailed uncertainty analyses may be carried out thanks to this new approach. Detailed further applications of the resulting metamodel are presented in Section 4.4.5.

The remainder of this chapter is organized as follows. In Section 4.2, metamodeling framework for TA models at metropolitan scale is proposed. In Section 4.3, a brief introduction of LADTA model is proposed, as well as the corresponding DTA simulations in Clermont-Ferrand. An *atomic* simulation with LADTA is designed to reduce input dimension yet preserve the main features of the complete LADTA simulations. This *atomic* simulation is called LCF (for *LADTA applied in Clermont-Ferrand*). The input of LCF is a vector of uncertain factors concerning time-varying traffic demands and network parameters. The evaluation of LCF model is presented by comparing model results with traffic flow observations during a reference day. In Section 4.4, the metamodeling approach is then applied to the LCF model, including the dimension reduction of the outputs and a statistical emulation. The CPU time per run for LCF model decreases from about 2 hours to 0.022 seconds on one core. The results of the metamodel are compared with the results computed by the original model, and also with loop detector traffic measurements on the network of Clermont-Ferrand during a one-month simulation for November 2014. Results show that the metamodel can well preserve input/output responses of the original LCF model. At last, we conclude with a discussion on future extensions of the metamodeling applications in global sensitivity analysis, uncertainty quantifications and on-road traffic emission estimations.

4.2 Methodology

Metamodeling consists in replacing a model, usually complex and computationally intensive, with a surrogate model (or emulator) whose computational cost is very low. A classical approach relies on Gaussian processes which provide a clear theoretical background [Rasmussen and Williams, 2006; Sacks et al., 1989] and quality implementations [Roustant et al., 2012]. The main idea is to sample the inputs of the model and compute the corresponding outputs. Under given assumptions, a Gaussian process emulator can be computed by Kriging of the samples [Kleijnen, 2009]. This emulator can then be seen as an interpolator between the samples. Other emulators can be computed with a

faster interpolator based on radial basis functions (RBF) [Dyn et al., 1986; Broomhead and Lowe, 1988], which is the method we chose in our work.

Kriging, RBF interpolation or alternatives are limited by the dimension of the input space. A few dozens of inputs is already a large number for an interpolation method. Hence the model to be emulated needs to carefully designed with a limited number of inputs. Also, in practice, one emulator is built for each output of the model. The dimension of the model outputs should remain moderate, so that the total computational cost remains low. When the dimension of the outputs is too high, a dimension reduction is necessary. In these cases, the model outputs are projected onto a reduced subspace, and an emulator of the reduced model is built. The metamodel combines an emulator of the reduced model and a reconstruction of the model outputs.

More precisely, a model is considered as a function $\mathcal{M}$ whose input is a vector of parameters $\boldsymbol{p} \in \mathbb{R}^K$ and output is a vector $\boldsymbol{y} \in \mathbb{R}^D$: $\boldsymbol{y} = \mathcal{M}(\boldsymbol{p})$, where K is the dimension of inputs and D is the dimension of outputs. $\mathcal{M}$ is computationally costly. The proposed metamodeling aims at replacing the original $\mathcal{M}$ with a metamodel $\widehat{\mathcal{M}}$: $\widehat{\boldsymbol{y}} = \widehat{\mathcal{M}}(\boldsymbol{p})$, so that the computational cost of $\widehat{\mathcal{M}}$ is very low and $\widehat{\boldsymbol{y}}$ is close to $\boldsymbol{y}$ with the same input vector $\boldsymbol{p}$. The detailed implementation of the metamodeling method combined with dimension reduction is presented as follows.

4.2.1 Dimension reduction

The objective of dimension reduction is to project any $\boldsymbol{y}$ onto the subspace spanned by the reduced basis $(\Psi_j)_{j=1,2,\ldots,N}$, with $N \ll D$. The reduced basis is represented by a matrix $\boldsymbol{\Psi} = [\Psi_1 ... \Psi_N]$. The reduced basis is chosen in order to include the directions of variability of the outputs. Principal component analysis (PCA) is used for building the reduced basis.

Firstly, m sample points $\boldsymbol{p}^{(i)}$ $(i = 1, 2, \ldots, m)$ are generated with m sufficiently large. The sample points are generated with Latin Hypercube Sampling (LHS) to ensure that $\boldsymbol{p}^{(i)}$ are evenly spread across all possible values in their variation domain. The latter should be carefully designed in order to take into account all the possible variations of the inputs, including their uncertainty ranges. Then a training set $\boldsymbol{Y} = [\boldsymbol{y}_1 \ldots \boldsymbol{y}_m] = [\mathcal{M}(\boldsymbol{p}^{(1)}) \ldots \mathcal{M}(\boldsymbol{p}^{(m)})]$ is obtained. The mean of the training set is denoted as $\bar{\boldsymbol{y}} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{y}_i$. By subtracting $\bar{\boldsymbol{y}}$ from each column of $\boldsymbol{Y}$, the centered training set is obtained and denoted as $\bar{\boldsymbol{Y}}$: $\bar{\boldsymbol{Y}} = [\boldsymbol{y}_1 - \bar{\boldsymbol{y}} \ldots \boldsymbol{y}_m - \bar{\boldsymbol{y}}]$. The projection coefficient on the j-th principal component is denoted as $\alpha_j = (\boldsymbol{y} - \bar{\boldsymbol{y}})^T \Psi_j$. Therefore, for any output $\boldsymbol{y}$, its projection (denoted as $\widetilde{\boldsymbol{y}}$) is written as $\boldsymbol{y} \simeq \widetilde{\boldsymbol{y}} = \bar{\boldsymbol{y}} + \sum_{j=1}^{N} \alpha_j \Psi_j$. The principal components Ψ_j are sorted in a way that the first component accounts for the largest

variance of $\boldsymbol{Y}$, and then each following component accounts for the largest possible variance. The approximation of the training set is denoted as $\widetilde{\boldsymbol{Y}} = [\widetilde{\boldsymbol{y}}_1 \dots \widetilde{\boldsymbol{y}}_m]$, and total quadratic error of $\widetilde{\boldsymbol{Y}}$ compared with the training set $\boldsymbol{Y}$ is:

$$\sum_{i=1}^{m} \|\boldsymbol{y}_i - \widetilde{\boldsymbol{y}}_i\|^2 = \sum_{i=1}^{m} \|\boldsymbol{y}_i - (\bar{\boldsymbol{y}} + \sum_{j=1}^{N} \alpha_j \Psi_j)\|^2 = \sum_{i=1}^{m} \|\boldsymbol{y}_i - \bar{\boldsymbol{y}} - \sum_{j=1}^{N} ((\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T \Psi_j)\Psi_j\|^2 = \sum_{k=N+1}^{D} \lambda_k, \tag{4.1}$$

where λ_k satisfies $\bar{\boldsymbol{Y}}\bar{\boldsymbol{Y}}^T\boldsymbol{\Psi}_k = \lambda_k\boldsymbol{\Psi}_k$, for $k = 1, 2, \dots, D$. The total variance of the training set is $\sum_{i=1}^{m} \|\boldsymbol{y}_i - \bar{\boldsymbol{y}}\|^2$, and the proportion of explained variance in $\bar{\boldsymbol{Y}}$ is then $var_{\text{explained}} = 1 - \sum_{k=N+1}^{D} \lambda_k / (\sum_{i=1}^{m} \|\boldsymbol{y}_i - \bar{\boldsymbol{y}}\|^2)$. In practice, we can determine the basis size N by fixing a target $var_{\text{explained}}$, and evaluate the performance of PCA by calculating the total quadratic error using equation 4.1. At last, a dimensional reduced model is obtained as:

$$\boldsymbol{y} \simeq \bar{\boldsymbol{y}} + \sum_{j=1}^{N} ((\boldsymbol{y} - \bar{\boldsymbol{y}})^T \Psi_j)\Psi_j = \bar{\boldsymbol{y}} + \boldsymbol{\Psi}\boldsymbol{\Psi}^T(\mathcal{M}(\boldsymbol{p}) - \bar{\boldsymbol{y}}) \tag{4.2}$$

After centering, the output of the model is always in the reduced subspace spanned by $(\boldsymbol{\Psi}_j)_{j=1,\dots,N}$. The dimension of the model is therefore reduced. As mentioned in the second paragraph of this section (Section 4.2), both inputs and outputs should be low of dimension for building the emulator. The outputs dimension affects the computational cost of the metamodel (see later in Section 4.2.3). The input dimension should also remain low enough, presumably lower than 100, for the emulation introduced below to properly work. In case $\boldsymbol{p}$ is of high dimension, it may be reduced by a PCA, just like for the outputs. We can also decrease the input dimension by defining inputs that we want to analyze the most, because they are associated with large uncertainties. In our case study, the input dimension is decreased by selecting the key inputs that we want to analyze (see later in Section 4.3.2).

4.2.2 Emulation of the reduced model

The reduced model in Equation 4.2 is of lower dimension, but computing the projection coefficient $\alpha = f(\boldsymbol{p}) = \boldsymbol{\Psi}^T\mathcal{M}(\boldsymbol{p})$ is still computationally costly because $\mathcal{M}(\boldsymbol{p})$ requires running the original model. The objective of building emulator is to replace $f(\boldsymbol{p})$ with a statistical emulator $\widehat{f(\boldsymbol{p})}$ so that $\widehat{f} \simeq f$ and its computational cost is low. We need a large number of samples to establish the direct relation between the inputs $\boldsymbol{p}$ and each of the independent projection coefficients $\alpha_j = f_j(\boldsymbol{p})$. The same training samples generated in PCA are used here: the m training points $\boldsymbol{p}^{(i)}$. The emulator

applied in our study consists of two parts: (i) a linear model (a linear regression), and (ii) an interpolation between the regression residuals of the samples $f_j(\boldsymbol{p}^{(i)})$,

$$\widehat{f_j}(\boldsymbol{p}^{(i)}) = \sum_{k=1}^{K} \beta_{j,k} p_k^{(i)} + \sum_{i=1}^{m} \omega_{j,i} \phi(d_j(\boldsymbol{p}, \boldsymbol{p}^{(i)})). \tag{4.3}$$

The regression represents the linear dependencies between the input parameters p_k $(k = 1, \ldots, K)$ and the projection coefficients α_j. In practice, the most important part is the interpolation of the residuals of the regression $f_j(\boldsymbol{p}^{(i)}) - \sum_{k=1}^{K} \beta_{j,k} p_k^{(i)}$. There are many approaches for the interpolation part of the emulator, and the method used in this study relies on radial basis functions (RBF).

The concept of RBF interpolation is that the influence of a sample point $\boldsymbol{p}^{(i)}$ to the approximation of $f_j(\boldsymbol{p})$ depends only on the distance between $\boldsymbol{p}^{(i)}$ and the target point $\boldsymbol{p}$: the bigger the distance, the less $\boldsymbol{p}^{(i)}$ should be influential. The Euclidean distance is used in our study. Therefore, the interpolation with RBF of the regression residuals is:

$$\widehat{f_j}(\boldsymbol{p}^{(i)}) - \sum_{k=1}^{B} \beta_{j,k} p_k^{(i)} = \sum_{i=1}^{m} \omega_{j,i} \phi(d(\boldsymbol{p}, \boldsymbol{p}^{(i)})), \tag{4.4}$$

where $d(\cdot, \cdot)$ is the Euclidean distance, ϕ is the chosen RBF and the $\omega_{j,i}$ are weights that depend on the residuals $f_j(\boldsymbol{p}^{(i)}) - \sum_{k=1}^{K} \beta_{j,k} p_k^{(i)}$. The $\omega_{j,i}$ are computed so that the relation (4.4) is exactly satisfied at the sample points $\boldsymbol{p}^{(i)}$. This involves solving a linear system.

4.2.3 Completion of the metamodel with reconstruction step

After dimension reduction and the emulation of the projection coefficients, an approximation of the original model $\mathcal{M}$ output can be written as follows:

$$\widehat{\boldsymbol{y}} = \bar{\boldsymbol{y}} + \boldsymbol{\Psi}(\widehat{\boldsymbol{f}}(\boldsymbol{p}) - \boldsymbol{\Psi}^T \bar{\boldsymbol{y}}) = \bar{\boldsymbol{y}} - \boldsymbol{\Psi}\boldsymbol{\Psi}^T \bar{\boldsymbol{y}} + \boldsymbol{\Psi} \begin{bmatrix} \widehat{f}_1(\boldsymbol{p}) \\ \widehat{f}_2(\boldsymbol{p}) \\ \vdots \\ \widehat{f}_N(\boldsymbol{p}) \end{bmatrix} \simeq \boldsymbol{y}, \tag{4.5}$$

where $\bar{\boldsymbol{y}}$ and $\boldsymbol{\Psi}\boldsymbol{\Psi}^T$ are already known according to PCA studies. Therefore, when the model is applied to a new input vector $\boldsymbol{p}$, the computational cost is that of each $\widehat{f}_j(\boldsymbol{p})$ $(j = 1, \ldots, N)$. Each component is emulated independently, so that the total cost is N times the cost of the emulation of one component. Even though one emulator can be very fast, we could not afford the computational cost for D emulators if D is high, hence the dimension reduction to $N \ll D$ on the outputs before the emulation.

4.3 Design of experiment with case study

This section presents the analytical DTA model named LADTA and the experimental design of LADTA applied in a large-scale network of the agglomeration of Clermont-Ferrand. We designed an *atomic* LADTA simulation applied in Clermont-Ferrand and named it as LCF (for *LADTA applied in Clermont-Ferrand*). The discretization of the complete LADTA simulation applied in Clermont-Ferrand into the *atomic* simulation of LCF model allows us to carry out DUE traffic assignment with LADTA for a large scale network with low dimensional input data. An evaluation of LCF model during a reference day is also presented by comparing LCF model results with traffic flow observations in the city of Clermont-Ferrand. The methods mentioned in Section 4.2 is applied to this *atomic* LCF model in the following Section 4.4.

4.3.1 Complete LADTA model

Leurent [2003] designed an analytical DTA model, named LADTA (for Lumped Analytical Dynamic Assignment). LADTA formulates the time-varying features in DUE problem by introducing temporal profiles of traffic demand, volume loaded, travel costs and routes for each user class. It also introduced FIFO point queues to represent volume-delay relation on links. The implementation LTK (for Ladta Tool Kit) succeeded in applying LADTA on the large-scale network of Paris metropolitan area[Aguiléra and Leurent, 2009]. Wagner [2012] provided a mathematically rigorous and general formulation of LADTA, together with numerical algorithms and applications. LADTA and LTK were applied to another network at metropolitan scale in [Chen et al., 2017] for the agglomeration of Clermont-Ferrand, and the simulation results were compared with traffic measurements.

Inputs and outputs of LADTA

$\mathcal{G} = (\mathcal{N}, \mathcal{A})$: the nodes and links of an oriented graph for a modeled network. The total link number is $D = card(\mathcal{A})$.

$\mathcal{Z}_O, \mathcal{Z}_D \subset \mathcal{N}$: the set of nodes representing the Origin-Destination zones. $z = card(\mathcal{Z}_O) = card(\mathcal{Z}_D)$ and z is the number of Origin/Destination zones.

$\mathcal{H}$: the simulation period of LADTA. $\mathcal{H} = [h_0, h_1]$.

Δh: time step for loop detectors to register aggregated traffic flow data. It is also assumed that during the interval $[h, h + \Delta h]$, the traffic flow on the network is constant, and so is the time-varying traffic demand.

$\boldsymbol{C} = [C_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$, $\boldsymbol{T}_0 = [T_{0a}(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$, $\boldsymbol{V}_0 = [V_{0a}(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$: a vector of link capacity, free-flow travel time and speed limits of each link a at time h.

$\boldsymbol{X} = [X_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$, $\boldsymbol{T} = [T_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$ and : a vector of traffic volume, travel time on each link a for time h. They are outputs of LADTA model.

$\boldsymbol{U} = [U_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$: a vector of generalized travel costs on each link a for time h.

$\boldsymbol{Q}(h) = (Q_{o,d}(h))_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D}$: the time-varying O-D matrix. $Q_{o,d}(h)$ is the cumulated traffic demand (total number of vehicles in veh) from o to d during $\mathcal{H}$. It is non-decreasing and can be deduced from time-varying traffic demand *density* $q_{o,d}(h)$ (in veh h^{-1}) at instant h for the O-D pair od. $q_{o,d}(h)$ is assumed to be a piecewise linear function of h and $q_{o,d}(h)$ is constant in the interval $[h, h+\Delta h]$. With known $q_{o,d}(h)$, bounded $\mathcal{H}$ and boundary condition of $Q_{o,d}(h_0)$ and $Q_{o,d}(h_1)$, $Q_{o,d}(h)$ can be obtained by $(Q_{o,d}(h))_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D} = \int_{h_0}^{h} q_{o,d}(\tilde{h})d\tilde{h}$ with $h \in \mathcal{H}$.

LADTA model structure

LADTA is a time-continuous link-based DUE assignment model. The analytical model proposed in [Leurent, 2003] can be briefly formulated as five sub-models:

(i) *Traffic flowing* model denoted as $\mathcal{F}$, stating that the travel time on each link is a function of the *current* traffic volume circulating on a: $\boldsymbol{T} = \mathcal{F}(\boldsymbol{X}, \boldsymbol{T_0})$. The capacity is constrained by $\boldsymbol{C}$, and the travel time of each link is based on a bottle-neck model.

(ii) *Service formation* (denoted as $\mathcal{S}$), modeling the total travel costs of a route, noted as $\boldsymbol{U}_R$, resulting from the travel costs of each link: $\boldsymbol{U}_R = \mathcal{S}(\boldsymbol{T})$.

(iii) *Route choice model* denoted as $\mathcal{U}$, essentially depending on $\boldsymbol{U}_R$: $\mathcal{U}(\boldsymbol{U}_R)$.

(iv) *Route volume loading* model for each least-cost route R. This will assign the traffic demand $\mathcal{Q}$ along each least-cost route: $\boldsymbol{X}_R = \mathcal{U}(\boldsymbol{U}_R)$.

(v) *Link volume loading* model noted as $\mathcal{D}$, which gives the volume loaded on each link, as a result of (iv). A link might belong to several different routes: $\boldsymbol{X} = \mathcal{D}(\boldsymbol{X}_R, \boldsymbol{T})$

The sub-models from (ii) to (iv) are modeled for all pairs of nodes in $\mathcal{N}\times\mathcal{N}$, by considering that each node is a destination point from all other nodes, for each $h \in \mathcal{H}$. Therefore, the DUE traffic assignment modeled by LADTA can be considered as a fixed-point problem as:

$$\text{find } \boldsymbol{X} \text{ such that } \boldsymbol{X} = \mathcal{D}((\mathcal{U}\circ\mathcal{S}(\boldsymbol{T})), \boldsymbol{T}) \text{ and } \boldsymbol{T} = \mathcal{F}(\boldsymbol{T}) \tag{4.6}$$

A detailed algorithmic scheme for LADTA is presented in [Aguiléra and Leurent, 2009] for each sub-model listed above. The implementation of LADTA (called as LTK) can be applied to large-scale networks [Aguiléra and Leurent, 2009; Chen et al., 2017].

4.3.2 LADTA model applied in the agglomeration of Clermont-Ferrand: the LCF model

The input dimension is high for TA simulation at metropolitan scale, especially for dynamic simulation because of the additional time dimension. As mentioned in Section 4.2, the input dimension should be low for metamodeling process. Therefore, it is needed to decrease the input dimension and focus on the most influential inputs concerning network characteristics and time-dependent O-D matrix. For a given TA problem, the dimension of inputs could be reduced by assuming that the parameters the demand and network are somehow fixed by the infrastructure of the given metropolitan area. Some scalar multiplicative coefficients are considered as uncertain inputs of DTA model applied to a given network.

With our case study, we build an *atomic* model based on LADTA applied in Clermont-Ferrand. This *atomic* model is called LCF (for *LADTA applied in Clermont-Ferrand*). For the same simulation period with the same input condition, the output of LCF model is nearly the same as that of the complete LADTA model. This subsection presents the construction and inputs of LCF model. The following subsection 4.3.3 presents the application of LCF model for a reference day simulation at the agglomeration of Clermont-Ferrand.

Inputs and outputs for LCF model

(1) Inputs

In our case study, the agglomeration of Clermont-Ferrand has more than 200,000 residents and an area of about $300\,\mathrm{km}^2$. The modeled network has $D = 19,628$ links and 8,844 nodes, as shown in Figure 4.2. In total, there are 469 inductive loop detectors, giving real-time vehicle counts at high-time resolution (every minute) at lane-level.

The inputs $\mathcal{H}$, $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, $\mathcal{Z}_O, \mathcal{Z}_D$, $\boldsymbol{C}$, $\boldsymbol{V_0}$ and $\boldsymbol{T_0}$ mentioned in Section 4.3.1 for the complete LADTA model are specified here for the network of the agglomeration of Clermont-Ferrand. The number of links is $D = card(\mathcal{A}) = 19628$. $\boldsymbol{C}, \boldsymbol{V_0}, \boldsymbol{T_0} \in \mathbb{R}^D$. The number of zones is $z = card(\mathcal{Z}_O) = card(\mathcal{Z}_D) = 124$. In addition, links are regrouped according to link capacities and speed limits. Below we introduce the notation related to the model inputs. Detailed description of the inputs regarding network parameters is presented in Section 4.4.1.

$\boldsymbol{V}_0^{low} = [V_{0a}]_{a \in \mathcal{A}_{low}}$: the vector of low speed limits ($V_0 \leq 50\,\mathrm{km\,h}^{-1}$), with $\mathcal{A}_{low}$ as the set of links with $V_0 \leq 50\,\mathrm{km\,h}^{-1}$.

$\boldsymbol{V}_0^{high} = [V_{0a}]_{a \in \mathcal{A}_{high}}$ with $\mathcal{A}_{high} = \mathcal{A} \backslash \mathcal{A}_{low}$: the vector of high speed limits.

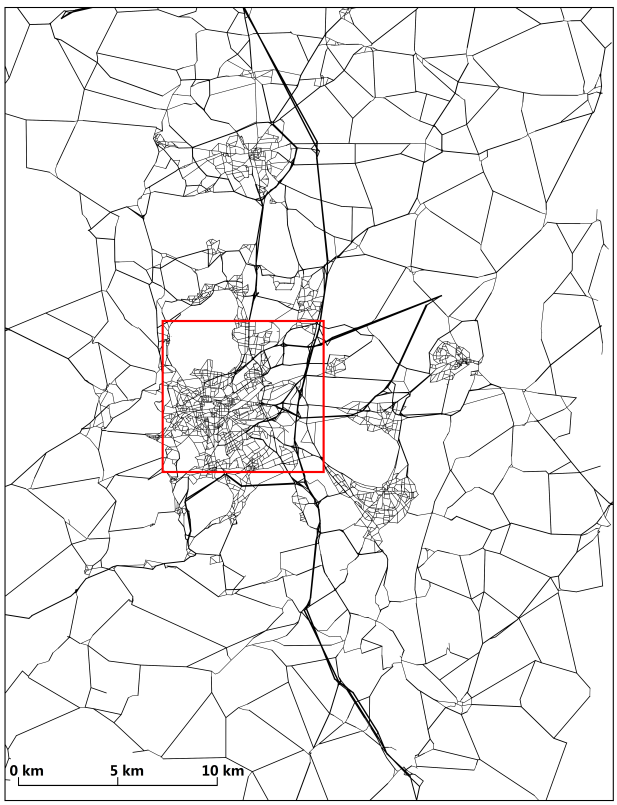


Figure 4.2: Domain of the agglomeration of Clermont-Ferrand. ($D = 19,628$ links, and 8,844 nodes, with $z = 124$ zones. It has about 200,000 residents in an area of about $300\,\mathrm{km}^2$. The red rectangle delimits the *center Clermont-Ferrand*. Concerning the measurements, there are in total 469 inductive loop detectors at lane resolution, giving vehicle counts at high-time resolution (each minute). In our case study, the average flow during each 15 min is used as measurement data.)

$\boldsymbol{C}^{small} = [C_a]_{a\in\mathcal{A}_{small}}$: the vector of small link capacities, with $\mathcal{A}_{small}$ as the set of links with $C \leq 900\,\mathrm{veh\,h^{-1}}$.

$\boldsymbol{C}^{big} = [C_a]_{a\in\mathcal{A}_{big}}$ with $\mathcal{A}_{big} = \mathcal{A}\backslash\mathcal{A}_{small}$: the vector of big link capacities.

The time-dependent traffic demand $\boldsymbol{Q}(h)$ mentioned in Section 4.3.1 can be obtained based on a static O-D matrix given for the agglomeration.

$\boldsymbol{Q}^{\text{peak}} = (q_{o,d}^{\text{peak}})_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D}$: O-D matrix of the agglomeration of Clermont-Ferrand during the evening peak hour 17:00 – 18:00, where $(q_{o,d}^{\text{peak}})_{(o,d)}$ is the average traffic demand during the evening peak. It is the static O-D matrix.

In addition, in order to study the spatial uncertainties lying in the O-D matrix, the O-D pairs are categorized into five groups according to their distance between Origin-Destination zones: $0\,\mathrm{km}$, $0-5\,\mathrm{km}$, $5-10\,\mathrm{km}$, $10-15\,\mathrm{km}$ and $>15\,\mathrm{km}$. can be divided

by 5 matrices and we have: $\boldsymbol{Q}^{\text{peak}} = \boldsymbol{Q}_{0}^{\text{peak}} + \boldsymbol{Q}_{0-5}^{\text{peak}} + \boldsymbol{Q}_{5-10}^{\text{peak}} + \boldsymbol{Q}_{10-15}^{\text{peak}} + \boldsymbol{Q}_{>15}^{\text{peak}}$.

(2) Outputs

$\boldsymbol{X}(h) = [X_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$: the vector of computed traffic volumes at time h. Note that in LADTA model, $\boldsymbol{X}(h)$ is the cumulated traffic flow (in *number of vehicles*) passing through the network. It is measurable on the whole period $\mathcal{H}$.

$\Delta h = 0.25$ hour: the time step of simulation in our case study. It is assumed that $(q_{o,d})_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D}$ is piecewise constant during $[h, h+\Delta h]$. Therefore, during the simulation period $\mathcal{H} = [h_0, h_1]$, if $(q_{o,d})$, $(Q_{o,d}(h_0))$, and $(Q_{o,d}(h_1))$ are known, the dynamic O-D matrix $\boldsymbol{Q}(h), h \in \mathcal{H}$ can be obtained by integrating $(q_{o,d})$ on $\mathcal{H}$.

$\boldsymbol{y} = \boldsymbol{X}(h+\Delta h) - \boldsymbol{X}(h)$ (in veh h^{-1}): a vector of average traffic flow during $[h, h+\Delta h]$. It is the final output of the LADTA model applied in Clermont-Ferrand.

The construction of LCF model

The LADTA model takes into account the influence of the time-varying traffic demand to compute time-dependent $\boldsymbol{X}(h)$. In the one-day DTA simulation with the complete LADTA model, the length of the simulation period is 24 h [Chen et al., 2017] (refer to Section 3.4.3 of Chapter 3). With a known *density* of the traffic demand $q_{o,d}(h)$, a given static O-D matrix $\boldsymbol{Q}^{\text{peak}}$ and the boundary condition for $Q_{o,d}(h_0)$ and $Q_{o,d}(h_1)$, the dynamic O-D matrix can be obtained by integrating $q_{o,d}(h)$ on $\mathcal{H}$. Let $P(h)$ define the *temporal variation coefficients*, representing the ratio between (i) the traffic demand during $[h, h+\Delta h]$ and (ii) during evening peak hour 17:00 - 18:00. With $\Delta h = 0.25$ h and the assumption that $q_{o,d}(h)$ is piece-wise constant, a one-day dynamic O-D matrix can be obtained by 96 $P(h)$ with $h \in I_{day} = \{h_0, h_0 + 0.25\,\text{h}, 0.5\,\text{h}, \ldots, h - 0.25\,\text{h}, h, h + 0.25\,\text{h}, h_1 - 0.25\,\text{h}\}$. I_{day} is the set of discrete time instants. The additional time dimension in LADTA model allows us to compute time-varying $\boldsymbol{X}$ on the network during the whole period $\mathcal{H}$, however, the simulation during a whole day requires big additional time dimension (96) and it is not recommended to build a meta-model with inputs of such dimension. In addition, since LADTA model is a DUE-based DTA model, it computes equilibrium at each instant during $\mathcal{H}$, the longer $\mathcal{H}$ is, the longer computational time is required. Therefore, instead of building a metamodel of LADTA with high input dimension and long computational time for a one-day DTA simulation, a metamodel is built based on an *atomic* DTA simulation with LADTA applied to Clermont-Ferrand with shorter simulation period. This *atomic* simulation is call LCF model, for ***L****ADTA applied in* ***C****lermont-****F****errand*.

The atomic LCF model takes time-varying traffic demand during a 3.25 hours period. It computes link-level traffic flow during an interval of $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$, with h_{simu} the

instant at which we want to compute traffic flow. The simulation period for an atomic simulation with LCF model is $\mathcal{H}_{\text{atomic}} = [h_{\text{simu}} - 3.25\,\text{h}, h_{\text{simu}} + 1.0\,\text{h}]$. The interval outside $\mathcal{H}_{\text{atomic}}$ is denoted as $\bar{\mathcal{H}}$. The computed $\boldsymbol{y}(h_{\text{simu}})$ during $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$ is assumed to be not affected by traffic demands during $\bar{\mathcal{H}}$. Let I_{atomic} denote the set of instants $\{h_{\text{simu}} - 2.25\,\text{h}, h_{\text{simu}} - 2.0\,\text{h}, \ldots, h_{\text{simu}} - 0.25\,\text{h}, h_{\text{simu}}, \ldots, h_{\text{simu}} + 0.75\,\text{h}\}$. Therefore, the piece-wise constant $(q_{o,d}(h))_{h \in \mathcal{H}_{\text{atomic}}}$ can be obtained by

$$q_{o,d}(\tilde{h}) = P(h) \times q_{o,d}^{\text{peak}}, h \in I_{\text{atomic}}, \tilde{h} \in \mathcal{H}_{\text{atomic}}. \tag{4.7}$$

$(P(h))_{h \in I_{\text{atomic}}}$ contains the 13 *temporal variation coefficients* for converting the static O-D matrix into the dynamic O-D matrix during $\mathcal{H}_{\text{atomic}}$. The boundary condition is set to $Q_{o,d}(h_{\text{simu}} - 2.25\,\text{h}) = 0.0$. Therefore, the dynamic O-D matrix with cumulated traffic demand $\boldsymbol{Q}$ during $\mathcal{H}_{\text{atomic}}$ can be obtained for carrying out LCF model. It is assumed that the temporal variation of the demand is independent of the spatial distribution of O-D pairs: at the same time h, $P(h)$ is the same for all $(q_{o,d}(h))_{(o,d) \in \mathcal{Z}_O \times \mathcal{Z}_D, h \in \mathcal{H}}$.

It is needed to decrease the input dimension for carrying out metamodeling. It is assumed that for the given agglomeration, the default values of $\boldsymbol{Q}(h)_{h \in \mathcal{H}_{\text{atomic}}}$, $\boldsymbol{V}_0$ and $\boldsymbol{C}$ remain unchanged. Some scalar multiplicative coefficients are chosen to represent variations and uncertainties for inputs mentioned in Section 4.3.2. The LCF model can then read, with demand and supply as inputs,

$$\begin{aligned}
\boldsymbol{y}(h_{\text{simu}}) &= \mathcal{F}(\boldsymbol{Q}(h)_{h \in \mathcal{H}_{\text{atomic}}}, \mathcal{G}) \\
&= \mathcal{F}(\boldsymbol{Q}(h)_{h \in \mathcal{H}_{\text{atomic}}}, \boldsymbol{V}_0, \boldsymbol{C}) \\
&= \mathcal{F}(P(h)_{h \in I_{\text{atomic}}}, \delta_0 \boldsymbol{Q}_0^{\text{peak}}, \delta_{0-5} \boldsymbol{Q}_{0-5}^{\text{peak}}, \delta_{5-10} \boldsymbol{Q}_{5-10}^{\text{peak}}, \delta_{10-15} \boldsymbol{Q}_{10-15}^{\text{peak}}, \delta_{>15} \boldsymbol{Q}_{>15}^{\text{peak}}, \\
&\qquad \mu_{low} \boldsymbol{V}_0^{low}, \mu_{high} \boldsymbol{V}_0^{high}, \lambda_{small} \boldsymbol{C}^{small}, \lambda_{big} \boldsymbol{C}^{big}).
\end{aligned} \tag{4.8}$$

In Equation (4.8), $P(h)$ can represent not only the temporal variation but also the uncertainty of the traffic demand. δ_0, δ_{0-5}, δ_{5-10}, δ_{10-15} and $\delta_{>15}$ are five *evening peak coefficients* to represent spatial uncertainty of the demand. μ_{low}, μ_{high}, λ_{small}, and λ_{big} are four coefficients applied to speed limits and link capacities. In addition, the direction of the traffic demand is mainly from working zones to residence zones in the given $\boldsymbol{Q}^{\text{peak}}$ since it contains traffic demands during the evening rush hour. The direction of the demand during the morning might be opposite. During morning period (before 12:00), we use $(\boldsymbol{Q}^{peak})^T$ as O-D matrix. The input for the transposition is a binary parameter η, with $\eta = 0$ when the simulation is carried out with the normal O-D matrix $\boldsymbol{Q}^{peak}$, and $\eta = 1$ when we use the transposed O-D matrix $(\boldsymbol{Q}^{peak})^T$ instead. We now focus

on the parameters $\boldsymbol{p}$ that either vary in time or parameterize the uncertainties in the inputs, and we write the LCF model as

$$\boldsymbol{y}(h_{\text{simu}}) = \mathcal{M}(P(h)_{h\in I_{\text{atomic}}}, \delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15}, \lambda_{big}, \lambda_{small}, \mu_{high}, \mu_{low}, \eta) = \mathcal{M}(\boldsymbol{p}), \tag{4.9}$$

where $\boldsymbol{p} \in \mathbb{R}^K$ is the input vector of the LCF model and its dimension is $K = 23$. One *atomic* simulation with LCF model takes about 2 hours. A metamodel named Meta-LCF is built for LCF model. With the same input vector $\boldsymbol{p}$, the Meta-LCF model also computes traffic flows at street level.

4.3.3 Case study with LCF model

In this section, we present the performance of traffic assignment simulation with the LCF for a reference day, Thursday, 20^{th} November 2014. The period of the one-day simulation with LCF is actually from 2:00 to 22:30, i.e., $h_{\text{simu}} \in \mathcal{H}_{daily} = [2.0, 22.5]$. There were then 82 runs of the LCF for the whole-day simulation. We then evaluated the original LCF model for the reference day, by comparing with observed traffic counts collected by inductive loop detectors. The outputs of traffic assignment with LCF model are the average traffic flow during each time interval $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$ on the reference day, on each road of the network in the agglomeration of Clermont-Ferrand. The comparisons were carried out between (i) the outputs of the simulations to be evaluated and (ii) the outputs of the reference simulation. The *global* criteria are computed with the spatio-temporal flows computed at all links (or all detectors) at all time intervals. The *temporal* criteria are calculated over the spatially-averaged flows with all discrete-time intervals of 15 minutes. The *spatial* criteria are calculated over the temporally-averaged flows with all links (or all detectors).

DTA with LCF model on the reference day

There are 469 detectors in the city of Clermont-Ferrand, giving real-time vehicle counts at lane-level. Spatially-averaged traffic measurements can be obtained for each interval of 15 min during a day, by averaging vehicle counts of all detectors during each interval. It is assumed that the temporal variation of the traffic demand is the same as the temporal variation of the spatially-averaged traffic flow measurement on the network. Therefore, the *temporal variation coefficients* $P(h)$ in $\boldsymbol{p}$ can be obtained by taking the ratio between the spatially-averaged traffic flow during $[h, h + \Delta h]$ and during the evening peak hour (17:00 – 18:00) of the measured traffic flow.

As mentioned in Section 4.3.2, a 3.25 hours temporal variation is needed for one *atomic* simulation with LCF model, to compute average flow during $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$. For example, for computing traffic flows during 8:00 - 8:15 (i.e. $\boldsymbol{y}(8.0)$), a vector of $(P(h))_{h\in\{5.75,6.00,\ldots,7.75,8.0,8.25,\ldots,8.75\}\text{ hour}}$ are needed. They can be obtained by taking the ratio of spatially-averaged traffic measurements during periods $\{[5:45, 6:00], [6:00, 6:15], \ldots, [7:45, 8:00], [8:00, 8:15], \ldots, [8:45, 9:00]\}$, and during evening peak.

It is assumed that there is no spatial variation in the O-D matrix and $\delta_0 = \delta_{0-5} = \delta_{5-10} = \delta_{10-15} = \delta_{>15} = 1.0$ for the reference day. It is also assumed that there is no variation in link capacities and speed limits neither: $\lambda_{big,small} = 1.0$ and $\mu_{low,high} = 1.0$. The O-D matrix transposition parameter $\eta = 1$ when LCF model is carried out before 12:00, and $\eta = 0$ otherwise. The input vector $\boldsymbol{p}_{\text{ref}}$ has then been initialized for each simulation time h_{simu}.

Model evaluation on the reference day

With $\boldsymbol{p}_{\text{ref}}$, the traffic assignment during the reference day was carried out and we compared the computed traffic flows (denoted as c_i) and measured vehicle counts (denoted as r_i) at each detector. As the computed flows are at *link level*, while the traffic counts are at *lane level*, we converted the simulated link flow to detector flows by dividing the number of lanes on the link. The formulas for calculating comparison scores are presented in Table 4.1. The scores for the simulations are presented in Table 4.2. The temporal variation of spatially-averaged flows during the whole-day is shown in Figure 4.3, for both the simulation links with LCF model and traffic flow measurements.

Table 4.1: Scores for the performance evaluation of a simulated flow. $(c_i)_i$ is a sequence of traffic flow to be evaluated. $(r_i)_i$ is the corresponding sequence of observations. n is the dimension of the sequences. $\bar{c}$ and $\bar{r}$ are respectively the means of $(c_i)_i$ and $(r_i)_i$.

Score	Formula
Mean value of the reference data	$\frac{1}{n}\sum_{i=1}^{n} r_i$
Bias	$\frac{1}{n}\sum_{i=1}^{n}(c_i - r_i)$
Normalized bias	$\frac{1}{n}\sum_{i=1}^{n}\frac{(c_i - r_i)}{\bar{r}}$
Correlation	$\frac{\sum_{i=1}^{n}(c_i-\bar{c})(r_i-\bar{r})}{\sqrt{\sum_{i=1}^{n}(c_i-\bar{c})^2}\sqrt{\sum_{i=1}^{n}(r_i-\bar{r})^2}}$
Root mean square error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(c_i - r_i)^2}$
Normalized mean square error (NRMSE)	$\frac{\text{RMSE}}{\bar{r}}$

Table 4.2: Comparison of traffic flow results between LCF and measurements for a whole-day simulation on the 20^{th} November 2014 during 2:00 - 22:30. In the *global* column, the scores are computed with all detectors and times. In the *temporal* column, the scores are calculated for the spatially-averaged flows during all time intervals of 15 minutes. In the *spatial* column, the scores are computed for the temporally-averaged flows at all detectors. The 82 runs of LCF for the whole-day simulation are launched in parallel using 82 cores.

Iteration number	50		
Simulation time	107min17s		
Comparison vectors	Global	Temporal	Spatial
Mean bias error ($\mathrm{veh\,h^{-1}}$)	7	7	7
RMSE ($\mathrm{veh\,h^{-1}}$)	161	14	123
NRMSE (%)	75.53	6.54	57.57
Correlation	0.70	1.00	0.59

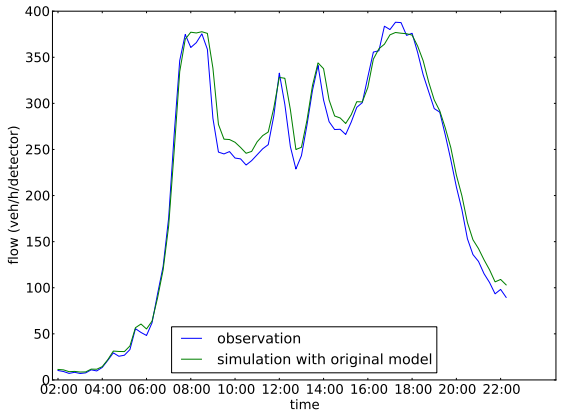


Figure 4.3: Temporal variation of the spatially-averaged flow on Thursday, the 20^{th} November 2014. For each time step of 15 minutes, we compute the traffic flow averaged over all detectors in the network. This figure illustrates the temporal distribution of the traffic over the simulation period.

The distribution of the traffic flow at all detectors and during each 15 minutes is shown in Figure 4.4 (left), and the distribution of the errors is also shown in Figure 4.4 (right). The results show that the spatio-temporal errors are mainly in $[-200, 200]$ ($\mathrm{veh\,h^{-1}\,detector^{-1}}$). Figure 4.4 (left) shows that overestimations of traffic flows are found on links with measured flows less than $100\,\mathrm{veh\,h^{-1}\,lane^{-1}}$, especially on with very

low traffic measurements ($< 100\,\text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$). There are some underestimation of flows on links with relatively heavier traffic ($> 500\,\text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$) measurements.

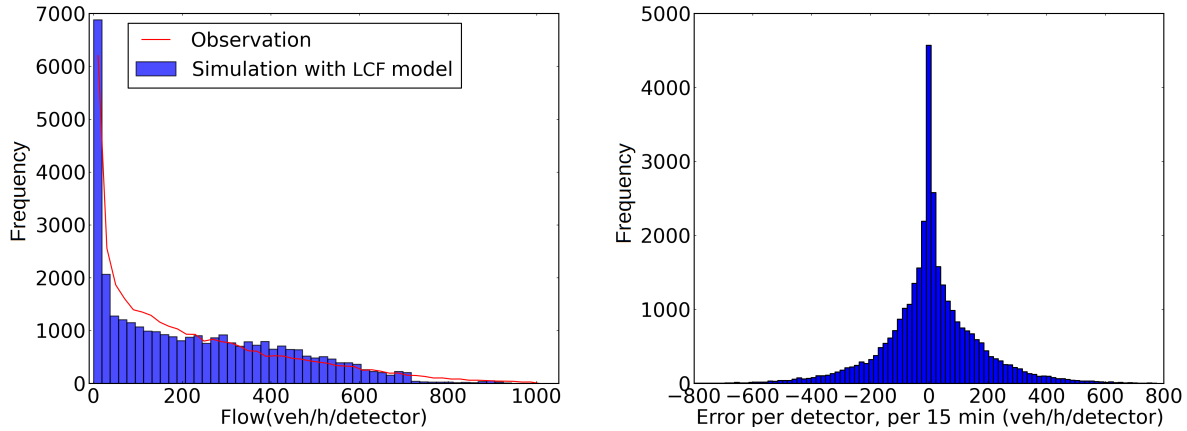


Figure 4.4: The distribution of the assigned traffic flow (left), and its error (in $\text{veh}\,\text{h}^{-1}$, right), computed using all 15-minute flows at the detectors, from 02:00 to 22:30 on the 20^{th} November 2014 (in $\text{veh}\,\text{h}^{-1}$).

The temporal and spatial performance is evaluated with the scatter plots in Figure 4.5. As shown in Figure 4.3 and Figure 4.5 (left), the LCF model produces a relatively good estimation for the daily temporal profile. We inputted the observed temporal profile, and the model did not deteriorate this information during the traffic assignment. However, the traffic assignment with LCF model is less accurate for simulating the users road choices (Figure 4.5, right). We have relatively big spatial errors, and consequently spatio-temporal errors (Figure 4.4 and Figure 4.5, right). This confirms the conclusions from Table 4.2.

The limitation of the LCF model in predicting the spatial distribution of the traffic in Clermont-Ferrand might be due to five reasons. (i) The assumption that the traffic are uniformally distrubuted on lanes of a same link ignores that there are surely differences of traffic flows on different lanes. In fact, we are not comparing exactly the same values. The flows computed by the model are the flows on each link, while the flows measured are flows on each lane. This difference brings *representativeness errors*. An example of this kind of error is presented in the following paragraph. (ii) The errors in the estimation of the O-D matrix, and the underlying difference of user classes in a static O-D matrix. In fact, in static O-D matrix during rush hours, the proportion of users with travel purpose *to/from work/school* is greater than other categories of travelers (leisure, personal, social, etc.). However, as we convert this matrix into a dynamic one during a whole day, the proportion of different users categories might change during

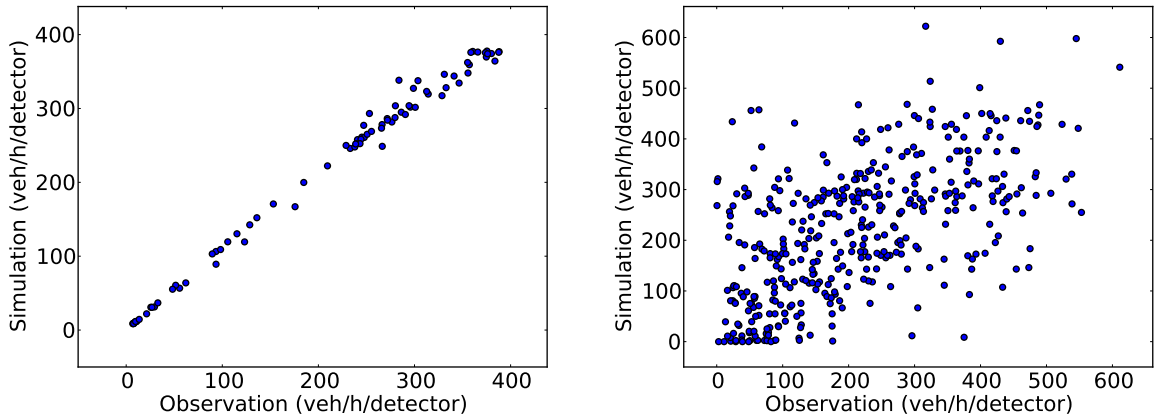


Figure 4.5: Evaluation of the daily temporal variation (left) and spatial variations (right) between the simulation outputs LCF and the measurements. In the temporal scatter plot, the flows are firstly averaged over all 469 detectors, and each 15-minute flow is then reported. In the spatial scatter plot, the flows are firstly averaged over the whole simulation period from 02:00 to 22:30, and then reported for each detector.

different period of a day. This might lead to spatial differences in O-D matrix between rush hours and off-peak periods. These spatial differences were not taken into account when converting the static O-D matrix into the dynamic one in our case study. Besides, there might be errors lying in the modeling of O-D matrix based on the household travel survey, and there are also uncertainties in the survey itself. (iii) We assume that all the users choose their paths by minimizing their travel time, while users may use different criteria in reality. (iv) For several paths between an O-D pair where there is no congestion, the model will assign all the demand on one and only one path. This is not exactly the case in reality. Randomness should be added in route choice strategies in these cases. (v) There is a toll highway (A89) connecting the north part of the agglomeration (Riom) to the center of Clermont-Ferrand, and the south part of the agglomeration (for instance, Auvergne). The toll price was not taken into account for calculating the travel cost. The spatial distribution on this highway may therefore be different between traffic assignment with LADTA and the measurements.

Here is an example for representativeness errors. Figure 4.6 (left) shows the spatial distribution of bias between the computed and observed flows. It helps us to locate the links with big errors when comparing with observations. In the red rectangle of Figure 4.6 (left), there are big overestimations on certain links. These links have three lanes, and they are located at the free north-south county road D2009 connecting the residence zone Riom, and the center of Clermont-Ferrand. However, there are only

left-turning detectors on theses links. The uniform distribution of traffic on lanes of a same given link is not applicable on these links. In fact, on this north-south country road, the traffic on straight lanes are significantly heavier than the traffic on left turn lanes. By deleting these 6 detectors only, the spatial NRMSE decreases from 57.57 % to 53.71 % (see results in Table 4.3 and in Table 4.2).

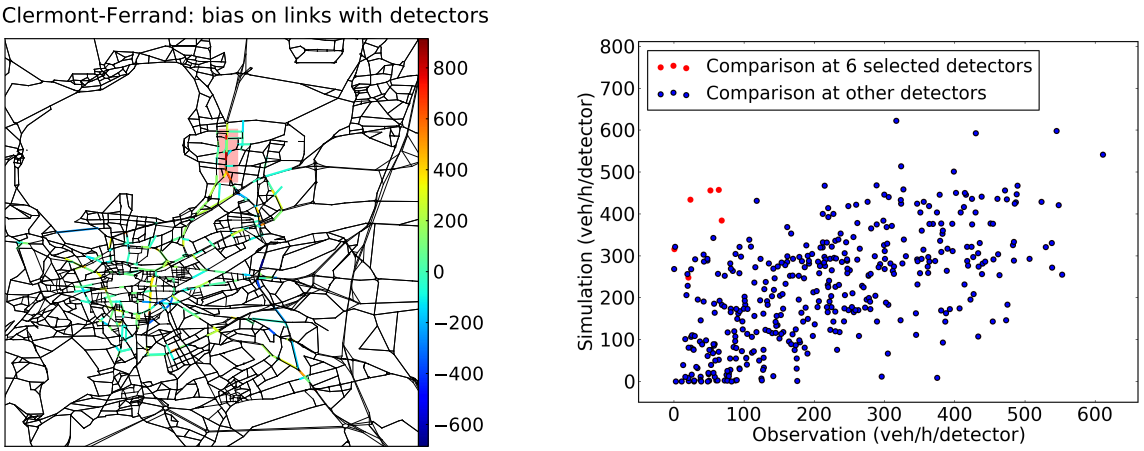


Figure 4.6: Left: map of bias (in veh h^{-1} $link^{-1}$) between the temporally-averaged flows computed by LCF on each link and the measured flows, on the 20^{th} November 2014. (In the red rectangle, the links with overestimated flows have detectors only on left-turn lanes.) Right: scatter plot of the temporally-averaged flows on each detector. (The red points match the simulation-observation comparison at detectors located on left-turning lanes in the red-rectangle zone of the left figure.)

4.4 Metamodeling for the dynamic traffic assignment model for the agglomeration of Clermont-Ferrand and performance evaluation

In this section, we applied the metamodeling method mentioned in Section 4.2 to the LCF model $\mathcal{M}(\boldsymbol{p})$ in Equation 4.9 whose input dimension is $K = 23$. A fast metamodel is built and it preserves the main features of the original LCF model: $\widehat{\mathcal{M}}(\boldsymbol{p}) = \widehat{\boldsymbol{y}} \simeq \boldsymbol{y} = \mathcal{M}(\boldsymbol{p})$. The output dimension $D = 19,628$: $\boldsymbol{y} \in \mathbb{R}^D$ and $\widehat{\boldsymbol{y}} \in \mathbb{R}^D$. The resulted metamodel is called Meta-LCF. Then the outputs of Meta-LCF is compared with those of LCF model, and the outputs of the two models are also compared with observation

Table 4.3: Comparison of traffic flow results between LCF and measurements for a whole-day simulation on the 20^{th} November 2014 during 2:00 - 22:30. In the *global* column, the scores are computed with all detectors and times. In the *temporal* column, the scores are calculated for the spatially-averaged flows during all time intervals of 15 minutes. In the *spatial* column, the scores are computed for the temporally-averaged flows at all detectors. The scores are computed without the 6 left-turning detectors.

Iteration number	50		
Comparison vectors	Global	Temporal	Spatial
Mean bias error ($\mathrm{veh\,h^{-1}}$)	6	6	5
RMSE ($\mathrm{veh\,h^{-1}}$)	155	13 (14)	116 (123)
NRMSE (%)	71.93	6.43	53.71
Correlation	0.72	1.00	0.63

data collected by inductive loop detectors in the network of Clermont-Ferrand, during a one-month simulation.

4.4.1 Input variation intervals

We define the variation intervals of each input element p_i in $\boldsymbol{p}$ of LCF model. The variation intervals are chosen to cover the possible uncertainties in inputs.

Traffic demand

For the *temporal variation coefficients*, we consider that $P(h)$ varies in $[0.0, 1.5]$ so that the traffic demand at instant h can vary from 0 % to 150 % of the average flow during evening peak hour. This range of variation allows us to build all the temporal profiles we need to reproduce the traffic demand during any period of the day, for any day type (working days, weekends and holidays).

One of the most important uncertainty source of LCF model is the O-D matrix of the agglomeration of Clermont-Ferrand. It is based on surveys on a limited part of the citizens. Based on the surveys, one can use a model to generate the O-D matrix, which can introduce large uncertainties in the O-D matrix. Besides, most of the travels in the O-D matrix are *to/from work*, from Monday to Friday. If we want to compute the traffic volume during weekends or holidays, the O-D matrix might be completely different.

We categorize 5 groups of O-D pairs based on the distance between their centroids: 0 km (the diagonal of the O-D matrix), $0-5$ km, $5-10$ km, $10-15$ km and > 15 km. For each of these 5 categories, the variation of the corresponding *evening peak coefficients*

is set as $[0.25, 1.50]$, in order to take into account all possible spatial variations and uncertainties lying in O-D matrix.

Link capacities and speed limits

The capacities and speed limits are characteristics of the road network of Clermont-Ferrand. As for link capacities, we categorize the links in two classes: the "big links" whose capacities are strictly bigger than $900\,\text{veh}\,\text{h}^{-1}$, and the "small links" with capacities less than or equal to $900\,\text{veh}\,\text{h}^{-1}$. For each category, we consider two multiplicative coefficients for the link capacity (λ_{big}, λ_{small}). Because of the uncertainties, the capacities are assumed to vary by $\pm 30\,\%$, and $\lambda_{big}, \lambda_{small} \in [0.7, 1.3]$ for big and small links.

Although the speed limits are defined by transportation regulations, they are also subject to uncertainties. In fact, the uncertainty may lie in the assumption that users travel at the speed limit in free flow condition. In some cases, users cannot always travel at the maximum speed even if there is no congestion according to the *fundamental diagram* presented in Chapter 1. In other cases, users might also exceed a little bit the speed limit (rural areas and highways). We then categorize network links according to their speed limit. A link is considered as *"high speed link"* if its speed limit is $> 50\,\text{km}\,\text{h}^{-1}$, which should correspond to a link in a rural area or to highways. A link is considered as *"low speed link"* if the speed limit is $\leq 50\,\text{km}\,\text{h}^{-1}$, which corresponds to a link in an urban area. Two multiplicative coefficients are applied to the speed limit of these two categories, μ_{high}, μ_{low}. The variation of these two coefficients is different since the road conditions and users behavior might be different in urban areas, rural areas or highways. We consider that the speed variation of *"high speed link"* is $80\% - 120\%$ of the modeled speed limits, and the speed variation of *"low speed link"* is $70\,\% - 110\,\%$. Figure 4.7 presents the link capacities and speed limitations in the agglomeration of Clermont-Ferrand.

The variations of each element in the vector $\boldsymbol{p}$ to build the training sets are presented in Table 4.4. Then sample points are generated with LHS of the p_i varying in their intervals. Note that η is fixed to 1 if we build a metamodel of LCF to be applied before 12:00, and it is fixed to 0 when building the metamodel for LCF applied to case studies after 12:00. The *variation* of η is not an *uncertain value*, as for the other 22 inputs.

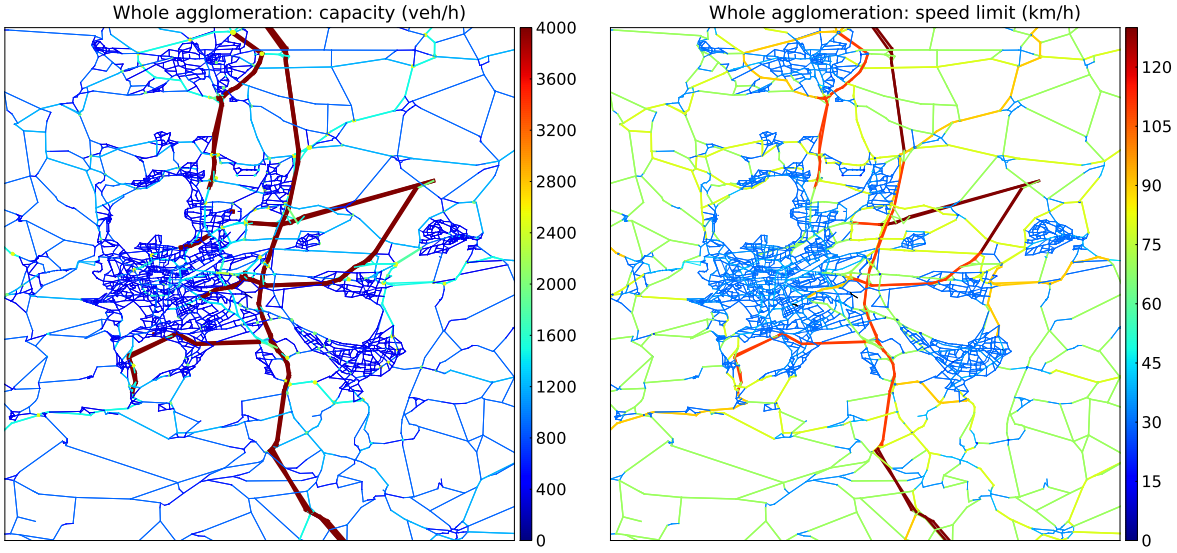


Figure 4.7: Capacity map (left) and speed limit map (right) for the agglomeration of Clermont-Ferrand.

Table 4.4: The inputs of LCF model and their variations for training sets

Input	Temporal profile	Capacity	Speed	Demands in O-D matrix	Transposed O-D matrix
Symbol	$(P(h)_{h\in\mathcal{H}_{\text{atomic}}}$	$\lambda_{\text{big}}, \lambda_{\text{small}}$	$\mu_{\text{high}}, \mu_{\text{low}}$	$\delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15}$	η
Number	13	2	2	5	1
Variation	$[0.00, 1.5]$	$[0.70, 1.30]$	$[0.80, 1.20]$, $[0.70, 1.10]$	$[0.25, 1.50]$	1 or 0

4.4.2 Metamodeling for LCF model

The sample size $m_{OD} = m_{OD_{transposed}} = 3003$ for the cases with the given O-D matrix as demand input ($\eta = 0$ in Equation 4.9), or with its transposition ($\eta = 1$ in Equation 4.9). We generated m samples using LHS approach for each case, i.e. for 6,006 training atomic simulations with LCF model in total. Then the PCA is carried out. The outputs of LCF are completely different if we use the original O-D matrix, or the transposed matrix. Therefore, we decided to carry out one PCA for each case, resulting in two different reduced bases. The projection of the outputs on the subspaces resulting from the PCA studies were compared with the outputs of the training simulations. We conditioned the explained variance to choose the basis dimension N as mentioned in Section 4.2.1. The errors presented in Section 4.2.1 are analyzed by using the root-

mean-square error (RMSE) between the outputs and their projection for evaluating PCA performance. Detailed evaluation of the dimension reduction is presented in the following Section 4.4.3

At last, we used the 3003 input samples, that were generated with LHS, to built N emulators of the N reduced models with input $\boldsymbol{p}$ and one of the N projection coefficients (on the principal components) as output. This was carried out once for the first half of the day (before 12:00), and once for the second half of the day (after 12:00). A linear trend was firstly fitted for each projection coefficient. Then we relied on the Python module SciPy to carry out the interpolation with radial basis functions (RBF). The cubic RBF $\phi : r \mapsto r^3$ was chosen for the interpolation. 13 and 15 emulators were built for the projection coefficients on the principal components, for simulations with LCF before 12:00 and after 12:00 respectively. The Meta-LCF was then built, and a detailed model evaluation is presented in the Section 4.4.4

An *atomic* simulation of the original LCF requires about 117 min on one core. 3003 evaluations of the original LCF model are needed to build a training set for a metamodel. Once the training set is available, the construction of emulators is very fast. The computational time for emulators construction only increases linearly with the reduced basis size. Once the Meta-LCF is built, it has a very low computational cost. The evaluation time of a single atomic simulation decreases to 0.022 s with Meta-LCF. The computational time of the DTA simulation during the reference day decreases from several hours to several seconds. Then we were able to carry out a one-month simulation with Meta-LCF model in less than 1 minute on one core. Even though there are some losses due to dimension reduction and emulation of the original model as shown in the following Section 4.4.3, the metamodel shows a very similar behavior as the original LCF model. The performance of Meta-LCF model is essentially the same as the original model when compared with loop detector measurements. Detailed results concerning Meta-LCF performance are presented in Section 4.4.4 and Section 4.4.5.

4.4.3 Performance evaluation of dimension reduction

The results in Table 4.5 show the performance of dimensionality reduction for different sizes of the basis. We denote $y_{a,i}$ as the computed flow on road a of with the sample point $p^{(i)}, i \in [1, 2, \ldots, m]$, and we denote $y'_{a,i}$ as the projection of $y_{a,i}$ on the reduced subspace. *Total* RMSE and NRMSE in Table 4.5 are calculated between $y_{a,i}$ and $y'_{a,i}$ at each link a for each input point $p^{(i)}, i \in [1, 2, \ldots, m]$. At each road $\tilde{a}$, the RMSE and NRMSE are calculated between $y_{\tilde{a},i}$ and $y'_{\tilde{a},i}$, with $i \in [1, \ldots, m]$, and we can get $\text{RMSE}_{\tilde{a}}$ and $\text{NRMSE}_{\tilde{a}}$. Then the average *spatial* NRMSE in Table 4.5 is the average

over all $\text{NRMSE}_{\tilde{a}}$, with $\tilde{a} \in \{1, 2, \ldots, 19628\}$. The results in Table 4.5 show that the PCA performance can be improved by increasing the basis size, but it will also increase the dimension of the subspace (i.e., the number of principal components). Consequently, this will increase the total computational cost for the metamodel (as mentioned in Section 4.2.3). It is necessary to choose a low-dimension reduced subspace, yet keep a reasonable good performance of the projections.

Table 4.5: Comparison of the traffic flow projection against the original simulations, for different sizes of the basis, using the original O-D matrix $\boldsymbol{Q}^{\text{peak}}$ (simulation during periods after 12:00, $\eta = 0$), and the transposed O-D matrix $(\boldsymbol{Q}^{\text{peak}})^T$ (simulation during periods before 12:00, $\eta = 1$).

Explained variance	0.80		0.85		0.90		0.95		0.98	
O-D matrix transposition η	0	1	0	1	0	1	0	1	0	1
Reduced basis size	5	4	7	6	15	13	63	56	228	218
Mean bias error (veh h^{-1})	0	0	0	0	0	0	0	0	0	0
Total RMSE (veh h^{-1})	54	52	47	47	39	38	28	27	18	17
Total NRMSE (%)	41.63	40.20	35.96	31.19	30.18	29.62	21.34	21.14	13.49	13.37
Avg spatial RMSE (veh h^{-1})	36	35	31	31	26	26	20	19	13	13
Avg spatial NRMSE (%)	31.9	31.591	28.86	29.41	25.23	25.22	20.39	20.53	14.81	15.05
Correlation	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	1.00	1.00

Now we focus on where the errors mainly come from. Figure 4.8 shows the dispersion of the spatial NRMSE_a against the average computed flow of each link a over all M training samples ($\frac{1}{m}\sum_{j=1}^{m} y_{a,k}$, with $k = 1, 2, \ldots, m$; $m = 3003$ and $a = 1, 2, \ldots, D$; $D = 19628$). It can be observed that bigger errors (with $\text{NRMSE}_a > 50\%$) are found on links whose traffic computed by LCF is relatively low (< 200 veh h^{-1}). Besides, Figure 4.9 shows the distributions of all computed 15-minute traffic flows (from all links), for the training simulations and for their projections. It is observed that the projection has its limitations in reproducing the lowest traffic flows computed by LCF, especially when the basis dimension is low. The projection performance gets better as basis dimension increases. However, in both Figure 4.8 and 4.9, we find that the increase of the basis dimension mainly improves the performance on links with lower flows. The improvement of the comparison criteria shown in Table 4.5 mainly comes from a better performance on these links, which are often less important in a DTA simulation. A PCA basis with explained variance of 0.90 can already reproduce a relatively good projection of the computed flows in training simulations. Therefore, for the construction of the

metamodel, we chose a PCA reduced basis with the explained variance equal to 0.90. The final chosen basis size for the case with the O-D matrix (i.e., simulations after 12:00) is $N_{non\_transposed} = 15$, and $N_{transposed} = 13$ for the simulation cases with the transposed O-D matrix (i.e., simulations period before 12:00). Figure 4.10 illustrates two of the 13 selected principal components, for the case of simulation with the transposed O-D matrix for morning periods.

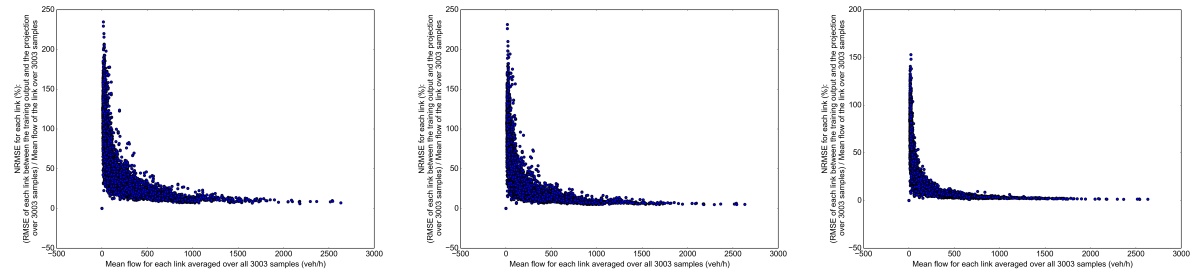


Figure 4.8: The dispersion of NRMSE_i (projection error) against the flow for all links over the 3,003 samples, for the simulation case with the non-transposed O-D matrix. Left: explained variance = 0.8, $N = 6$; center: explained variance = 0.9, $N = 15$; right: explained variance = 0.98, $N = 228$. (We compute the RMSE for each link i over the average flows of 3,003 samples, and divide the RMSE by the average flow to get NRMSE_i.)

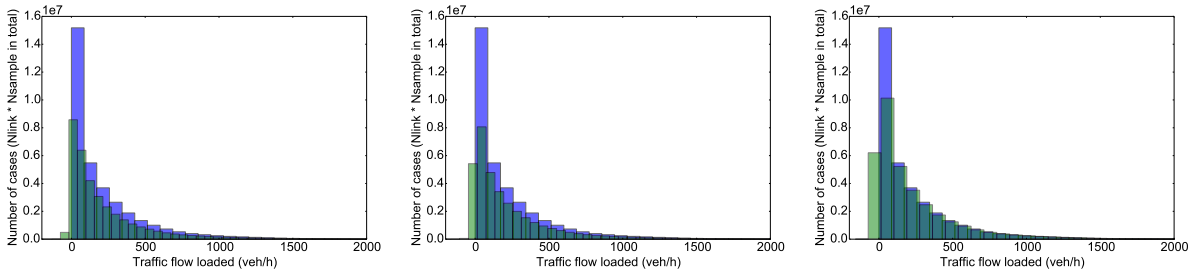


Figure 4.9: The distribution of traffic assigned by LCF (blue) and its projection (red), for all links of all 3,003 samples. Left: explained variance = 0.8, $N = 6$; center: explained variance = 0.9, $N = 15$; right: explained variance = 0.98, $N = 228$.

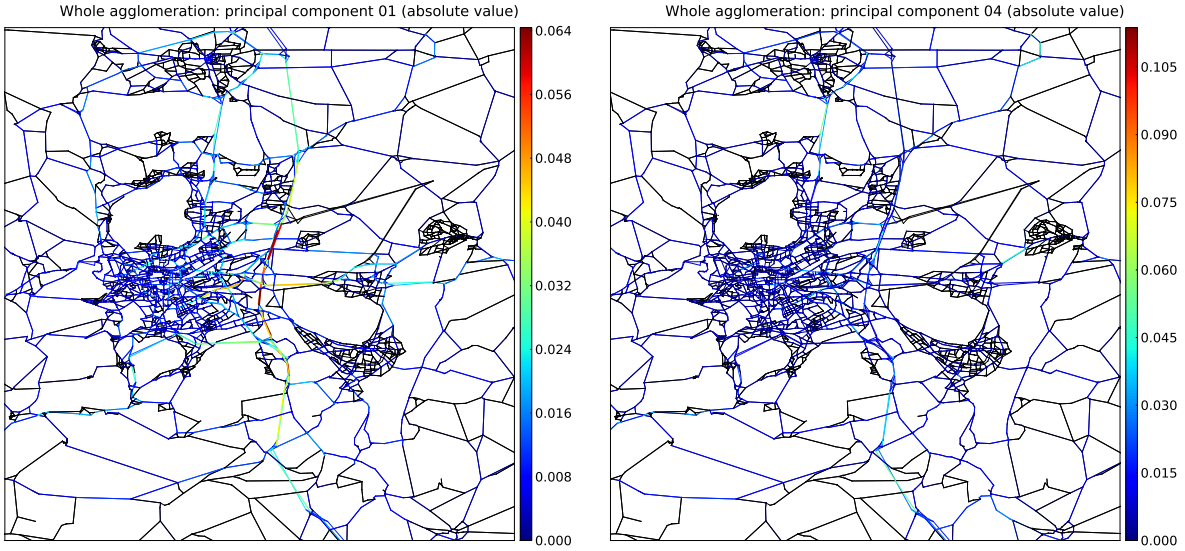


Figure 4.10: The figures represent two of the 13 selected principal components. (The first principal component (left) explains the most important part of the variance, and highlights links with high capacities. The fourth principal component (right) represents the effect of the links with higher speed limit, i.e. shorter free-flow travel time.)

4.4.4 Comparison between the metamodel of LCF and the original LCF model

In this subsection, we carried out a one-month DTA simulation with the original LCF model and the Meta-LCF model for November 2014. Both models rely on the same inputs, and the performance of Meta-LCF model is first evaluated by comparing its outputs to those of the original model. The results of both models are also compared with inductive loop detector observations in the following Section 4.4.5.

The traffic demand varies from one day to another. We introduce here a new multiplicative coefficient: the *total demand coefficient* ξ_{day} for the simulation *day*. The demand variation for a given simulation *day* can then be represented by *temporal variation coefficients* ($P_{day}(h)$) and ξ_{day}. It is assumed that there is no spatial difference in the O-D matrix so that $\delta_0 = \delta_{0-5} = \delta_{5-10} = \delta_{10-15} = \delta_{>15} = 1.0$. It is also assumed that, during one month, the inputs regarding street capacities and speed limits take the default value so that $\lambda = \mu = 1.0$ in Equation 4.9. $P_{day}(h)$ and δ_{day} are obtained as follows. It is assumed that the temporal variation of traffic flow measured by loop detectors can reflect the time variation of traffic demand in the dynamic O-D matrix. We used the observations from September 2013 to August 2014 as learning samples, and

built 16 *typical temporal profiles*. The 16 different day types in Clermont-Ferrand are (i) seven weekdays during normal period, (ii) seven weekdays during school vacation period, (iii) a public holiday during normal period and (iv) a public holiday during school vacation period. Then for each day type *day*, $P_{day}(h)$ is obtained by taking the ratio between (i) the measured spatially-averaged traffic flow (in veh h^{-1}) during $[h, h + \Delta h]$ and (ii) the average traffic flow (in veh h^{-1}) during 17:00 - 18:00 on that day. ξ_{day} are obtained by taking the ratio between (i) the measured traffic flow (in veh h^{-1}) during the evening peak hour of the typical day *day* and (ii) that on the reference day. Therefore, for computing traffic flow during $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$ on the typical *day*, the input vector $\boldsymbol{p}$ in the Equation 4.9 is set as

$$\boldsymbol{p}_{day}(h_{\text{simu}}) = \begin{cases} (\xi_{day}(P_{day}(h))_{h\in I_{\text{atomic}}}, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1) & h_{\text{simu}} < 12\,\text{h} \\ (\xi_{day}(P_{day}(h))_{h\in I_{\text{atomic}}}, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0) & \text{otherwise.} \end{cases} \tag{4.10}$$

For computing traffic flows at link resolution at time resolution of every 15 minutes from 02:00 to 22:30, h_{simu} is in the set $\{2.0, 2.25, 2.5, \ldots, 22.0, 22.25, 22.5\}$ h. Figure 4.11 gives an example of spatially-averaged profile of all the Tuesdays from September 2013 to August 2014.

From Figure 4.11, considering the Tuesdays in normal periods, the differences between the traffic flows at peak hours of different days are very small. Moreover, the total volume of traffic during a day remains almost unchanged. The same feature is observed for other weekdays, so that the temporal variation during a given weekday can be represented by an average temporal profile (yellow line). However, Figure 4.11 (right) shows that during the vacation periods, bigger differences are found in both the total daily volume and the temporal variation of the traffic. These may lead to errors in our simulations.

We compared the traffic flows computed by Meta-LCF and those computed by the original LCF model during November 2014 from 2:00 to 22:30 (local time) each day: $\boldsymbol{y}(h)$ computed by the two models with $h \in \{2.0, 2.25, 2.5, \ldots, 22.0, 22.25, 22.5\}$ (hour). During November 2014, there were (i) one day of public holiday during school vacation (the 1^{st} November), (ii) one public holiday during normal period (the 11^{th} November), (iii) one Sunday during school vacation (the 2^{nd} November) and (iv) weekdays during normal period. The corresponding typical day profiles are used as demand inputs. The outputs are traffic flows at link-level during all 15-minutes intervals $[h, h + \Delta h]$, and are denoted $(y_{i,h})$ for the original model and $(x_{i,h})$ for the metamodel.

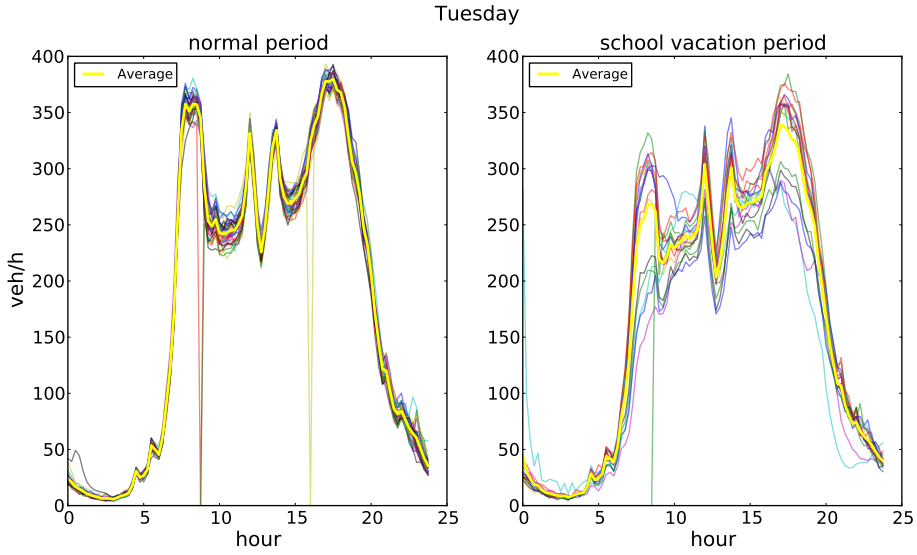


Figure 4.11: Temporal profiles of spatially-averaged flows of all Tuesdays from September 2013 to August 2014. The profiles in yellow are average profiles that we use to represent typical temporal profiles as inputs of LCF and Meta-LCF models. Each thin line represents the temporal profile of traffic of a Tuesday during normal periods (left) and school vacation periods (right), and the yellow line represent the average temporal profile of traffic flow over all Tuesdays during each kind of period. Several points reach zero when all detectors failed. This rarely happens, less than 5% of the time, and these failed values have not been taken into account for calculating of the average temporal profiles.

Table 4.6 summarizes the comparison between $(y_{i,h})$ and $(x_{i,h})$ against different basis sizes. The results show that the performance of the metamodel improves slightly when increasing the basis size, i.e., the number of principal components. These results are consistent with the projection performance in PCA studies. The computational time for constructing emulators increases with the increase of basis size as well.

Figure 4.12 shows the comparison between the flows computed by LCF and Meta-LCF model at link-level. The Meta-LCF model can well predict both the spatially-averaged flows over all time steps, and the temporal-averaged traffic flows over all links.

Figure 4.13 shows the spatial distribution map of traffic flows computed by the original LCF model and the Meta-LCF for the one-month simulation. For each link, the temporally-averaged flows are calculated and displayed in the Figure 4.13. It shows that the spatial distribution of the flows computed by the two models are similar. Figure 4.14 presents the comparison criteria of the metamodel against the original model. The nor-

Table 4.6: Comparison of the traffic flows computed by the metamodel against the results from the original model, for different sizes of the reduced basis. The simulations were carried out for the same period of the original LCF model evaluation, i.e., November 2014, from 2:00 to 22:30 each day. (The *global* comparison criteria are calculated for spatio-temporal flows at all links and times. The *temporal* comparison are carried out based on spatially-averaged flows for all time intervals. The *spatial* comparison criteria are calculated for temporally-averaged flows on all links.)

Explained variance	0.80	0.85	0.90	0.95	0.98
Basis size (simulation periods before 12:00 and after 12:00)	4/4	6/7	13/15	56/63	218/228
Computational cost for building Meta-LCF (s)	54.319	85.734	182.412	773.625	3070.521
Computational cost per run of Meta-LCF (s)	0.0147	0.0151	0.0220	0.0442	0.1471
Global bias ($\mathrm{veh\,h^{-1}}$)	0	0	2	2	2
Global RMSE ($\mathrm{veh\,h^{-1}}$)	47	45	44	43	44
Global NRMSE (%)	54.15	51.70	51.20	50.00	50.53
Global Correlation	0.98	0.98	0.98	0.98	0.98
Temporal bias ($\mathrm{veh\,h^{-1}}$)	0	0	2	2	2
Temporal RMSE ($\mathrm{veh\,h^{-1}}$)	2	2	3	3	4
Temporal NRMSE (%)	2.68	3.17	4.55	4.47	4.70
Temporal Correlation	1.00	1.00	1.00	1.00	1.00
Spatial bias ($\mathrm{veh\,h^{-1}}$)	0	0	2	2	2
Spatial RMSE ($\mathrm{veh\,h^{-1}}$)	24	23	20	19	19
Spatial NRMSE (%)	28.52	26.50	23.37	22.12	22.44
Spatial Correlation	0.99	0.99	0.99	0.99	0.99

malized bias and NRMSE for each link is calculated according to Table 4.1. Comparing Figure 4.13 and Figure 4.14, it is observed that most of the large errors are found on links where the flows computed by the original model are low (less than $200\,\mathrm{veh\,h^{-1}}$, see Figure 4.13 as reference). This is the similar conclusion as in Section 4.4.3, where errors lying in projection of PCA studies are higher on links with lower computed flows.

4.4.5 Comparison between the metamodel of LCF and the traffic flow measurements

We compared measured traffic flows and the computed flows for the simulation of November 2014, with both the original LCF model and Meta-LCF. The scores in

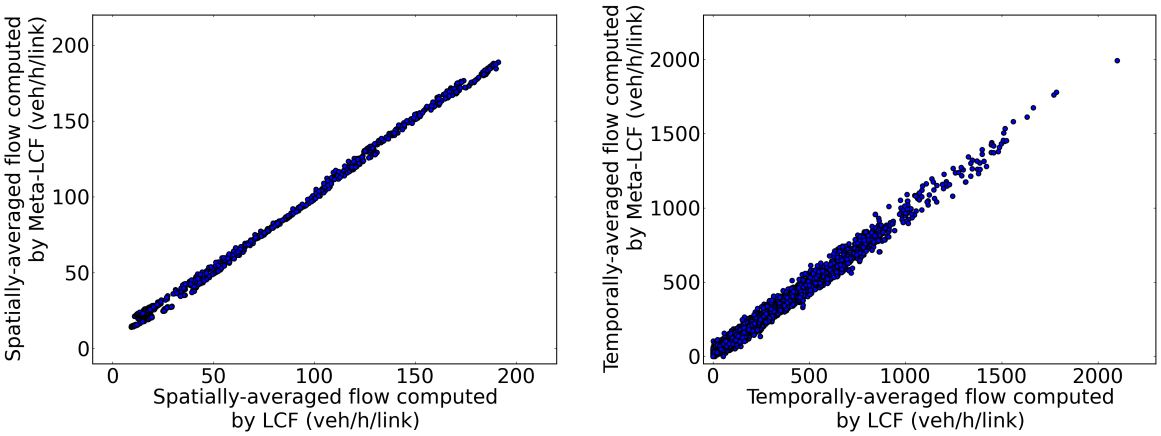


Figure 4.12: Evaluation of the spatially-averaged traffic flows (left) at all time steps, and temporally-averaged traffic flows (right) on all links computed by LCF and Meta-LCF model. (For the spatially-averaged flows, the flows are firstly averaged over all links and then reported during each 15-minutes interval. For the temporally-averaged flow, the flows are firstly averaged over the whole simulation period during one month and then reported for each link.)

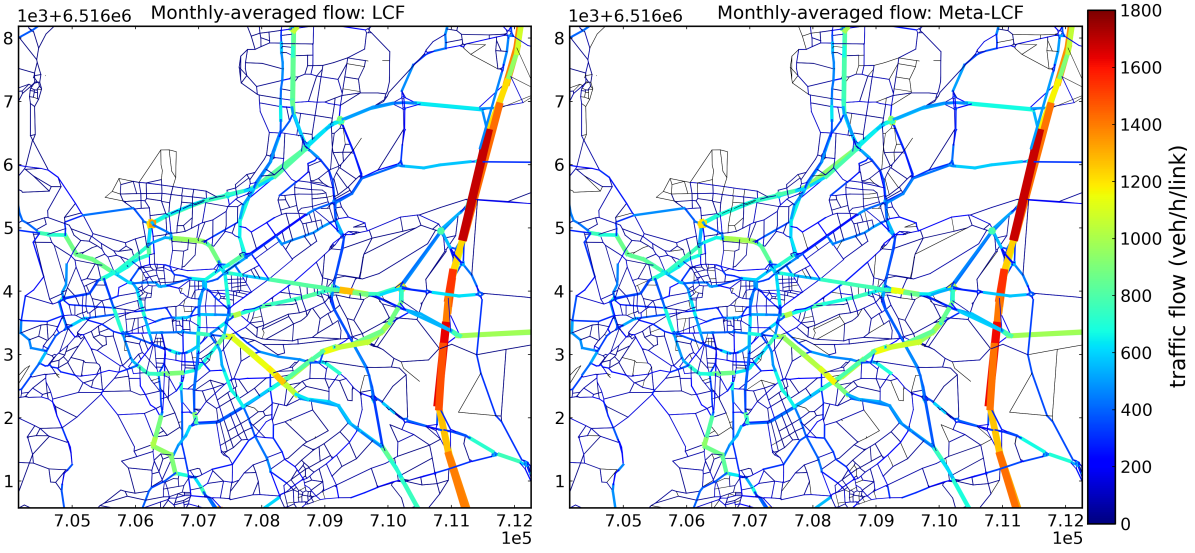


Figure 4.13: Maps of temporally-averaged flow computed by models during the whole month for the center Clermont-Ferrand. Left: flows computed by the original model; right: flows computed by the metamodel. For the links that share the same nodes but are oriented in opposite directions, the sum of the computed flows are calculated and displayed. The projection system is Lambert-93.

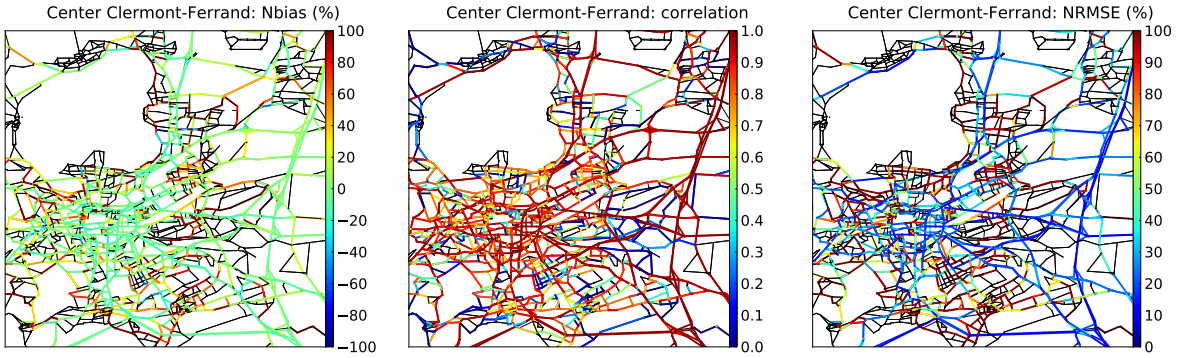


Figure 4.14: Maps of normalized bias (left, in %), correlation (center) and normalized RMSE (NRMSE, right, in %), over the center of Clermont-Ferrand, between the flows computed by the Meta-LCF and the original LCF model. (For each link, the normalized bias and the NRMSE are the bias and RMSE divided by the temporally-averaged flows computed by the original model. Bigger errors are found on links whose computed flows are lower than 200 veh h^{-1}.)

Table 4.7 are computed similarly as in Section 4.4.4, except that the comparison is at detector level, as in the LCF model evaluation from Section 4.3.3. It is observed that the metamodel has slightly better performance than the original model. Furthermore, the performance of the metamodel is insensitive to the number of principal components.

From results in Table 4.7 and Table 4.6, we can conclude that the losses due to dimension reduction and emulation of the original model are barely significant for the prediction of the observed traffic flows on the network. The metamodeling approach preserves the essential behavior of the original LCF model for operational applications to the network of Clermont-Ferrand even during long-term simulation of one month. Figure 4.15 is a zoom of Figure 4.1 in Section 4.1. It shows the temporal variation of the spatially-averaged flows over two weeks of loop detector measurements (blue), flows computed by the original LCF model (green) and flows computed by Meta-LCF model (red). Figure 4.16 (left) shows the scatter plot of the spatially-averaged flows computed by Meta-LCF against measured flows at all 15 minute time intervals. The results in Figure 4.1, Figure 4.15 and Figure 4.16 (left) show that both the original LCF model and the Meta-LCF can well predict the temporal variation of the spatially-averaged flows for a long-term simulation.

However, an overestimation during the morning peak hour is observed in the simulation on Monday the 10^{th} November as shown in Figure 4.15. In fact, it is a day between a weekend and a public holiday and some people do not work on this day. Therefore,

Table 4.7: Scores of the one-month simulation during November 2014 against the observations. These evaluate the traffic flows computed by the original model and the metamodel with different sizes for the reduced basis. (The *global* scores are calculated for spatio-temporal flows at all detectors and time steps. The *temporal* scores are calculated for spatially-averaged flows for all time intervals. The *spatial* scores are calculated based on temporally-averaged flows on all detectors.)

Model	Original model	metamodel				
Explained variance of PCA	–	0.80	0.85	**0.90**	0.95	0.98
Global bias (veh h^{-1} detector^{-1})	**18**	7	9	**11**	10	10
Global RMSE (veh h^{-1} detector^{-1})	**154**	140	140	**142**	143	143
Global NRMSE (%)	**84.13**	76.68	76.61	**77.77**	78.00	78.22
Global Correlation	**0.69**	0.72	0.72	**0.71**	0.71	0.71
Temporal bias (veh h^{-1} detector^{-1})	**18**	7	9	**11**	10	10
Temporal RMSE (veh h^{-1} detector^{-1})	**25**	20	20	**23**	23	23
Temporal NRMSE (%)	**14.05**	11.02	11.23	**12.50**	12.62	12.60
Temporal Correlation	**0.99**	0.99	0.99	**0.99**	0.99	0.99
Spatial bias (veh h^{-1} detector^{-1})	**20**	9	10	**13**	12	12
Spatial RMSE (veh h^{-1} detector^{-1})	**116**	106	106	**108**	108	108
Spatial NRMSE (%)	**64.02**	58.90	58.75	**59.89**	59.79	59.93
Spatial Correlation	**0.58**	0.59	0.59	**0.59**	0.59	0.59

the traffic demand is less than a normal Monday during working periods. However, we took the temporal variation of traffic demand from a normal working Monday as input, so that the demand for LCF and Meta-LCF model is bigger than the actual demand. In addition, results show that bigger errors are found in the prediction during public holiday (the 11^{th} November) and weekends (the 15^{th} – 16^{th} and 22^{th} – 23^{th} of November 2014). This is due to a bigger variance of temporal profile of holidays and weekends than that of normal working weekdays. Figure 4.17 shows the temporal profiles of all the Saturdays and Sundays during working period from September 2013 to August 2014. The bias between the average profile which we used as inputs of Meta-LCF and the actual profiles of different weekend are bigger than those of working days as shown in Figure 4.11.

Concerning the spatial features, spatial scores in Table 4.7 and scatter plot in Figure 4.16 (right) show the dispersion of the computed temporally-averaged flows, against measured flows for all detectors. Both the original LCF model and the Meta-LCF has

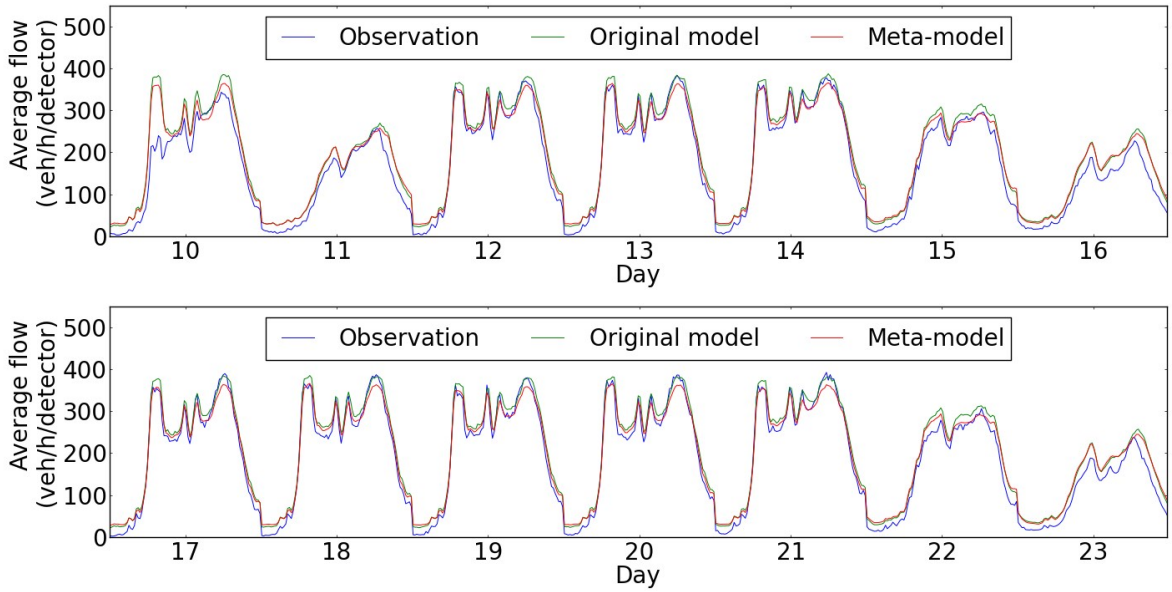


Figure 4.15: Temporal variation of the spatially-averaged flows (in veh h^{-1}) of observation (blue), flows computed by the original model (green) and flows computed by metamodel (red), with a time step of 15 min from the 10^{th} to 23^{th} November 2014 for the agglomeration of Clermont-Ferrand.

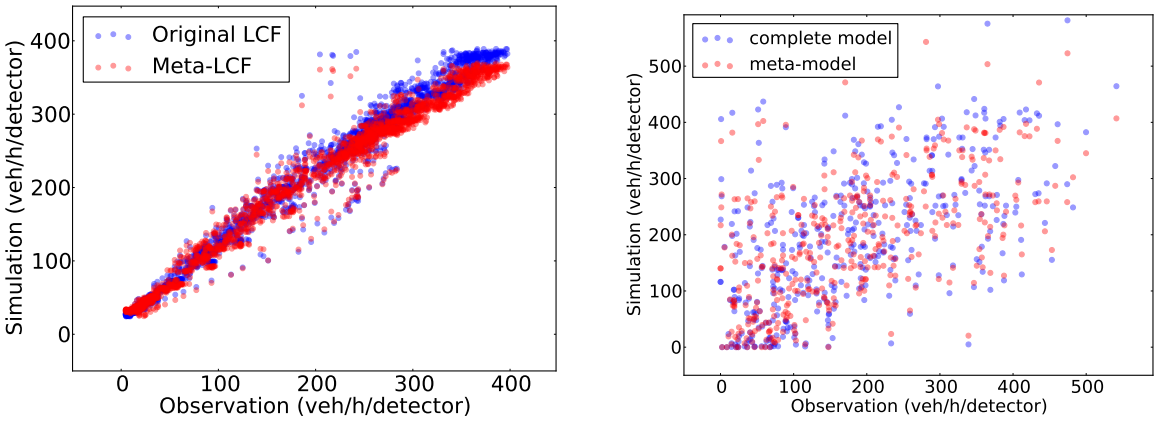


Figure 4.16: Left: dispersion diagram of spatially-averaged flows computed by models (blue for the original one, and red for the metamodel), against measured flows, at each time step. Right: dispersion diagram of temporally-averaged flows computed by models (blue for the original one, red for the metamodel), against measured flows, at each detector.

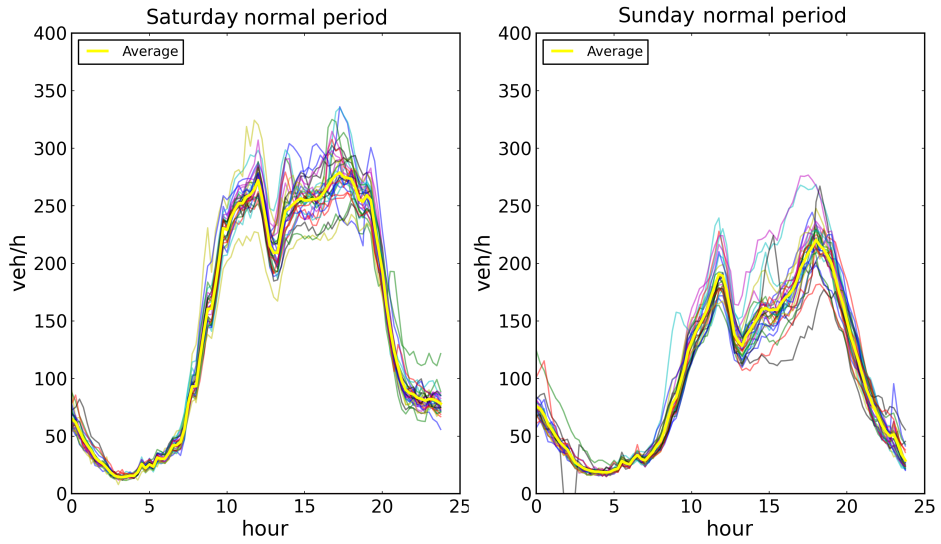


Figure 4.17: Temporal profiles of spatially-averaged flows of all Saturdays (left) and Sundays (right) during normal period from September 2013 to August 2014. Each thin line represents a temporal profile of traffic on each Saturday and Sunday, and the bold yellow line is the average temporal profile of traffic flow over all profiles of Saturdays/Sundays.

limitations in predicting the spatial distribution of the traffic on the network. However, it is reassuring to find that the metamodel can preserve the main behavior of the original model. These results are consistent with those shown in Table 4.7.

Figure 4.18 (left) and Figure 4.19 (left) present the total distribution of the flows computed by the original model and the metamodel, for all detectors and time steps. The distributions are similar for both the original model and metamodel. Differences between simulations and measurements can mostly be found on links with low observed traffic ($< 200\text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$), but there are still some errors on links with heavy flow observed traffic ($> 700\text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$). This limitation for predicting the spatial distribution of the traffic might due to the same reasons as mentioned in Section 4.3.3. Figure 4.18 (right) and Figure 4.19 (right) show the spatio-temporal errors between the computed flows by two models, and the flows measured by detectors. The spatio-temporal errors are almost distributed symmetrically for both models. The mean value of errors between the original model and observations is $\mu_{LCF} = 48\,\text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$. The standard deviation is $\sigma_{LCF} = 173\,\text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$. The spatio-temporal errors between the original LCF model and the observation are mostly distributed in the interval $\mu_{LCF} \pm \sigma_{LCF} = [-125, 221]\ \text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$, for one-month simulation of time step of 15 minutes. For the spatio-temporal errors between the metamodel and observations, the mean value is

$\mu_{Meta} = 40\,\text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$. The standard deviation is $\sigma_{Meta} = 160\,\text{veh}\,\text{h}^{-1}\,\text{lane}^{-1}$. The errors are mostly distributed in the interval $\mu_{Meta} \pm \sigma_{Meta} = [-120, 200]$ veh $\text{h}^{-1}\,\text{lane}^{-1}$.

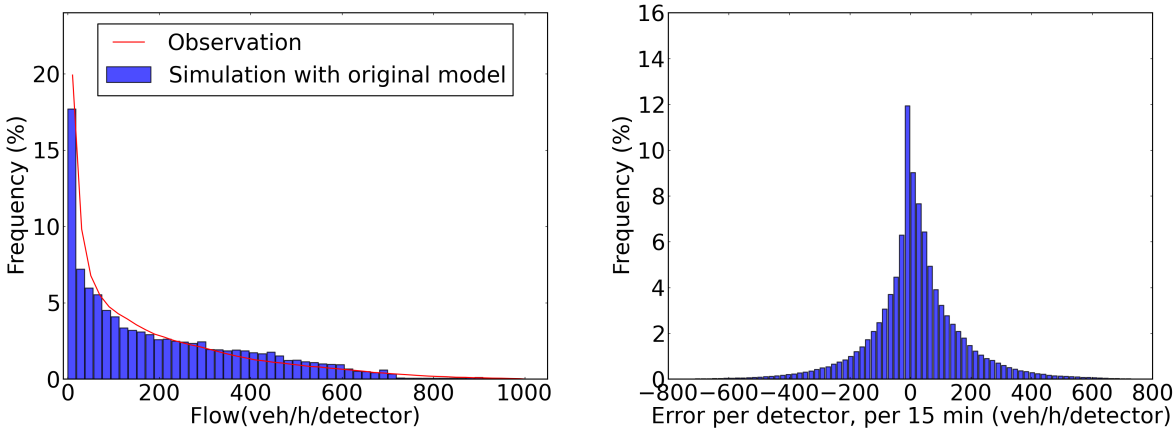


Figure 4.18: Left: the total distribution of the flows computed for each Δh on each detector, by the simulation with **original model**. Right: the distribution of error between the flows computed by original model and flows measured by detectors.

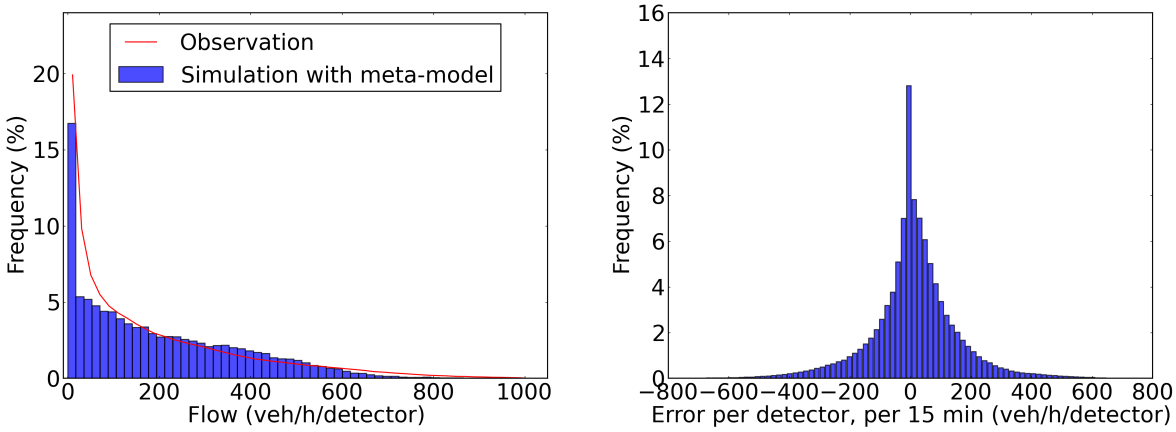


Figure 4.19: Left: the total distribution of the flows computed for each Δh on each detector, by the simulation with **metamodel**. Right: the distribution of error between the flows computed by metamodel and flows measured by detectors.

Conclusions

In this chapter, a metamodeling framework is presented in order to build surrogates of TA models applied to large-scale network. The proposed metamodeling method

is then used in a DTA model (called LADTA) applied in a metropolitan network of Clermont-Ferrand in France. The simulation process was firstly replaced by an *atomic* simulation with LCF (for *LADTA applied in Clermont-Ferrand*) model, whose input is a vector of 23 multiplicative coefficients to perturb the most uncertain inputs. The outputs of LCF are the flows during 15 min on each road of the network in Clermont-Ferrand, and the output dimension is about 19,000. This LCF model was then replaced by a metamodel named Meta-LCF by using the proposed metamodeling method. The computational cost of the resulted Meta-LCF model is very low. The evaluation time of LCF decreases from about 2 hours to 0.022 seconds. The Meta-LCF can reproduce the traffic flows in Clermont-Ferrand with values very close to the computed flows of the original LCF. Time-varying traffic demand for one month was derived from traffic flows measured by detectors. We were able to carry out long-term DUE simulation in Clermont-Ferrand with typical day temporal profiles. The original model and its metamodel have similar performance when compared with traffic flow observations, and the results show that the metamodel preserves well the main features of the original LCF model.

For the metamodeling implementation to the LCF model, we firstly ran 3003×2 training simulations with LCF (3003 before 12:00, 3003 after). The intervals for the inputs variations were then designed in order to take into account all possible variations and input uncertainties. The input samples for the training simulations were generated by Latin hypercube sampling (LHS). With the outputs of the training simulations, we carried out principal component analysis (PCA) to build a reduced basis of LCF outputs. The number of the principal components N was less than 15, and we could keep more than 90 % of the output flow variance. Next, for each principal component, the relations between each of the projection coefficients and the inputs of LCF were reproduced by statistical emulators. The training points for building the emulators were the same as in PCA studies. The emulator for each principal component was composed of a regression part and an interpolation between the residuals of the regression at the training points, by using radial basis functions (RBF). Combining the emulators of the projection coefficients and a reconstruction of the outputs in the reduced space, we were able to construct a metamodel, called Meta-LCF, which is a surrogate for the original LCF.

We compared the outputs of Meta-LCF and the original LCF model for simulating traffic flow on each link of the agglomeration of Clermont-Ferrand during the one-month simulation of November 2014. The computational time for a one-month simulation decreases from 48 hours using 110 cores, to less than 1 minute on one core. The flows

computed at link level during each 15 minutes time step show a good correlation of 0.98 between the metamodel and the original LCF model. Moreover, the metamodel can well reproduce the temporal variation of the spatially-averaged flows. The temporal NRMSE is about 4.5 %. The performance of spatial distribution of the traffic flows computed by Meta-LCF is not as good as the temporal performance. The spatial NRMSE is about 23.4 %, but it is found that the main errors appeared on links where the flows computed by LCF are low ($< 200\,\mathrm{veh\,h^{-1}}$). The errors between the metamodel and the original model could be decreased by increasing the total number of samples. The performance of the metamodel could also be improved by better designing the *distance function* $d_j(\cdot,\cdot)$ of RBF for the interpolation part of emulation. However, despite the spatial errors, the Meta-LCF built in this work is able to preserve the main features of the original LCF. We compared the one-month simulation outputs of the two models with the traffic flows measured by detectors on the network. The scores are similar for both models. They can reproduce the temporal variation of the spatially-averaged traffic flows over the whole simulation period, with good correlation (0.99) and low NRMSE (about 12.5 %) during the one-month simulation. There are still spatial errors with NRMSE bigger than 50 %. The main spatial errors might due to representativeness errors, the uncertainty in the O-D matrix and the road choice assumptions.

One of the main purposes to build the metamodel of a DTA model is to facilitate the variance-based global sensitivity analysis (such as Sobol' method) and uncertainty quantification of a road traffic emission model, in the process of air quality simulation at the metropolitan scale, during long-term periods. In fact, in the fields of air quality modeling, the on-road emissions are among the most important inputs and show large uncertainties. To predict traffic emissions, the traffic flows and speeds during long-term periods are needed as inputs of emission models. Concerning uncertainty quantification, even though the LADTA model is a deterministic model, the inputs of the model is often with big uncertainties. The errors, especially spatial error of LADTA when compared with traffic flow observations are inevitable as shown in [Chen et al., 2017]. The objective of uncertainty analysis is to generate an ensemble of LADTA simulations with input perturbations using Monte-Carlo method, and then to evaluate the uncertainty of the inputs by analyzing if the traffic flow observations during a long-period time can be well predicted (statistically) by the generated ensemble. The computational cost for applying DTA models is often high and to generate an ensemble of DTA simulations is challenging. The metamodeling methods make this study possible by largely reducing the evaluation time of a DTA simulation yet preserving main behaviors of the original DTA models. In our case study with the LCF model, with the same inputs of the original LCF model,

the resulted Meta-LCF can reproduce traffic flows at street level of the given network and the outputs of Meta-LCF model are of the same performance as those of the original one when compared with traffic flow observations. This encouraging results give new insights to carry out a large number of LCF simulations with very low computational cost, making quantitative sensibility analysis and uncertainty quantification possible. The quantitative global sensitivity analysis (GSA) on the DTA model is presented later in Chapter 5. The uncertainty quantification of the computed traffic flows from DTA simulation is presented later in Chapter 7. In addition, since the Meta-LCF allows us to compute the emissions of road traffic during a long-term period with low computational costs, we can build an on-road traffic emission model combining the Meta-LCF and the emission factor model COPERT IV, for the same agglomeration. Therefore, the GSA study and uncertainly quantification of the emission simulation chain Meta-LCF-COPERT can be carried out, which will be presented in Chapter 6 and Chapter 7.

In future works, more inputs could be added to LCF model, in order to take into account more uncertainty sources. (i) On the demand side, we assumed that users of all O-D pairs share the same temporal profile of the departure time. Different profiles of departure times according to different categories of O-D pairs can be designed to investigate the influence of the departure-time choice. Moreover, in order to represent the spatial uncertainty lying in the O-D matrix, more categories of O-D pairs could be considered. (ii) On the supply side, more categories of links can be added according to their capacities, speed limits and spatial positions as well. These designs of supply inputs would allow us to take into account the spatial variations of the network features. In addition, with the help of the built emission simulation chain of very low computational cost, the uncertainties lying in the traffic assignment model and emission model can be propagated through the whole air quality simulation chain, from the traffic assignment to air quality simulations. It is then possible to carry out the uncertainty quantification of the air quality simulations

Chapter 5

Global sensitivity analysis for a dynamic traffic assignment model at metropolitan scale

Summary

In this chapter, a global sensitivity analysis (GSA) is carried out on the dynamic traffic assignment (DTA) simulation with LADTA model applied to the agglomeration of Clermont-Ferrand. The Sobol' method is used. It is a variance-based GSA method aiming at variance decomposition. For an input p_i, two indices are used to measure the effect of this input on one output: the *first-order* and *total-effect* sensitivity indices. The *first-order* index reflects the main effect of p_i. The *total-effect* index reflects the higher-order interactions involving p_i and the other parameters. These two indices are computed for the traffic flows and average speeds at street resolution, with respect to a number of uncertain inputs of DTA simulations in the agglomeration of Clermont-Ferrand with LADTA model. GSA results show that the computed traffic flows are very sensitive to the direction of traffic demands in the Origin-Destination (O-D) matrix. They are also sensitive to the volume of traffic demands, especially: (i) the temporal variation of traffic demand during the previous 0.5 hour before the target simulation time and (ii) the traffic demands between O-D pairs for which the inter-zone distance is between 0 - 5 km. The computed flows are not very sensitive to the link capacity or speed limit in our case study. For the GSA study on the computed average speeds in the agglomeration, they are sensitive to (i) the direction of traffic demand, and (ii) the uncertain inputs corresponding to street speed limits. The computed speeds are not significantly sensitive to other inputs on most of the streets in our case study. However, there are still some streets on which the traffic speeds are sensitive to the inputs that are influential to the computed traffic flows. The spatial distributions of output sensitivity with respect to uncertain inputs is also illustrated by maps of sensitivity indices. These maps allow us to visualize how the uncertain inputs can

influence traffic flows and travel speeds throughout the whole network at street resolution.

Résumé

On s'intéresse dans ce chapitre à l'analyse de sensibilité globale (*global sensitivity analysis*, GSA) de l'affectation dynamique du trafic avec le modèle LADTA, dans l'agglomération de Clermont-Ferrand. La méthode de Sobol' est utilisée. On calcule deux types d'indice de sensibilité: (i) l'indice de Sobol du premier ordre et (ii) l'indice total. Les indices sont calculés pour chaque entrée du modèle, à la résolution de la rue sur l'ensemble de l'agglomération, pour étudier la sensibilité du débit et de la vitesse simulés par le modèle LADTA. Les résultats de l'analyse montrent que les débits du trafic simulés sont très sensibles à la direction dans la demande de trafic décrite par la matrice d'Origine-Destination (matrice O-D). Ils sont également sensibles aux demandes totales et plus précisément, (i) aux demandes de trafic pendant les 30 minutes précédant l'heure de simulation et (ii) aux demandes entre les paires O-D dont la distance inter-zone est de 0 à 5 km. Dans notre cas d'étude pour l'agglomération de Clermont-Ferrand, les débits simulés ne sont très sensibles ni aux capacités ni aux vitesses maximales autorisées sur le réseau. Pour les vitesses moyennes simulées par LADTA, elles sont sensibles (i) à la direction de la demande de trafic, et (ii) aux vitesses maximales autorisées sur le réseau de l'agglomération. Dans la plupart des rues, les vitesses simulées sont peu sensibles aux autres entrées. Cependant, il existe des rues où les vitesses du trafic sont sensibles aux entrées qui sont influentes sur le débit simulé par LADTA. La distribution spatiale des indices de sensibilité est illustrée par des cartes, qui nous permettent de visualiser comment les entrées incertaines peuvent influencer les débits et vitesses simulés, à la résolution de la rue sur toute l'agglomération.

Contents

5.1 Introduction

This chapter introduces the definition of quantitative global sensitivity analysis based on Sobol' method [Sobol, 1993] and the computation of global sensitivity indices. Then the method is applied to the dynamic traffic assignment (DTA) in the agglomeration of Clermont-Ferrand using the metamodel built in Chapter 4.

5.2 Global sensitivity analysis with Sobol' method

5.2.1 Definition of global sensitivity analysis

The sensitivity analysis aims to "study how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" [Saltelli et al., 2004, 2008]. Sobol' method is one of the variance-based methods for global sensitivity analysis (GSA). The variance-based measures are useful to explore the model over different combinations of values for the uncertain inputs [Saltelli et al., 2008]. They show how the variance of the output depends on the uncertain input factors, and can then be decomposed accordingly. A model is denoted by $y = g(p_1, p_2, \ldots, p_i, \ldots, p_K)$. The decomposition can be applied to the expectation and the variance of the model output y. This allows us to analyze the first-order effect of the factor p_i, and the higher-order interaction effects of combination of factors [Sobol, 2001]. The first-order effect expresses what would happen to the uncertainty of y if we could fix a factor $p_i = p_i^*$, where p_i^* is a particular value. $Var_{\boldsymbol{p}_{\sim i}}(y \mid p_i = p_i^*)$ is the resulting variance of y, taken over $\boldsymbol{p}_{\sim i}$ (all factors but p_i). By taking the average of the measure $Var_{\boldsymbol{p}_{\sim i}}(y \mid p_i = p_i^*)$ over all possible points p_i^*, the dependence of this measure on p_i^* disappears. This average is denoted by $E_{p_i}(Var_{\boldsymbol{p}_{\sim i}}(y \mid p_i))$. The unconditional variance of y can be decomposed into (i) the main effect of p_i and (ii) a residual with respect to the other factors:

$$Var(y) = Var_{p_i}(E_{\boldsymbol{p}_{\sim i}}(y \mid p_i = p_i^*)) + E_{p_i}(Var_{\boldsymbol{p}_{\sim i}}(y \mid p_i = p_i^*)). \tag{5.1}$$

A small $E_{p_i}(Var_{\boldsymbol{p}_{\sim i}}(y \mid p_i = p_i^*))$, or a large $Var_{p_i}(E_{\boldsymbol{p}_{\sim i}}(y \mid p_i = p_i^*))$, will imply that p_i is an influential factor. The conditional variance $Var_{p_i}(E_{\boldsymbol{p}_{\sim i}}(y \mid p_i = p_i^*))$ is called the first-order effect of p_i on y, and the sensitivity measure

$$S_i = \frac{Var_{p_i}(E_{\boldsymbol{p}_{\sim i}}(y \mid p_i = p_i^*))}{Var(y)} \tag{5.2}$$

is the first-order sensitivity index of p_i on y. A high value of S_i implies that p_i is an

influential factor. However, a small value of S_i does not naturally refer a non-influential variable, since the first-order index ignores the interaction effects of factor combinations within p_i. That is why we need the total-effect sensitivity index S_{Ti}. S_{Ti} estimates the total contribution to the output variation due to the factor p_i, i.e., its first-order effect plus all higher-order effects due to interactions with p_i. The condition $S_{Ti} = 0$ is necessary and sufficient for p_i to be a non-influential factor [Saltelli et al., 2008]. This means that the uncertainty of p_i would not significantly affect the output variance $Var(y)$. In order to estimate the total effect of p_i, we can condition on $\boldsymbol{p}_{\sim i}$ except p_i and $Var(y)$ can be decomposed as :

$$Var(y) = E_{\boldsymbol{p}_{\sim i}}(Var_{p_i}(y \mid \boldsymbol{p}_{\sim i} = \boldsymbol{p}^*_{\sim i})) + Var_{\boldsymbol{p}_{\sim i}}(E_{p_i}(y \mid \boldsymbol{p}_{\sim i} = \boldsymbol{p}^*_{\sim i})). \tag{5.3}$$

The measure $E_{\boldsymbol{p}_{\sim i}}(Var_{p_i}(y \mid \boldsymbol{p}_{\sim i} = \boldsymbol{p}^*_{\sim i})) = Var(y) - Var_{\boldsymbol{p}_{\sim i}}(E_{p_i}(y \mid \boldsymbol{p}_{\sim i} = \boldsymbol{p}^*_{\sim i}))$ is the remaining variance of y that would be left on average if we could determine the "true" values of $\boldsymbol{p}_{\sim i}$. This remaining variance is then due to not conditioning the factor p_i, and it can be considered as the total effect of p_i. $E_{\boldsymbol{p}_{\sim i}}(Var_{p_i}(y \mid \boldsymbol{p}_{\sim i} = \boldsymbol{p}^*_{\sim i}))$ is calculated over all possible combinations of $\boldsymbol{p}_{\sim i}$ since $\boldsymbol{p}_{\sim i}$ are uncertain factors and their "true" values are unknown. Therefore, the total effect index for p_i is obtained after dividing the measure $E_{\boldsymbol{p}_{\sim i}}(Var_{p_i}(y \mid \boldsymbol{p}_{\sim i} = \boldsymbol{p}^*_{\sim i}))$ by $Var(y)$:

$$S_{Ti} = \frac{E_{\boldsymbol{p}_{\sim i}}(Var_{p_i}(y \mid \boldsymbol{p}_{\sim i} = \boldsymbol{p}^*_{\sim i}))}{Var(y)} = 1 - \frac{Var_{\boldsymbol{p}_{\sim i}}(E_{p_i}(y \mid \boldsymbol{p}_{\sim i} = \boldsymbol{p}^*_{\sim i}))}{Var(y)}. \tag{5.4}$$

5.2.2 Computation of first-order and total-effect Sobol' indices

$Var(E(y \mid p_i))$ in Equation 5.2 can be calculated as

$$Var(E(y \mid p_i)) = \int E^2(y \mid p_i = \tilde{p}_i)\rho(\tilde{p}_i)d\tilde{p}_i - E^2(y) = U_i - E^2(y), \tag{5.5}$$

where $\rho(\tilde{p}_i)$ is the probability that $p_i = \tilde{p}_i$. Saltelli [2002] proposes an efficient strategy for computing S_i and S_{Ti} by using Monte Carlo estimates. The sample size used for the Monte Carlo estimate is n, and there are K uncertain inputs. Two input sample matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are generated:

$$\boldsymbol{M}_1 = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1K} \\ p_{21} & p_{22} & \dots & p_{2K} \\ \dots & & & \\ p_{n1} & p_{n2} & \dots & p_{nK} \end{pmatrix}, \qquad \boldsymbol{M}_2 = \begin{pmatrix} p'_{11} & p'_{12} & \dots & p'_{1K} \\ p'_{21} & p'_{22} & \dots & p'_{2K} \\ \dots & & & \\ p'_{n1} & p'_{n2} & \dots & p'_{nK} \end{pmatrix} \tag{5.6}$$

Saltelli [2002] proposes estimators $\widehat{Var}(y)$, $\widehat{E}^2(y)$, $\widehat{U}_i$ and $\widehat{U}_{\sim i}$ for computing the first-order and total-effect indices presented in Equation (5.2) and Equation (5.4):

$$S_i \simeq \frac{\widehat{U}_i - \widehat{E}^2(y)}{\widehat{Var}(y)} \ , \tag{5.7}$$

$$S_{Ti} \simeq \frac{\widehat{U}_{\sim i} - \widehat{E}^2(y)}{\widehat{Var}(y)}. \tag{5.8}$$

$\widehat{Var}(y)$, $\widehat{E}^2(y)$, $\widehat{U}_i$ and $\widehat{U}_{\sim i}$ can be obtained by Monte Carlo estimates from values of y computed on the sample in $\boldsymbol{M}_1$, $\boldsymbol{M}_2$, and $\boldsymbol{N}_i$, $\boldsymbol{N}_{\sim i}$, defined as [Saltelli, 2002; Homma and Saltelli, 1996]

$$\boldsymbol{N}_i = \begin{pmatrix} p'_{11} & p'_{12} & \dots & p_{1i} & \dots & p'_{1K} \\ p'_{21} & p'_{22} & \dots & p_{2i} & \dots & p'_{2K} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ p'_{n1} & p'_{n2} & \dots & p_{ni} & \dots & p'_{nK} \end{pmatrix} \qquad \boldsymbol{N}_{\sim i} = \begin{pmatrix} p_{11} & p_{12} & \dots & p'_{1i} & \dots & p_{1K} \\ p_{21} & p_{22} & \dots & p'_{2i} & \dots & p_{2K} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p'_{ni} & \dots & p_{nK} \end{pmatrix} \tag{5.9}$$

Since both $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are generated randomly, they are considered to be independent of each other. If we consider $\boldsymbol{M}_1$ as the "sample" matrix and $\boldsymbol{M}_2$ as the "re-sample" matrix, in $\boldsymbol{N}_i$, all the values are "re-sampled" except for the column i for the factor p_i. $\widehat{U}_i$ can be obtained from values of y computed on matrices $\boldsymbol{M}_1$ and $\boldsymbol{N}_i$:

$$\widehat{U}_i = \frac{1}{n-1} \sum_{r=1}^{n} g(p_{r1}, p_{r2}, \dots, p_{rk}) g(p'_{r1}, p'_{r2}, \dots, p'_{r(i-1)}, p_{ri}, p'_{r(i+1)}, \dots, p'_{rK}). \tag{5.10}$$

The matrix $\boldsymbol{N}_{\sim i}$ in Equation (5.9) can be considered as $\boldsymbol{M}_1$ with only the factor p_i re-sampled. Therefore, $\widehat{U}_{\sim i}$ can be obtained from values of y computed on matrices $\boldsymbol{M}_1$ and $\boldsymbol{N}_{\sim i}$:

$$\widehat{U}_{\sim i} = \frac{1}{n-1} \sum_{r=1}^{n} g(p_{r1}, p_{r2}, \dots, p_{rk}) g(p_{r1}, p_{r2}, \dots, p_{r(i-1)}, p'_{ri}, p_{r(i+1)}, \dots, p_{rK}). \tag{5.11}$$

$\widehat{Var}(y)$ is computed from $\boldsymbol{M}_1$. For the estimator of $E^2(y)$ in Equation (5.7) and (5.8), Homma and Saltelli [1996] propose that better estimates for the first-order terms are obtained if $E^2(y)$ is estimated from values of y computed on both $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ matrices:

$$\widehat{E}^2 = \frac{1}{n} \sum_{r=1}^{n} g(p_{r1}, p_{r2}, \dots, p_{rk}) g(p'_{r1}, p'_{r2}, \dots, p'_{rK}). \tag{5.12}$$

In summary, for computing S_i and S_{Ti}, Saltelli [2002] proposes an efficient strategy by using Monte Carlo estimates. The sample size used for the Monte Carlo estimate is n, and there are K uncertain inputs. The total number required of model evaluations is $n(2K+2)$: nK for the first-order terms, nK for the total-effect terms, and $2n$ for $\widehat{E}(y)$ [Saltelli et al., 2000]. Saltelli et al. [2008] recommend that for carrying out GSA with Sobol' method, it is better to have a lower dimension model in inputs and outputs. The recommended CPU time per run of the model should be less than 1 min. However, for simulating traffic flows and emissions of an urban area at street level, the computational cost is much more higher than recommended. Also, the inputs of classic traffic assignment models are of very high dimension. To reach similar conditions as recommended carried out the GSA with Sobol' method for the dynamic traffic assignment model built in Chapter 4.

5.3 Application to the dynamic traffic assignment model for the agglomeration of Clermont-Ferrand

At the metropolitan scale, traffic assignment (TA) models are used to predict the traffic flows at street level. The inputs of TA models are traffic demand and a modeled network. The network is modeled with origins/destinations, nodes and links. The link capacity constraints and speed limits are given at street level. In the following parts, a "link" represents a "street" or "road" of a city road network. The traffic demand is represented by an Origin-Destination matrix (O-D matrix), summarizing the traffic flow from each origin zone to each destination zone during a certain period. A pair of Origin-Destination zones is refered to an O-D pair. The most commonly used TA models are based on the Wardrop User Equilibrium (UE) principle [Wardrop, 1952], stating that users choose the least cost path to travel through the network from an origin to a destination. There are static TA models and dynamic TA models. In this paper, we used a dynamic TA model in order to take into account time-dependent inputs. The model used here is called LADTA [Leurent, 2003; Aguiléra and Leurent, 2009]. It is applicable to large-size network [Aguiléra and Leurent, 2009]. LADTA model is computationally costly. A meta-modeling approach was therefore applied to LADTA model in the agglomeration of Clermont-Ferrand in Chapter 4. The outputs of the resulting meta-model is very close to those of the complete model, yet the computation time is significantly decreased. Then GSA studies are carried out on this meta-model instead of the complete LADTA model.

5.3.1 Metamodel of LADTA applied to the agglomeration of Clermont-Ferrand

We recall here the modeled network of the agglomeration of Clermont-Ferrand in Figure 5.1. There are more than 19,000 links on the network. For each link, the computed traffic flow is a scalar so that we can compute sensitivity indices at link resolution.

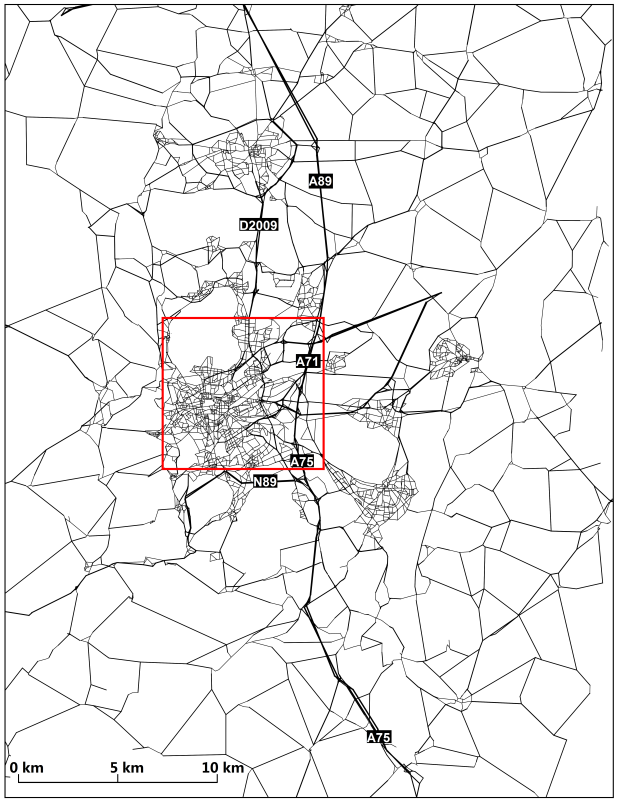


Figure 5.1: Domain of the agglomeration of Clermont-Ferrand. The red rectangle delimits the *city of Clermont-Ferrand*. A89, A71 and A75 are three highways in this agglomeration. They cross the city of Clermont-Ferrand as well. D2009 and N89 are two national roads, connecting the city and the suburbs.

The metamodel we use in this chapter is the one built in Chapter 4. The metamodel is built based on *LADTA model applied to the agglomeration of Clermont-Ferrand* (LCF model). The metamodel is called Meta-LCF model. We recall here the inputs, outputs and equations of the LCF model and the Meta-LCF model.

Inputs

$\mathcal{G} = (\mathcal{N}, \mathcal{A})$: the nodes ($\in \mathcal{N}$) and links ($\in \mathcal{A}$) of the oriented graph $\mathcal{G}$ modeled for the network of the agglomeration of Clermont-Ferrand. $D = card(\mathcal{A}) = 19628$. D is the number of links.

$\mathcal{Z}_O, \mathcal{Z}_D \subset \mathcal{N}$: the set of nodes representing the Origin-Destination zones. $z = card(\mathcal{Z}_O) = card(\mathcal{Z}_D) = 124$.

$\boldsymbol{C} = [C_a]_{a\in\mathcal{A}}$, $\boldsymbol{L} = [L_a]_{a\in\mathcal{A}}$, $\boldsymbol{V_0} = [V_{0a}]_{a\in\mathcal{A}}$: the vector of link capacities, link lengths and speed limits. $\boldsymbol{C}, \boldsymbol{L}, \boldsymbol{V_0} \in \mathbb{R}^D$.

$\boldsymbol{T_0} = [T_{0a}]_{a\in\mathcal{A}}$: a vector of free-flow travel time. $\boldsymbol{T_0} \in \mathbb{R}^D$.

In order to represent the uncertainty of the network parameters, links are regrouped according to link capacities and speed limits.

$\boldsymbol{V_0^{low}} = [V_{0a}]_{a\in\mathcal{A}_{low}}$: the vector of low speed limits ($V_0 \leq 50\,\mathrm{km\,h^{-1}}$), with $\mathcal{A}_{low}$ as the set of links with $V_0 \leq 50\,\mathrm{km\,h^{-1}}$. Links in the set $\mathcal{A}_{low}$ are considered "low-speed links".

$\boldsymbol{V_0^{high}} = [V_{0a}]_{a\in\mathcal{A}_{high}}$ with $\mathcal{A}_{high} = \mathcal{A}\backslash\mathcal{A}_{low}$: the vector of high speed limits. Links in the set $\mathcal{A}_{high}$ are considered "high-speed links".

$\boldsymbol{C^{small}} = [C_a]_{a\in\mathcal{A}_{small}}$: the vector of small link capacities, with $\mathcal{A}_{small}$ as the set of links with $C \leq 900\,\mathrm{veh\,h^{-1}}$.

$\boldsymbol{C^{big}} = [C_a]_{a\in\mathcal{A}_{big}}$ with $\mathcal{A}_{big} = \mathcal{A}\backslash\mathcal{A}_{small}$: the vector of big link capacities.

$\mathcal{H} = [h_0, h_1]$: the simulation period of LADTA.

Δh: time step for loop detectors to register aggregated traffic flow data. It is also assumed that during the interval $[h, h + \Delta h]$, the traffic flow on the network is constant, and so is the time-varying traffic demand. In our case study, $\Delta h = 0.25$ hour.

$\boldsymbol{Q}(h) = (Q_{o,d}(h))_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D}$: the time-varying O-D matrix. $Q_{o,d}(h)$ is the cumulated traffic demand (total number of vehicles in veh) from o to d during $\mathcal{H}$. It is non-decreasing and can be deduced from time-varying traffic demand *density* $q_{o,d}(h)$ (in $\mathrm{veh\,h^{-1}}$) at instant h for the O-D pair od. $q_{o,d}(h)$ is assumed to be a piece-wise linear function of h and $q_{o,d}(h)$ is constant in the interval $[h, h + \Delta h]$. With known $q_{o,d}(h)$ and bounded $\mathcal{H}$, $(Q_{o,d}(h))_{(o,d)} = \int_{h_0}^{h} q_{o,d}(\tilde{h})d\tilde{h}$ with $h \in \mathcal{H}$.

$\boldsymbol{Q}^{\mathrm{peak}} = (q_{o,d}^{\mathrm{peak}})_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D}$: O-D matrix of the agglomeration of Clermont-Ferrand during the evening peak hour 17:00 – 18:00, where $(q_{o,d}^{\mathrm{peak}})_{(o,d)}$ is the average traffic demand during the evening peak. It is the static O-D matrix.

Outputs

$\boldsymbol{X}(h) = [X_a(h)]_{a\in\mathcal{A}}$, $\boldsymbol{T}(h) = [T_a(h)]_{a\in\mathcal{A},h\in\mathcal{H}}$: the vector of cumulated traffic volumes at time $h \in \mathcal{H}$ computed by LADTA.

$\boldsymbol{y} = \boldsymbol{X}(h+\Delta h) - \boldsymbol{X}(h)$ (in veh h^{-1}): a vector of average traffic flow during $[h, h+\Delta h]$. It is the final output of the LADTA model.

LCF model: LADTA applied in Clermont-Ferrand

The LADTA model takes into account the influence of the time-varying traffic demand to compute time-dependent $\boldsymbol{X}(h)$. Unlike in a static assignment model, $\boldsymbol{X}(h)$ computed by LADTA varies continuously as a function of h. The longer $\mathcal{H}$ is, the more traffic demand inputs are required in order to represent time-dependent $Q_{o,d}(h)$. In addition, the longer $\mathcal{H}$ is, the longer computation time is required since LADTA computes dynamic equilibrium at each instant $h \in \mathcal{H}$. Therefore, an *atomic* LADTA simulation for the agglomeration of Clermont-Ferrand is defined in Chapter 4. It allows us to reduce input dimension and computational cost, yet preserve the time-varying influence of traffic demand before the target time h_{simu} on the computed flows at h_{simu}. This *atomic* LADTA simulation is called LCF, for ***L**ADTA applied to **C**lermont-**F**errand*. In the LCF model, the simulation period is decreased to $\mathcal{H}_{atomic} = [h_{simu} - 2.25, h_{simu} + 0.75]$. $\mathcal{H}_{atomic}$ is chosen based on the assumption that the traffic flow at h_{simu} is not affected by traffic demands before 2.25 hours and after 1.0 hour for the agglomeration of Clermont-Ferrand. LCF model still takes cumulated traffic demand $(Q_{o,d}(h))_{h\in\mathcal{H}_{atomic}}$ as input and computes $\boldsymbol{X}(h), h \in \mathcal{H}_{atomic}$. However, we focus only on the computed traffic flow at time h_{simu} and consider the computed $\boldsymbol{y}(h_{\text{simu}})$ as the output of one DTA simulation with LCF model.

$(Q_{o,d}(h))_{h\in\mathcal{H}_{atomic}}$ can be obtained if we know the *density* of traffic demand on the network during $\mathcal{H}_{atomic}$: $(q_{o,d}(h))_{h\in\mathcal{H}_{atomic}}$. $(q_{o,d}(h))_{h\in\mathcal{H}_{atomic}}$ is assumed to be piece-wise linear and constant on the interval $[h, h+\Delta h], \forall h \in \mathcal{H}_{atomic}$. With $\Delta h = 0.25$, we define the set of instants as $I_{\text{atomic}} = \{h_{\text{simu}} - 2.25, h_{\text{simu}} - 2.0, \ldots, h_{\text{simu}} - 0.25, h_{\text{simu}}, \ldots, h_{\text{simu}} + 0.75\}$. Then $q_{o,d}(h)_{h\in\mathcal{H}_{\text{atomic}}}$ can be obtained if we know $q_{o,d}(h)_{h\in I_{\text{atomic}}}$. In our case study, the static O-D matrix $\boldsymbol{Q}^{\text{peak}}$ is given. Therefore, based on this static matrix, $q_{o,d}(h)_{h\in I_{\text{atomic}}}$ can be obtained by $q_{o,d}(h)_{h\in I_{\text{atomic}}} = P(h)_{h\in I_{\text{atomic}}} \times q_{o,d}^{\text{peak}}$, where $P(h)_{h\in I_{\text{atomic}}}$ are *temporal variation coefficients* representing the ratio between traffic demand (i) during $[h, h+\Delta h], \forall h \in I_{\text{atomic}}$ and (ii) during the evening peak hour. It is also assumed that the temporal variation of the demand is independent of the spatial distribution of O-D pairs: at the same time h, $P(h)$ is the same for all $(q_{o,d}(h))_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D, h\in\mathcal{H}}$. The dimension of I_{atomic} is 13. Therefore, 13 coefficients $P(h)_{h\in I_{\text{atomic}}}$ can define the temporal-variation of traffic demand and build the input dynamic O-D matrix for carrying out DTA with LCF model during $\mathcal{H}_{atomic}$.

The O-D pairs are categorized into five groups according to their distance between

Origin-Destination zones: 0 km, $0-5$ km, $5-10$ km, $10-15$ km and >15 km. The matrix $\boldsymbol{Q}^{\text{peak}}$ can be divided by 5 matrices and we have: $\boldsymbol{Q}^{\text{peak}} = \boldsymbol{Q}_0^{\text{peak}} + \boldsymbol{Q}_{0-5}^{\text{peak}} + \boldsymbol{Q}_{5-10}^{\text{peak}} + \boldsymbol{Q}_{10-15}^{\text{peak}} + \boldsymbol{Q}_{>15}^{\text{peak}}$. With all the inputs mentioned in Section 5.3.1, the temporal-variation coefficients $P(h)$ the LCF model can then be represented as

$$\begin{aligned}\boldsymbol{y}(h_{\text{simu}}) &= \mathcal{F}(\boldsymbol{Q}(h)_{h\in\mathcal{H}_{\text{atomic}}}, \mathcal{G}) \\ &= \mathcal{F}(\boldsymbol{Q}(h)_{h\in\mathcal{H}_{\text{atomic}}}, \boldsymbol{V}_0, \boldsymbol{C}) \\ &= \mathcal{F}(P(h)_{h\in I_{\text{atomic}}}, \delta_0\boldsymbol{Q}_0^{\text{peak}}, \delta_{0-5}\boldsymbol{Q}_{0-5}^{\text{peak}}, \delta_{5-10}\boldsymbol{Q}_{5-10}^{\text{peak}}, \delta_{10-15}\boldsymbol{Q}_{10-15}^{\text{peak}}, \delta_{>15}\boldsymbol{Q}_{>15}^{\text{peak}}, \\ &\qquad \mu_{low}\boldsymbol{V}_0^{low}, \mu_{high}\boldsymbol{V}_0^{high}, \lambda_{small}\boldsymbol{C}^{small}, \lambda_{big}\boldsymbol{C}^{big}).\end{aligned} \tag{5.13}$$

In Equation (5.13), $P(h)$ represents the temporal variation of the demand and is perturbed to account for part of the uncertainty in O-D matrix input. δ_0, δ_{0-5}, δ_{5-10}, δ_{10-15} and $\delta_{>15}$ are five *evening peak coefficients* to represent spatial uncertainty of the demand in the static O-D matrix for evening peak hour. They can be perturbed to account for spatial uncertainty lying in the static O-D matrix. μ_{low}, μ_{high}, λ_{small}, and λ_{big} are four coefficients applied to link capacities and speed limits. They can be perturbed to account for uncertainties in link capacities and speed limits. In addition, the direction of the traffic demand is mainly from working zones to residence zones in the given $\boldsymbol{Q}^{\text{peak}}$ since it contains traffic demands during the evening rush hour. The direction of the demand during the morning might be opposite. During morning period (before 12:00), we use $(\boldsymbol{Q}^{peak})^T$ as O-D matrix. The input for the transposition is a binary parameter η, with $\eta = 0$ when the simulation is carried out with the normal O-D matrix $\boldsymbol{Q}^{peak}$, and $\eta = 1$ when we use the transposed O-D matrix $(\boldsymbol{Q}^{peak})^T$ instead. The LCF model can then be represented by the multiplicative coefficients corresponding to its inputs. The LCF model represented by multiplicative coefficients is written as

$$\boldsymbol{y}(h_{\text{simu}}) = \mathcal{M}(P(h)_{h\in I_{\text{atomic}}}, \delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15}, \lambda_{big}, \lambda_{small}, \mu_{high}, \mu_{low}, \eta) = \mathcal{M}(\boldsymbol{p}), \tag{5.14}$$

where $\boldsymbol{p} \in \mathbb{R}^K$ is the input vector of the LCF model and its dimension is $K = 23$. The objective of global sensitivity analysis is to analyze the sensitivity of computed traffic flows and average speeds from the metamodel, with respect to the 23 inputs representing the uncertainty in the LCF model.

One *atomic* simulation with LCF model takes about 2 hours. After applying metamodeling methods in Chapter 4, the final metamodel of the LCF model is called Meta-LCF model and it is written as

$$\widehat{\boldsymbol{y}} = \bar{\boldsymbol{y}} + \boldsymbol{\Psi}(\widehat{\boldsymbol{f}}(\boldsymbol{p}) - \boldsymbol{\Psi}^T\bar{\boldsymbol{y}}) = \bar{\boldsymbol{y}} - \boldsymbol{\Psi}\boldsymbol{\Psi}^T\bar{\boldsymbol{y}} + \boldsymbol{\Psi}\begin{bmatrix}\widehat{f}_1(\boldsymbol{p})\\ \widehat{f}_2(\boldsymbol{p})\\ \vdots\\ \widehat{f}_N(\boldsymbol{p})\end{bmatrix} \simeq \boldsymbol{y}, \tag{5.15}$$

where $\bar{\boldsymbol{y}}$ and $\boldsymbol{\Psi}\boldsymbol{\Psi}^T$ are already known. When the model is applied to a new input vector $\boldsymbol{p}$, the computational cost is that of each $\widehat{f}_j(\boldsymbol{p})$ with $(j = 1, \ldots, N)$. With the same input vector $\boldsymbol{p}$, the outputs $\widehat{\boldsymbol{y}}$ are also average traffic flows during $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$ at street level, and $\widehat{\boldsymbol{y}}$ is close to $\boldsymbol{y}$.

For the metamodeling process in our case study, the training size m is 3003 for each case using $\boldsymbol{Q}$ or using the its transposition $\boldsymbol{Q}^T$ as input O-D matrix. Once the training set is get, the $\bar{\boldsymbol{y}}$ and $\boldsymbol{\Psi}\boldsymbol{\Psi}^T$ are obtained from the PCA study mentioned in Chapter 4. The resulting Meta-LCF model takes only 0.022 second to compute $\widehat{\boldsymbol{y}}$, with $\bar{\boldsymbol{y}}$ and $\boldsymbol{\Psi}\boldsymbol{\Psi}^T$ already known. By using the same training set, a metamodel for computing average speeds is also built. The metamodels for computing traffic flow and average speed at link resolution allow us carry out large numbers of model evaluations. The global sensitivity analysis for computed traffic flow and average speeds with respect to each element of the input $\boldsymbol{p}$ is then carried out by computing Sobol' indices introduced in Section 5.2.

5.3.2 Variation domain of Meta-LCF model inputs for computing Sobol' indices

The inputs $\boldsymbol{p} = (p_1, p_2, \ldots, p_i), i \in \{1, 2, \ldots, 23\}$ of Meta-LCF and their variation intervals are listed in Table 5.1. The variation interval of each input is set so that it can cover the possible uncertainty. They are the same as we build the training set for the metamodeling of LCF model in Section 4.4.1 of Chapter 4. Monte-Carlo simulations of Meta-LCF are then applied by taking each input factor p_i randomly in its variation interval, following the method described in Section 5.1. The total sampling size is $n = 40000$.

Table 5.1: The variation intervals of Meta-LCF model inputs

Input	Temporal profile	Capacity	Speed	Demands in O-D matrix	Transposed O-D matrix
Symbol	$(P(h)_{h\in\mathcal{H}_{\text{atomic}}}$	$\lambda_{\text{big}}, \lambda_{\text{small}}$	$\mu_{\text{high}}, \mu_{\text{low}}$	$\delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15}$	η
Number	13	2	2	5	1
Variation	$[0.00, 1.5]$	$[0.70, 1.30]$	$[0.80, 1.20]$, $[0.70, 1.10]$	$[0.25, 1.50]$	1 or 0

5.4 Global sensitivity analysis results for the DTA simulation in the agglomeration of Clermont-Ferrand

For the simulation time h_{simu}, the outputs of the Meta-LCF are the computed flows $\boldsymbol{y}(h_{\text{simu}})$ and average speed $\boldsymbol{V}(h_{\text{simu}})$ at link level during $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$. The GSA study aims at investigating how the computed traffic flow and average speed *on each link* is influenced by the 23 elements of the input $\boldsymbol{p}$. For each link, the computed traffic flow/speed can be considered as a scalar output of the Meta-LCF model with $K = 23$ inputs: $y_a = g_1(p_1, p_2, \ldots, p_i, \ldots, p_{23})$, $V_a = g_2(p_1, p_2, \ldots, p_i, \ldots, p_{23})$. With respect to the same input, the sensitivity of the computed output on different links may be different. The first-order and total-effect sensitivity indices ($S_{i,a}$ and $S_{Ti,a}$) for traffic flows and average speeds are computed at all links.

The GSA results are represented by plotting sensitivity indices in two kinds of figures: boxplot and maps. Firstly, the boxplots allow us to compare the influence of all input factors among each other, throughout the whole network. Each point in a boxplot corresponds to $S_{a,i}$ or $S_{T_{a,i}}$ on link a of input factor p_i. The box contains the indices of values between the 1^{st} quartile (Q_1) and the 3^{rd} quartile (Q_3). The interquartile range (IQR) is Q_3 - Q_1. The red line corresponds to the median value of $S_{a,i}$ or $S_{T_{a,i}}$. The Whisker is designed as 3.0: a value $< Q_1 - 3.0 \times IQR$ or $> Q_3 + 3.0 \times IQR$ will be considered as *out of range* and presented by points. Secondly, since the sensitivity indices are calculated at link level for each factor p_i, a map of all $S_{a,i}$ for the whole network of Clermont-Ferrand is displayed. The maps help us to visualize how the input factors can influence the computed flows and average speeds in function of link spatial locations.

5.4.1 Sensitivity analysis for computed traffic flows

The sensitivity of the computed flows at all links with respect to all 23 input in Equation 5.14 are displayed in Figure 5.2.

From Figure 5.2, it is observed that for the whole network, the computed flows are very sensitive to the direction of traffic demand, represented by the transposition parameter η. Concerning other factors, the computed flows during $\mathcal{H}_{\text{simu}} = [h_{\text{simu}}, h_{\text{simu}} + 0.25 \text{ hour}]$ are sensitive to the traffic demand during $[h_{\text{simu}} - 0.25 \text{ hour}, h_{\text{simu}}]$ and during $[h_{\text{simu}} - 0.5 \text{ hour}, h_{\text{simu}} - 0.25 \text{ hour}]$. The traffic demands before $h_{\text{simu}} - 0.5$ hour and after h_{simu} are essentially not influential to the output traffic flows, as shown by total-effect

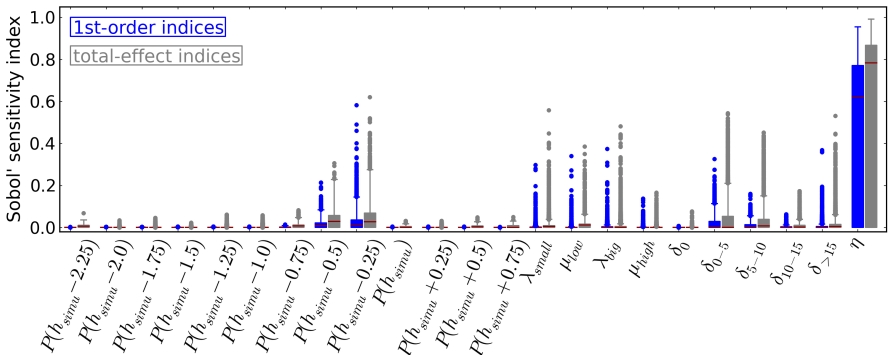


Figure 5.2: First-order (blue) and total-effect (grey) sensitivity indices of the inputs of the Meta-LCF model for the computed traffic flows

indices. Concerning demands between O-D pairs, the computed flows are sensitive to traffic demands of the O-D pairs whose inter-distance is between 0 km and 5 km. That is because most of the travels in the given static O-D matrix are in this category of distance. Concerning link capacities and speed limits, they are not very influential to the flows computed by LCF model. The median values for $\lambda_{small}, \lambda_{big}, \mu_{low}, \mu_{high}$ are nearly zero. This might due to the limited variation ranges for these inputs (less than $\pm 30\%$).

The spatial differences of input influence on computed flows are illustrated by two examples. Figure 5.3 shows the spatial distribution of $S_{a,i}$ resulting from GSA, for traffic demands of O-D pairs whose inter-distance is less than 5 km ($\delta_{0\text{-}5}$, left) and more than 15 km ($\delta_{>15}$, right) in the whole agglomeration.

Figure 5.3 shows that the computed traffic flows on roads in the city of Clermont-Ferrand are sensitive to the demands between low-distance O-D pairs, while the computed flows on highways and roads of suburbs are sensitive to the traffic demands between high-distance O-D pairs. This is due to (i) the route choices of travelers and (ii) the spatial distribution of working/residential zones in the agglomeration of Clermont-Ferrand. In fact, O-D pairs with low-distance are mainly located in the city of Clermont-Ferrand so that the corresponding traffic is assigned on roads in the city, and the flows on these roads are more sensitive to the demands between low-distance O-D pairs than long-distance demand. For users traveling between long-distance O-D pairs, the roads outside the city with high capacities are chosen, so that the flows on these roads are more sensitive to the traffic demands between long-distance O-D pairs than demands between low-distance O-D pair. Demands of long-distance O-D pairs represent often the traffic demands between the city of Clermont-Ferrand and suburb residential areas, such as the north part or the south-east part of the agglomeration.

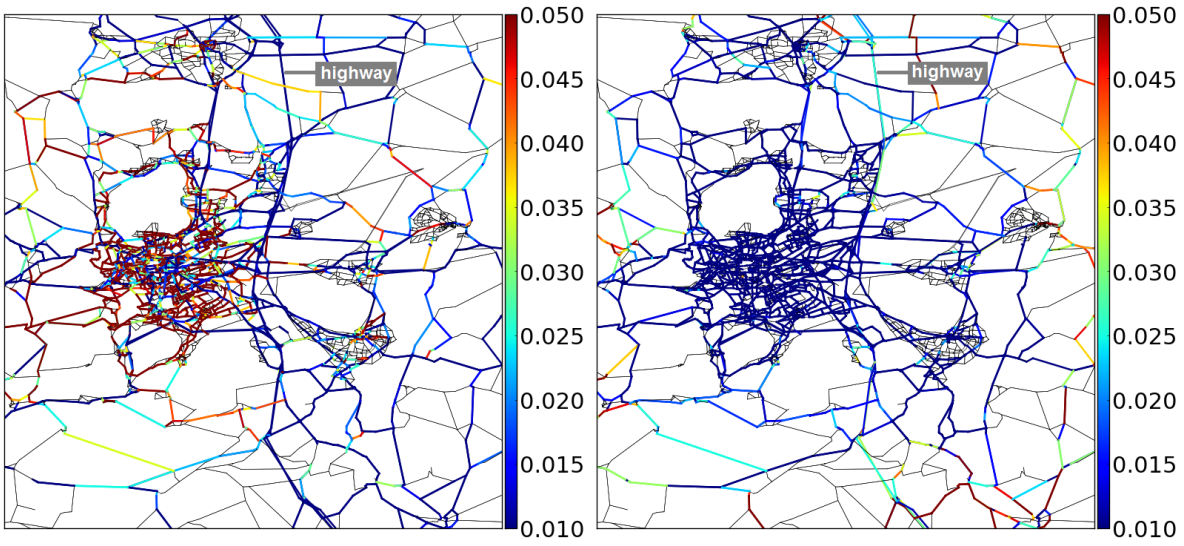


Figure 5.3: The spatial distribution of $S_{\delta_{0\text{-}5}}$ (left) and $S_{\delta_{>15}}$ (right) in the whole agglomeration of Clermont-Ferrand. $\delta_{0\text{-}5}$ represents the uncertainty in the traffic demand of O-D pairs between which the distance is non-zero and lower than 5 km. $\delta_{>15}$ represents the uncertainty in the traffic demand of O-D pairs between the distance is greater than 15 km. For the links that share the same nodes but are oriented in opposite directions, only the largest Sobol' index of the two directions is displayed.

That is why the input coefficient δ_{0-5} would influence mostly the flows on roads in the city, while the coefficient $\delta_{>15}$ would affect mostly the traffic flows in suburb areas.

Figure 5.4 presents the spatial distribution of $S_{a,i}$ for the traffic demand during $[h_{\text{simu}} - 0.5 \text{ hour}, h_{\text{simu}} - 0.25 \text{ hour}]$ ($P(h_{\text{simu}} - 0.5 \text{ hour})$ in $\boldsymbol{p}$, left), and demand during $[h_{\text{simu}} - 0.25 \text{ hour}, h_{\text{simu}}]$ ($P(h_{\text{simu}} - 0.25 \text{ hour})$ in $\boldsymbol{p}$, right).

It can be observed in Figure 5.4 that the flows computed on the highways and roads outside of the city Clermont-Ferrand are more sensitive to $P(h_{\text{simu}} - 0.5 \text{ hour})$ than flows computed on roads in the city. This implies that traffic demand during $\mathcal{H}_{\text{pre\_0.5\_0.25 hour}} = [h_{\text{simu}} - 0.5 \text{ hour}, h_{\text{simu}} - 0.25 \text{ (hour)}]$, can affect the *current* computed flows $y(h_{\text{simu}})$ during $\mathcal{H}_{\text{simu}}$ on highways and roads in suburbs. In fact, it is obvious that the demands between more distant O-D pairs require longer time to travel from the origin zone to the destination zone. For example, if users travel at a speed more than 60 km h^{-1}, they spend more than 15 min for traveling more than 15 km. Therefore, the variations of long-distance demands of travelers departing during $\mathcal{H}_{\text{pre\_0.5\_0.25 hour}}$, can still influence the computed traffic flows $y(h_{\text{simu}})$ during $\mathcal{H}_{\text{simu}}$, especially on highways and roads in suburbs for the same reasons presented in previous paragraph. This is an

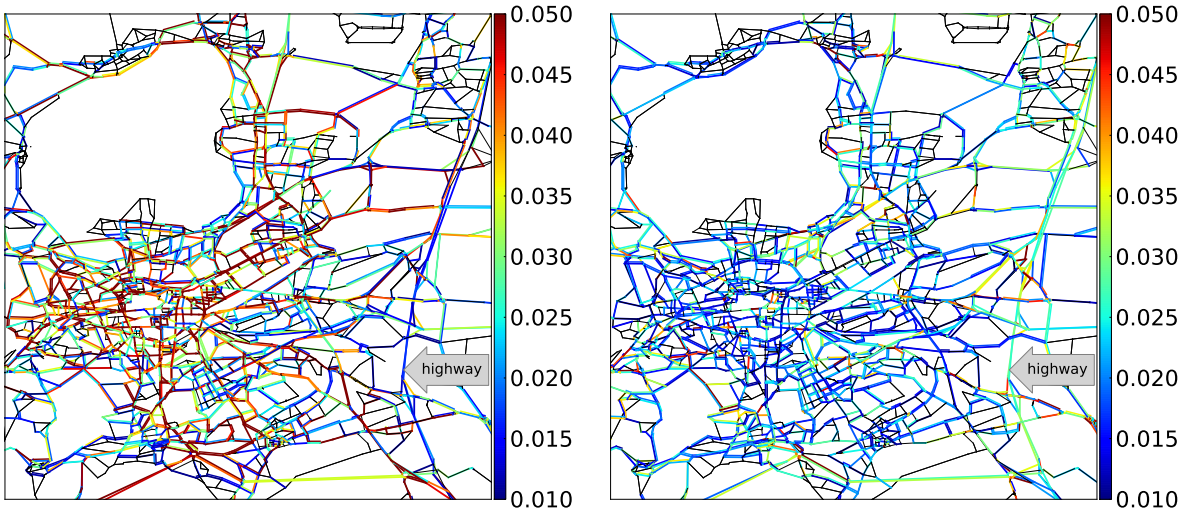


Figure 5.4: The spatial distribution of $S_{a,P(h_{\text{simu}}-0.25\text{ hour})}$ (left) and $P(h_{\text{simu}} - 0.5\text{ hour})$ (right) in the city of Clermont-Ferrand. $P(h_{\text{simu}} - 0.25\text{ hour})$ represents the uncertainty in traffic demand during 15 min before the target simulation time. $S_{a,P(h_{\text{simu}}-0.5\text{ hour})}$ represents the uncertainty in the traffic demand during the previous 30 min to 15 min before the target simulation time. Black lines represent links for which the Sobol' first-order indices are less than 0.01.

interesting result that we can obtain thanks to the dynamic traffic assignment. Note that the static traffic models cannot take the previous temporal variations of the traffic demands into account for the simulation hour.

5.4.2 Sensitivity analysis for computed average speeds

The first-order and total-effect indices of the computed speeds with respect to the 23 inputs are presented in Figure 5.5. Results show that the direction of the traffic demands in O-D matrix is also an important factor to the output travel speeds. Besides, the computed average speeds are also influenced by the uncertainty parameter associated with the $\boldsymbol{V}_0^{low}$: μ_{low}. It corresponds to the uncertainty of speed limits on links whose speed limits are less than 50 km h^{-1}. If we focus on the values *out of range* in boxplots of the first-order indices (blue points), we can find that the factors which are influential on the computed traffic flows mentioned in Section 5.4.1 have also some contribution to the variance of output speeds on certain links.

In order to analyze the influence of inputs besides O-D matrix transposition, Figure 5.6 presents the spatial distribution of the first-order sensitivity index on each link

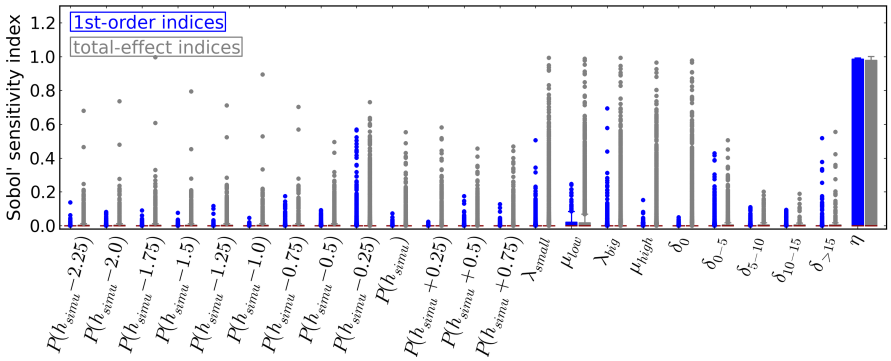


Figure 5.5: First-order and total effect Sobol' sensitivity index of the inputs of the Meta-LCF model for the computed average speeds

with respect to μ_{low} and μ_{high}. It is normal (and reassuring) to find that the computed average speeds on "low-speed links" are sensitive to the variation of the speed limits on this category of links, and the computed speeds on "high-speed links" are sensitive to the variation of the speed limits of these links, if we compare the results with the speed limit map shown in Figure 5.7.

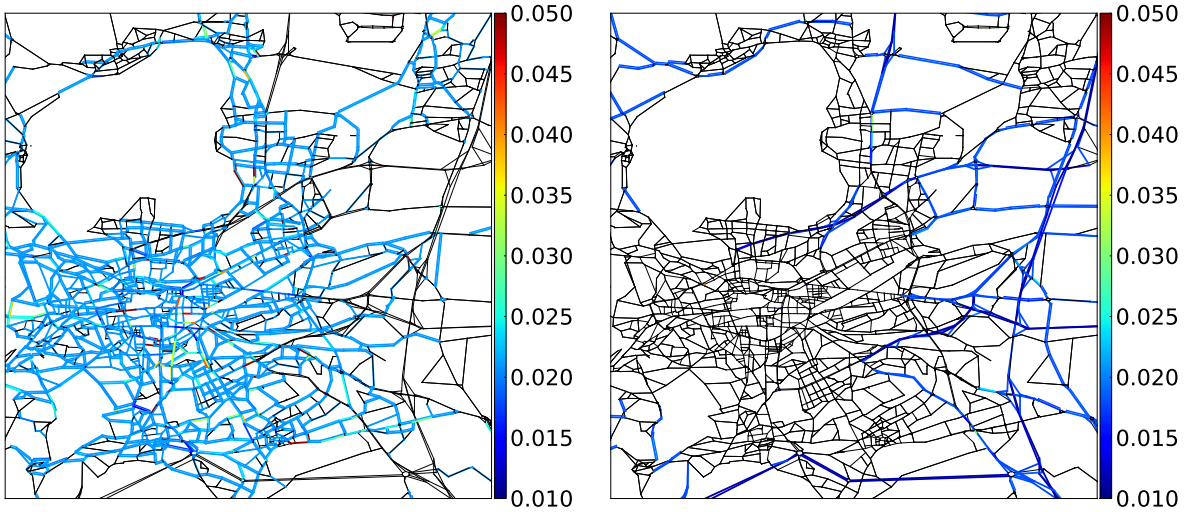


Figure 5.6: The spatial distribution of sensitivity of the computed average speeds by Meta-LCF model, to the variations of speed limits of 'low-speed links' (left), and 'high-speed links' (right), in the agglomeration of Clermont-Ferrand. Refer to Figure 5.7 to see the maps of speed limits in the city of Clermont-Ferrand. Black lines represent links for which the Sobol' first-order indices are less than 0.01.

Besides μ_{low} and μ_{high}, the traffic demand during the previous 30 min and especially

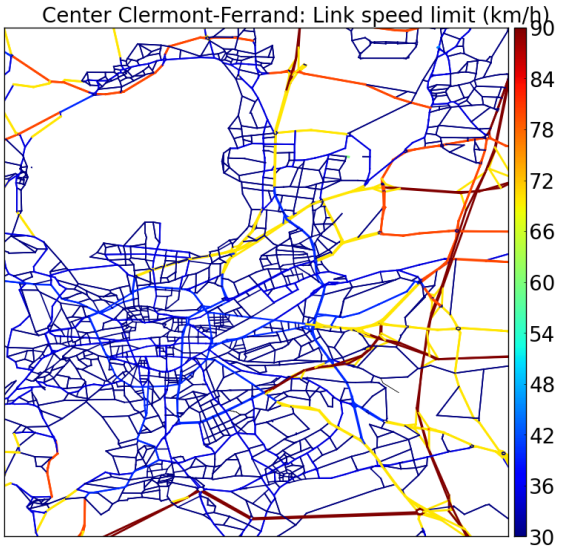


Figure 5.7: Maps of speed limits in $\mathrm{km\,h^{-1}}$ in the city of Clermont-Ferrand.

during the previous 15 min are also influential to the computed average speed on some links, if we focus on the first-order indices "out of range" plotted in the boxplot of Figure 5.5. Figure 5.8 shows the influence of $P(h_{\text{simu}} - 0.25 \text{ hour})$ on the average speeds of Meta-LCF model for the city of Clermont-Ferrand.

We have already observed that computed traffic flow is very sensitive to the traffic demand. The traffic demand cannot directly affect the computed average speed, but the resulting traffic flow can influence the average speed on the network since the travel time on a link is computed as a function of (i) the free flow travel time of the link which depends on its speed limit, and (ii) traffic flow on the link computed by the traffic assignment model. Therefore, if the traffic flow on a link is very sensitive to the some inputs of the Meta-LCF model, the computed averaged speed might also be sensitive to the same inputs. Even though there are many links where the computed traffic flow is sensitive to the traffic demand during the previous 15 min, as shown in Figure 5.4 (left) in Section 5.4.1, Figure 5.8 (left) shows that only for some of them the computed average speeds are sensitive to the traffic flows. In fact, if we compare the sensitivity map with link capacity map in Figure 5.8 (right), we can see that only on links with low capacity ($< 1000\,\mathrm{veh\,h^{-1}}$), especially on low-capacity links connected to links with higher capacities, the computed average speeds are sensitive to the input traffic demand. This suggests that it is more likely to observe congestion phenomena on these links. The computed average speeds are not very sensitive to the traffic flows on links with capacities more than $2000\,\mathrm{veh\,h^{-1}}$. These results are due to the assumption in the LADTA model, that vehicles can travel with the same speed as the speed limit,

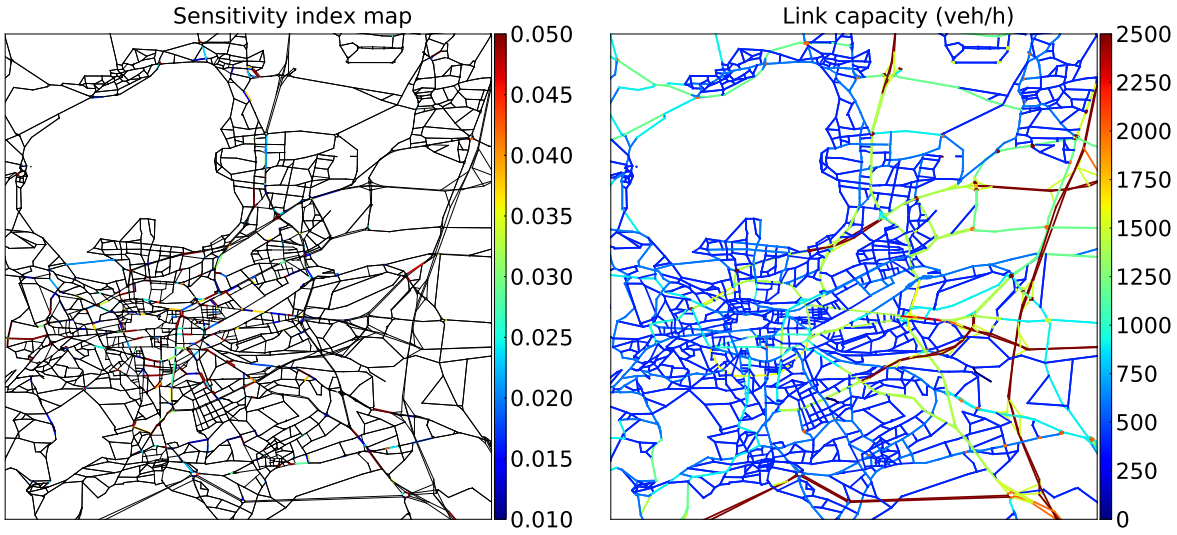


Figure 5.8: First-order Sobol' sensitivity index for the computed average speed with respect to (i) $P(h_{\text{simu}} - 0.25 \text{ hour})$ (left), and to (ii) link capacity (right) at street resolution for the city of Clermont-Ferrand. $P(h_{\text{simu}} - 0.25 \text{ hour})$ represents the uncertainty in traffic demand during 15 min before the target simulation time. In the left figure, black lines represent links for which the Sobol' first-order indices are less than 0.01.

when the computed flows do not exceed the link capacities. With the demand variation presented in Table 5.1, the computed traffic flows on these big-capacity roads rarely exceed the link capacities.

Conclusions

With the metamodeling approach and Sobol' method, a global sensitivity analysis (GSA) was carried out for the meta-model of a DTA model applied in the agglomeration of Clermont-Ferrand (France). The Sobol' method was chosen for GSA studies. The GSA is carried out to the Meta-LCF computing traffic flows and average speeds at street level.

GSA results show that the computed traffic flows are very sensitive to the direction of traffic demands in the O-D matrix. They are also sensitive to the volume of traffic demands: (i) the temporal variation during the previous 0.5 hour before the simulation time and (ii) the total demands between O-D pairs for which the inter-zone distance is between 0 - 5 km. The computed flows are not very sensitive to the link capacity or speed limit in our case study. Besides, the resulting Sobol' indices also depend on

spatial location and characteristics of the links. The computed flows on highways are sensitive to the traffic demands departing from 0.5 hour before the simulation period and the traffic demands between long-distance O-D zones, while the computed flows on links in the city center are sensitive to the traffic demands during the previous 0.25 hour and the demands between short-distance O-D zones. For the average speeds computed by Meta-LCF model however, besides the direction of traffic demand and the uncertainty inputs corresponding to the street speed limits, the computed speeds are not very sensitive to other inputs of the Meta-LCF model on most of the streets in our case study. However, results show that on some street where the capacity is low, the computed speeds are sensitive to the traffic demands. This suggests that traffic jams are more likely to happen on these streets.

The GSA study results help us to identify the most influential inputs in dynamic traffic assignment model. In the uncertainty quantification, the transposition input η is not an uncertain input since it depends on the simulation time of LADTA simulation. The traffic demands are influential inputs for computed traffic flows. This refers that it is very likely the uncertainty in the traffic demand will propagate to the computed traffic flows. The computed average speeds are not very sensitive to the 22 inputs except η. This refers that the uncertainty of computed average speeds due to the uncertainty of the 22 inputs might not be significant. The quantified uncertainty analysis results are presented in Chapter 7.

Chapter 6

Air pollutant emissions for the agglomeration of Clermont-Ferrand and sensitivity analysis

Summary

In this chapter, we build a modeling chain for estimating on-road traffic emissions by coupling the dynamic traffic assignment (DTA) model LADTA with the emission factor model COPERT IV, in the whole agglomeration of Clermont-Ferrand at street resolution. A qualitative local sensitivity analysis (LSA) with the *one-at-a-time* method is carried out on the modeling chain. The LSA studies the sensitivity of the computed on-road traffic emissions with respect to the LADTA inputs as well as the inputs of COPERT IV model: (i) total traffic demand, (ii) speed limits of the road network, and (iii) vehicle fleet composition. The results of LSA show that the emissions are very sensitive to these factors, especially during the transition from a free traffic network to a congested one. The COPERT IV model is also coupled with the metamodel built in Chapter 4. With the modeling chain coupling COPERT IV and the metamodel of LADTA, we are able to carry out emission estimations at urban area during a long-term period (one year) with high spatial and temporal resolutions. Then the global sensitivity analysis with Sobol' method is carried out to study the sensitivity of on-road traffic emissions computed by the modeling chain in the whole agglomeration at street resolution, with respect to both (i) the uncertain inputs of LADTA model, and (ii) the vehicle fleet composition inputs for COPERT IV. The Sobol' first-order and total-effect sensitivity indices show that the hot emissions of on-road traffic are very sensitive to all the inputs that are influential to the traffic flow computed by the DTA model. Besides, the emissions are also sensitive to gasoline car share in passenger cars, and the proportion of heavy duty vehicles. Furthermore, spatial differences are found among the sensitivity indices throughout the metropolitan area. For example, on streets without Heavy Duty Vehicles (HDVs),

the most influential factor that affects the computed NOx (hot) emissions is the share of gasoline vehicles, while on streets with HDVs, the share of HDVs is the most influential factor. In addition, the factors that influence the computed (hot) emissions of different pollutants are not the same. For example, the computed CO emissions are not as sensitive to the HDV share as the computed NOx emissions.

Résumé

Dans ce chapitre, on construit une chaîne de modélisation en couplant le modèle d'affectation dynamique du trafic LADTA avec le modèle d'émission COPERT IV, à la résolution de la rue. Cela permet d'estimer les émissions dues au trafic routier à l'échelle urbaine et à une résolution temporelle fine. Ensuite, une étude de sensibilité locale est effectuée sur cette chaîne de modélisation. On étudie la sensibilité des émissions par rapport aux entrées du modèle LADTA ainsi qu'à celles du modèle COPERT IV: (i) la demande totale du trafic, (ii) les vitesses maximales autorisées sur le réseau, et (iii) la composition du parc. Les résultats montrent que les émissions sont très sensibles à ces entrées, surtout lorsqu'une congestion du trafic se produit. COPERT IV est également couplé avec le métamodèle construit dans le chapitre 4. Cela nous permet d'effectuer des simulations d'émissions à l'échelle urbaine pendant une longue période (un an). L'analyse de sensibilité globale avec la méthode de Sobol' est aussi effectuée sur cette chaîne de modélisation couplant le métamodèle de LADTA et COPERT IV. Les indices de sensibilité montrent que les émissions en sortie sont sensibles à toutes les entrées qui influencent le débit simulé par LADTA. De plus, les émissions sont également sensibles à la composition du parc: (i) la proportion de véhicules essence et (ii) la proportion de poids lourds (PL) sur le réseau. Par ailleurs, les émissions simulées dans différentes rues sont sensibles à différentes données d'entrée. Par exemple, dans les rues où il n'y a pas de PL, les émissions sont sensibles à la proportion de véhicules essence. Au contraire, dans les rues où les PL sont présents, les émissions sont beaucoup plus sensibles à la proportion de PL. De plus, les facteurs les plus influents pour les émissions estimées ne sont pas les mêmes pour différents polluants. Par exemple, la proportion de PL est moins influente pour les émissions de CO que pour les émissions de NOx.

Contents

This chapter is reproduced from the articles: (i) Chen, R., Aguiléra, V., Mallet, V., Cohn, F., Poulet, D., and Brocheton, F. (2017). A sensitivity study of road transportation emissions at metropolitan scale. *Journal of Earth Science & Geotechnical Engineering*, 7(1):151-173. (ii) Chen, R., Mallet, V., Aguiléra, V., Cohn, F., Poulet, D., and Brocheton, F. (2017). Global Sensitivity Analysis in the Simulation of Road Traffic Emissions at Metropolitan Scale. In *Proceedings of the 22th International Transportation and Air Pollution Conference (TAP 2017)*.

6.1 Introduction

The emissions of road traffic are one of the main sources of air pollutants in urban area. In Île-de-France for instance, the road traffic is responsible for more than 55 % of nitrogen oxides (NOx) and more than 30 % of particulate matter [Airparif, 2014]. One of the biggest contributors to urban air pollution is the on-road traffic emissions [Smit et al., 2010; Franco et al., 2013], and the associated uncertainty is then an important factor for the air quality simulation. The on-road traffic emissions are often modeled as the product of (i) the number of vehicles and (ii) the emission factors (EF, in $g\,km^{-1}$), for different pollutants, vehicle types and technologies, etc. In metropolitan areas, the former can be estimated by traffic assignment models, and the latter are often modeled by empirical functional relations between pollutant emissions and characteristics of on-road traffic. Various models are built for EFs, and five main types of EF models are categorized by Smit et al. [2010] according to the required input variables, such as (i) 'average-speed' models (e.g. COPERT, MOBILE, EMFACA), where EFs are a function of the mean traveling speed; (ii) 'traffic-situation' models (e.g. HBEFA) where EFs are determined by descriptions of a particular traffic situation (e.g. 'stop-and-go-driving', 'free flow driving'); or (iii) 'modal' models (e.g. PHEM) where EF (in $g\,s^{-1}$) are produced via engine of vehicle operating models at the highest resolution (down to seconds), etc. Different types should be applied for different contexts to adapt the level of detail required by researchers, city planners or policy makers.

Among different types of EF models, there is no conclusive evidence that demonstrates that more complex models systematically perform better in terms of prediction error than less complex models [Smit et al., 2010]. In fact, complex models require more detailed inputs in order to predict more accurate EFs, but the uncertainty of extra inputs would also add complexity and uncertainty to the EF models themselves. Complete sensitivity analysis of each EF models are needed, in order to evaluate the reliability and uncertainty of the computed EFs. For example, Kioutsioukis et al. [2004]; Kouridis et al. [2010] carried out a global sensitivity analysis (GSA) on COPERT model via the Monte Carlo method and the extended FAST method. They found that vehicle fleet share are among the most important factors that influence the output emissions of most pollutants (PM, VOC, NOx, etc.). However, this is an individual sensitivity analysis for an isolated EF model and the uncertainty from the traffic inputs and traffic models were not taken into account. In fact, concerning emission estimations, the required traffic inputs (number of vehicles and travel speeds) are also computed by models. The uncertainty of the traffic models and their inputs might be propagated in the simulation coupling traffic models and EF models, and affect the final computed emissions.

Therefore, it is necessary to carry out a sensitivity analysis through the whole modeling chain from traffic modeling to emission estimation.

The main purpose of this chapter is to build the modeling chain from dynamic traffic assignment model to emission estimation at street level. Then it is possible to carry out the qualitative and quantitative sensitivity analysis on the modeling chain, in order to study how both the traffic model inputs and emission model inputs can influence the computed emissions at an metropolitan area. Concerning the traffic assignment model, a dynamic traffic assignment model called LADTA [Leurent, 2003] is used in order to better represent the congestion phenomenon, and the temporal-spatial variation of traffic flow and travel time. LADTA computes time-varying traffic flows and average speed at street resolution. These outputs correspond to traffic input required by COPERT model. Therefore, we built a modeling chain from traffic assignment to emission estimation: LADTA-COPERT IV. The emissions are also at street resolution, and they are computed at the same time resolution as that of the computed traffic flow/speed from LADTA model. A case study on a working Tuesday in the agglomeration of Clermont-Ferrand is carried out. With the modeling chain LADTA-COPERT IV, emissions are computed at street resolution at each time interval of 15 minutes during a whole day. Then a qualitative sensitivity analysis is carried out to the modeling chain LADTA-COPERT IV [Chen et al., 2017]. We vary the total traffic demand, speed limits and vehicle fleet composition in the agglomeration to see how they can influence the computed NOx emission due to on-road traffic.

For the quantitative sensitivity analysis, the global sensitivity analysis (GSA) with Sobol' method presented in Chapter 5 is used here. Since the GSA with Sobol' method requires a large number of model evaluations, a modeling chain coupling the metamodel of LADTA (cf. the Meta-LCF model built in Chapter 4) model and COPERT IV is built: Meta-LCF-COPERT IV. The efficiency of Meta-LCF model allows us to compute street resolution traffic flow and average speed during any interval of 15 minutes. A whole-year emission simulation is then carried out for NOx and CO. The low model evaluation time of Meta-LCF-COPERT IV also allows us carry out GSA with Sobol' method. We perturb independently the inputs of meta traffic assignment model and COPERT IV model in order to analyze the contribution of each input to the computed emissions at street resolution. The case study is carried out in the agglomeration of Clermont-Ferrand, but the method can be applied to other networks.

This chapter is organized as follows. Firstly, in Section 6.2, the modeling chain coupling the complete LADTA model and the COPERT IV model is built. Then in Section 6.3 the built modeling chain is applied to the agglomeration of Clermont-Ferrand,

which is an important industrial city located in the center part of France. In the case study, we computed traffic flows, average traffic speeds, NOx and CO emissions at street level for the whole agglomeration during the whole year 2014. At last, qualitative and quantitative sensitivity analysis is carried out for the modeling chain from dynamic traffic assignment to emission estimations in Section 6.4. The inputs of LADTA model as well as vehicle fleet parameters are analyzed in order to study the influence of inputs from both DTA model and COPERT IV model to the final estimated emissions. The GSA of estimated emissions is carried out for the modeling chain of Meta-LCF-COPERT IV. Both the first-order effect and the total effect are studied. The spatial distributions of the sensitivity of different input factors are also presented.

6.2 Modeling chain of street-level on-road traffic emission estimation

For computing on-road traffic emissions, one of the most common used methods is the tier 3 method of the EMEP CORINAIR emission inventory guidebook of European Environment Agency [EEA, 2016]. The total exhaust emissions of on-road traffic are calculated as the sum of hot emissions and cold-start emissions. In this study, only hot emissions are considered. According to the tier 3 method of EMEP inventory guidebook, the hot emissions can be estimated by the following formulation:

$$E_{hot,i,j}(\mathrm{g}) = e_{hot,i,j}(\mathrm{g\,km^{-1}}) \times N_j(\mathrm{veh}) \times M_j(\mathrm{km\,veh^{-1}}), \tag{6.1}$$

where $E_{hot,i,j}$ and $e_{hot,i,j}$ are respectively the hot emission and hot emission factor (HEF) of pollutant i, for vehicle of technology j. N is the number of vehicles of technology j. M_j is the mileage of vehicle of technology j. The tier 3 method of EMEP inventory guidebook was initially introduced to estimate the exhaust on-road emissions of a whole country during long periods of time. The emissions at street level can be computed by combining the HEF with traffic flows and average speeds at street level. Then the Equation 6.1 is applied to each road a of the studied area and we get

$$E_{hot,i,j,a}(\mathrm{g}) = e_{hot,i,j,a}(\mathrm{g\,km^{-1}}) \times N_{j,a}(\mathrm{veh}) \times L_a(\mathrm{km\,veh^{-1}}), \tag{6.2}$$

where $E_{hot,i,j,a}$ and $e_{hot,i,j,a}$ are respectively the hot emissions and the HEFs on the road a for pollutant i, for vehicle of technology j. $N_{j,a}$ is the number of vehicles on the road a. M_j in Equation 6.1 is replaced with the road length of a, which is denoted as L_a in Equation 6.2. In order to compute $E_{hot,i,j,a}$, we need to compute the traffic volume,

as well as the emission factor of each pollutant, associated with each type of vehicle. The emission factors can be modeled by COPERT IV formula in EMEP inventory standards [EEA, 2016; Gkatzoflias et al., 2007]. In this study, we only consider hot exhaust emissions. According to COPERT IV, the hot emission factor (HEF) e_{hot} is a function $g_{i,j}$ of the vehicle average speed v, for a given pollutant i and a given class of vehicle technology j. For example, for pollutant NOx and for gasoline passenger cars (PCs) with emission standard of Euro1 to Euro4,

$$e_{\text{hot}} = g_{\text{NOx,euro1\_euro4}}(v) = \frac{a + c \times v + e \times v^2}{1 + b \times v + d \times v^2}, \tag{6.3}$$

for gasoline passenger cars (PC) of Euro 1 to Euro 4. While for gasoline PC of Euro 5 to Euro 6c, the formulation is then:

$$e_{\text{hot}} = g_{\text{NOx,euro5\_euro6}}(v) = a \times v^5 + b \times v^4 + c \times v^3 + d \times v^2 + e \times v + f. \tag{6.4}$$

In Equation 6.3 and 6.4, the parameters $a, b, \ldots, f$ depend on the technology of emission standards for gasoline PC. The function $g_{i,j}$ itself as well as its parameters depend on vehicle type and pollutant. If the vehicle technologies and vehicle speeds are given at street level, e_{hot} can be computed at street level as well. The computation of emissions at street level combining with traffic information is implemented and published as an open-source program *Pollemission* [Chen and Mallet, 2016].

6.3 Emission results for the agglomeration of Clermont-Ferrand

Both the complete LADTA model and the metamodel Meta-LCF compute time-varying traffic flow and average speed at street resolution. The temporal resolution is $\Delta h = 0.25$ hour. Let $y_{a,h}$ and $v_{a,h}$ denote the computed traffic flow and average speed at link a during $[h, h + \Delta h]$, respectively. They are computed by the complete model LADTA or the metamodel of LADTA applied to the agglomeration of Clermont-Ferrand. Based on Equation 6.2, the hot emission at link a during $[h, h + \Delta h]$ can then be computed as

$$\begin{aligned} E_{hot,i,j,a,h}(\text{g}) = e_{hot,i,j,a,h}(\text{g km}^{-1}) \times y_{j,a,h}(\text{veh}) \\ g_{hot,i,j}(v_{a,h}) \times y_{j,a,h}, \end{aligned} \tag{6.5}$$

where $g_{hot,i,j}$ is the HEF function for pollutant i and vehicle type j. It is calculated based on average speed $v_{a,h}$. Firstly, the modeling chain LADTA-COPERT IV is used to compute NOx emission during a working Tuesday in the agglomeration. Then a qualitative sensitivity analysis is carried out to the LADTA-COPERT IV to see how the NOx emission is influenced by the input total demand and network speed limits during each interval of 15 minutes of the whole day. Secondly, $y_{a,h}$ and $v_{a,h}$ computed by Meta-LCF model is coupled with COPERT IV model. The modeling chain Meta-LCF-COPERT IV is used to compute CO and NOx emission at street resolution during the whole year 2014. There are in total 23 inputs for computing $y_{a,h}$ and $v_{a,h}$, and 6 added inputs with vehicle fleet inputs for COPERT IV model. A GSA using Sobol' method is then applied to Meta-LCF-COPERT IV to analyze the contribution of all the 29 inputs of the modeling chain to the computed NOx and CO emission at street resolution.

6.3.1 Inputs

Dynamic traffic assignment inputs for LADTA and Meta-LCF model

We recall here the inputs of LADTA model and Meta-LCF model.

$\mathcal{G} = (\mathcal{N}, \mathcal{A})$: the nodes ($\in \mathcal{N}$) and links ($\in \mathcal{A}$) of the oriented graph $\mathcal{G}$ modeled for the network of the agglomeration of Clermont-Ferrand. $D = card(\mathcal{A}) = 19628$. D is the number of links.

$\mathcal{Z}_O, \mathcal{Z}_D \subset \mathcal{N}$: the set of nodes representing the Origin-Destination zones. $z = card(\mathcal{Z}_O) = card(\mathcal{Z}_D) = 124$.

$\boldsymbol{C} = [C_a]_{a\in\mathcal{A}}$, $\boldsymbol{L} = [L_a]_{a\in\mathcal{A}}$, $\boldsymbol{V_0} = [V_{0a}]_{a\in\mathcal{A}}$: the vector of link capacities, link lengths and speed limits. $\boldsymbol{C}, \boldsymbol{L}, \boldsymbol{V_0} \in \mathbb{R}^D$.

$\boldsymbol{L} = [L_a]_{a\in\mathcal{A}}$, $\boldsymbol{T_0} = [T_{0a}]_{a\in\mathcal{A}}$ and $\boldsymbol{C} = [C_a]_{a\in\mathcal{A}}$: vector of link length, free-flow travel time and link capacity.

$\mathcal{H} = [h_0, h_1]$: the simulation period of LADTA.

Δh: time step for loop detectors to register aggregated traffic flow data. It is also assumed that during the interval $[h, h + \Delta h]$, the traffic flow on the network is constant, and so is the time-varying traffic demand. In our case study, $\Delta h = 0.25$ hour.

$\boldsymbol{Q}(h) = (Q_{o,d}(h))_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D}$: the time-varying O-D matrix. $Q_{o,d}(h)$ is the cumulated traffic demand (total number of vehicles in veh) from o to d during $\mathcal{H}$. It is non-decreasing and can be deduced from time-varying traffic demand *density* $q_{o,d}(h)$ (in $\mathrm{veh\,h^{-1}}$) at instant h for the O-D pair *od*. $q_{o,d}(h)$ is assumed to be a piece-wise linear function of h and $q_{o,d}(h)$ is constant in the interval $[h, h + \Delta h]$. With known $q_{o,d}(h)$ and

bounded $\mathcal{H}$, $(Q_{o,d}(h))_{(o,d)} = \int_{h_0}^{h} q_{o,d}(\tilde{h})d\tilde{h}$ with $h \in \mathcal{H}$.

$\boldsymbol{Q}^{\text{peak}} = (q_{o,d}^{\text{peak}})_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D}$: O-D matrix of the agglomeration of Clermont-Ferrand during the evening peak hour 17:00 – 18:00, where $(q_{o,d}^{\text{peak}})_{(o,d)}$ is the average traffic demand during the evening peak. It is the static O-D matrix.

In Meta-LCF model, in order to represent the uncertainty of the network parameters, links are regrouped according to link capacities and speed limits.

$\boldsymbol{V}_0^{low} = [V_{0a}]_{a\in\mathcal{A}_{low}}$: the vector of low speed limits ($V_0 \leq 50\,\text{km}\,\text{h}^{-1}$), with $\mathcal{A}_{low}$ as the set of links with $V_0 \leq 50\,\text{km}\,\text{h}^{-1}$. Links in the set $\mathcal{A}_{low}$ are considered "low-speed links".

$\boldsymbol{V}_0^{high} = [V_{0a}]_{a\in\mathcal{A}_{high}}$ with $\mathcal{A}_{high} = \mathcal{A}\backslash\mathcal{A}_{low}$: the vector of high speed limits. Links in the set $\mathcal{A}_{high}$ are considered "high-speed links".

$\boldsymbol{C}^{small} = [C_a]_{a\in\mathcal{A}_{small}}$: the vector of small link capacities, with $\mathcal{A}_{small}$ as the set of links with $C \leq 900\,\text{veh}\,\text{h}^{-1}$.

$\boldsymbol{C}^{big} = [C_a]_{a\in\mathcal{A}_{big}}$ with $\mathcal{A}_{big} = \mathcal{A}\backslash\mathcal{A}_{small}$: the vector of big link capacities.

For a whole-day DTA simulation with the complete LADTA model, $\mathcal{H} = [0.0, 24.0]$ hour for a working Tuesday. $(Q_{o,d}(h))_{(o,d)\in\mathcal{Z}_O\times\mathcal{Z}_D}$ is the dynamic O-D matrix converted from the given static matrix $\boldsymbol{Q}^{\text{peak}}$. LADTA then computes $(y_{a,h})_{a\in\mathcal{A}}$ and $(v_{a,h})_{a\in\mathcal{A}}$ for $h \in \mathcal{H}$. Detailed description for the whole-day simulation with LADTA is presented in Section 3.3 and Section 3.4.3 of Chapter 3.

For the metamodel, we recall the Meta-LCF model for the agglomeration of Clermont-Ferrand in Equation 6.6. It computes traffic flows y in $\text{veh}\,\text{h}^{-1}$ at street resolution during a time step of $[h_{simu}, h_{simu} + \Delta h]$, $\Delta h = 0.25$ (h in hour) on a day d. The road network in the agglomeration of Clermont-Ferrand is modeled by oriented graph and the roads are represented by *links*.

$$\begin{aligned}\boldsymbol{y}_d(h_{simu}) &= \widehat{\mathcal{M}}(\xi_d P_d(h)_{h\in I_{\text{atomic}}}, \delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15}, \lambda_{big}, \lambda_{small}, \mu_{high}, \mu_{low}, \eta)\\ &= \widehat{\mathcal{M}}(\boldsymbol{p}),\end{aligned} \tag{6.6}$$

where I_{atomic} is a set of instants $\{h_{simu}-2.25, h_{simu}-2.0, \ldots, h_{simu}, \ldots, h_{simu}+0.75\}$ (in hour). LADTA requires time-dependent traffic demand as input and computes time-varying traffic flow. $q_{o,d}(h)$ with $h \in \mathcal{H}_{atomic}$ can be obtained from loop detector measurements of the simulation period (refer to Section 4.3.2 in Chapter 4). We assume that the dynamic demand and the computed flow are constant during an interval of 15 minutes. Then $Q_{o,d}(h)$ can be obtained by integrating $q_{o,d}(h)$ during $[h_{simu}-2.25, h_{simu}+1.0]$ (in hour). Then $\boldsymbol{y} = (y_{a,h_{simu}})_{a\in\mathcal{A}}$ and $\boldsymbol{v} = (v_{a,h_{simu}})_{a\in\mathcal{A}}$ can be computed using Meta-LCF. We also assume that the traffic flow computed during $[h_{simu}, h_{simu}+\Delta h]$ is not influenced by

the traffic demand during the time intervals $[-\infty, h_{simu} - 2.25]$ and $[h_{simu} + 1.0, +\infty]$. $P_d(h)_{h \in I_{\text{atomic}}}$ are the 13 *temporal variation coefficients*. They are ratios between the traffic demand during $[h, h + \Delta h]$ and during the evening peak of a given simulation day d. ξ_d is the *total demand coefficient* for adjusting the total traffic demand according to different day d. δ_0, δ_{0-5}, δ_{5-10}, δ_{10-15} and $\delta_{>15}$ are five *spatial variation coefficients* to represent spatial uncertainty of the demand according to the Origin-Destination distance. $\lambda_{big}, \lambda_{small}, \mu_{high}, \mu_{low}$ represent uncertainty of road capacities and speed limits. $\eta = 1.0$ if $h_{simu} <= 12$ and $\eta = 0.0$ otherwise.

Vehicle fleet inputs for COPERT IV model

The hot emission factors (HEFs) depend on the following factors according to COPERT IV model: (i) vehicle category (passenger car, heavy duty vehicle, etc.), (ii) emission standard technology (pre-Euro, Euro 1, Euro 2, etc.), (iii) engine type (gasoline, diesel, etc.), (iv) engine capacity ($< 1.4\,\text{L}$, $1.4\,\text{L} - 2.0\,\text{L}$, $> 2.0\,\text{L}$, etc.), (v) pollutant type and (vi) vehicle average speed. The first four factors are often categorized as *vehicle fleet* inputs. LADTA model and the Meta-LCF model compute the number of vehicles and average speeds at street level. Hot emissions can then be computed at street level if vehicle fleet inputs are provided at the same level.

In our case study, it was assumed that the vehicle fleet inputs were the same on all the streets of the network. As the O-D matrix of traffic simulation is only provided for passenger cars (PCs), the outputs of the traffic assignment model are only traffic flows of PCs. Five scalar coefficients are used to represent the vehicle fleet inputs. (i) θ_{gaso} is the proportion of gasoline PCs. (ii) $\gamma_{\text{gaso\_1.4}}$ is the share of gasoline PCs with engine capacity less than 1.4 L, among all gasoline PCs. (iii) $\varepsilon_{\text{diesel\_2.0}}$ is the proportion of diesel PCs with engine capacity less than 2.0 L, among all diesel PCs. (iv) $\zeta_{\text{gaso\_euro4}}$ is the proportion of gasoline PCs with emission standard of Euro 4 and higher. (v) $\varphi_{\text{diesel\_euro4}}$ is the proportion of diesel PCs with emission standard of Euro 4 and higher. All these parameters are less than or equal to 1. For the agglomeration of Clermont-Ferrand, the default vehicle fleet data for PCs is that of France, provided in [André et al., 2013] and presented in Table 6.1 and 6.2.

In Table 6.1 and 6.2, the distribution of engine capacity ($\tau_{<1.4l}$, $\tau_{1.4-2.0l}$, $\tau_{>2.0l}$) is different according to the engine type. In France, the share of diesel PCs is bigger than that of gasoline PCs: $\theta_{\text{diesel}} > \theta_{\text{gaso}}$. In our case study, it was assumed that $\theta_{\text{gaso}} + \theta_{\text{diesel}} = 1$. The default share of emission standard is assumed to be independent of the engine type (Table 6.2). As we vary $\gamma_{\text{gaso\_euro4}}$ or $\varepsilon_{\text{diesel\_euro4}}$ for sensitivity studies in Section 6.3.2, we keep the proportion of each vehicle technology among them,

Table 6.1: Distribution of engine type and engine capacity in France. Source: [André et al., 2013]. For vehicles with emission standard Euro 6, the share is unknown and supposed to be zero.

Engine type (proportion in all PCs, in %)	Engine capacity	Proportion in each category (in %)
gasoline ($\theta_{\text{gaso}} = 29.2$)	$< 1.4\,\text{L}$ ($\tau_{\text{gaso},<1.4l}$)	59.9
	$1.4 - 2.0\,\text{L}$ ($\tau_{\text{gaso},1.4-2.0l}$)	34.6
	$> 2.0\,\text{L}$ ($\tau_{\text{gaso},>2.0l}$)	3.8
Diesel ($\theta_{\text{diesel}} = 70.6$)	$< 1.4\,\text{L}$ ($\tau_{\text{diesel},<1.4l}$)	9.8
	$1.4 - 2.0\,\text{L}$ ($\tau_{\text{diesel},1.4-2.0l}$)	77.8
	$> 2.0\,\text{L}$ ($\tau_{\text{diesel},>2.0l}$)	12.5

Table 6.2: Distribution of emission standard technology in France [André et al., 2013]

Emission standard technology	Proportion in all PCs (in %)
Pre-Euro ($\tau_{\text{pre-euro}}$)	3.6
Euro 1 (τ_{euro1})	6.6
Euro 2 (τ_{euro2})	14.0
Euro 3 (τ_{euro3})	33.5
Euro 4 (τ_{euro4})	39.9
Euro 5 (τ_{euro5})	2.4

i.e, $\tau_{\text{pre}} : \tau_{\text{euro1}} : \tau_{\text{euro2}} : \tau_{\text{euro3}} = 3.6 : 6.6 : 14.0 : 33.5$, and $\tau_{\text{euro4}} : \tau_{\text{euro5}} = 39.9 : 2.4$. Note that the share of Euro-6 PCs is unknown and assumed to be zero. In the same category of vehicles with emission standards lower or higher than Euro 4, the sum of the shares for each class is 1.0: $\tau_{\text{pre}} + \tau_{\text{euro1}} + \tau_{\text{euro2}} + \tau_{\text{euro3}} = 1.0$, and $\tau_{\text{euro4}} + \tau_{\text{euro5}} = 1.0$. As a result, with one parameter $\zeta_{\text{gaso\_euro4}}$ or $\varphi_{\text{diesel\_euro4}}$, the proportion of vehicles with other emission standards can be obtained. Similarly, concerning the engine capacity, the proportion of gasoline PCs with engine capacity $1.4\,\text{L}-2.0\,\text{L}$ and $> 2.0\,\text{L}$ is assumed to be constant, i.e., $\tau_{\text{gaso},1.4-2.0l} : \tau_{\text{gaso},>2.0l} = 34.6 : 3.8$ with $\tau_{\text{gaso},1.4-2.0l} + \tau_{\text{gaso},>2.0l} = 1.0$. The proportion of diesel PCs with engine capacity $< 1.4\,\text{L}$ and $1.4\,\text{L} - 2.0\,\text{L}$ is also assumed to be constant: $\tau_{\text{diesel},<1.4l} : \tau_{\text{diesel},1.4-2.0l} = 9.8 : 77.8$, with $\tau_{\text{diesel},<1.4l} + \tau_{\text{diesel},1.4-2.0l} = 1.0$. Therefore, the proportion of gasoline or diesel PCs with different engine capacities can be expressed by only two scalars: $\gamma_{\text{gaso\_1.4}}$ and $\varepsilon_{\text{diesel\_2.0}}$. For example, assuming that the vehicle fleet data are default values presented in Table 6.1 and 6.2, the computation of hot emissions of NOx on link a, for Euro 4 gasoline PCs, with engine capacity less

than 1.4 L can be written as

$$E_{\text{NOx, euro4, gaso\_1.4l, a}} = \zeta_{\text{gaso\_euro4}} \times \frac{\tau_{\text{euro4}}}{\tau_{\text{euro4}} + \tau_{\text{euro5}}} \times (\gamma_{\text{gaso\_1.4}} \theta_{\text{gaso}} N_a) \times e_{\text{NOx, euro4, gaso\_1.4l}} \times L_a, \tag{6.7}$$

where N_a is the total number of vehicles on link a computed by LADTA or Meta-LCF, L_a is the length of link a (in km) and $e_{\text{NOx, euro4, gaso\_1.4l}}$ is the corresponding HEF. The total emissions on the link a is then the sum of computed hot emissions of all types of PCs with all different technologies. The complete modeling chain from traffic assignment to emission estimation is referred to as LADTA-COPERT IV, and Meta-LCF-COPERT IV.

6.3.2 Results

NOx emissions on a working Tuesday

A dynamic traffic assignment (DTA) simulation is carried out for a working Tuesday. During each time interval of 15 minutes, an emission map of NOx is plotted. Figure 6.1 shows the temporal evaluation of NOx (in g/km/15min) from 07:00 to 08:30.

Figure 6.2 compares the computed emission and traffic flow during the period of 07:00 to 07:45. We can see that the emission is mainly influenced by the traffic flow on the link. The spatial distribution of NOx emissions is highly correlated with that of the traffic flow. In fact, based on Equation 6.2, the street-resolution hot emission is the production of (i) hot emission factor and (ii) traffic volume on the street. Figure 6.2 shows that the traffic volume on the street influences more significantly the computed NOx emission than the emission factor, with fixed vehicle fleet composition data on the network.

Annual emissions of NOx and CO

With traffic counts collected by detectors, the temporal variation of on-road traffic in Clermont-Ferrand can be obtained for all kinds of days (working days, weekends, public holidays and school vacations) during any periods. A whole-year traffic assignment simulation with Meta-LCF was carried out. The computed traffic flows and average speeds at street level were then coupled with COPERT IV model. Together with the default vehicle fleet data listed in Table 6.1 and 6.2, a whole year of hot emissions for PCs were estimated. Figure 6.3 presents the temporal variation of the daily vehicle counts, total daily on-road traffic (hot) emissions of PCs for NOx and CO during 2014.

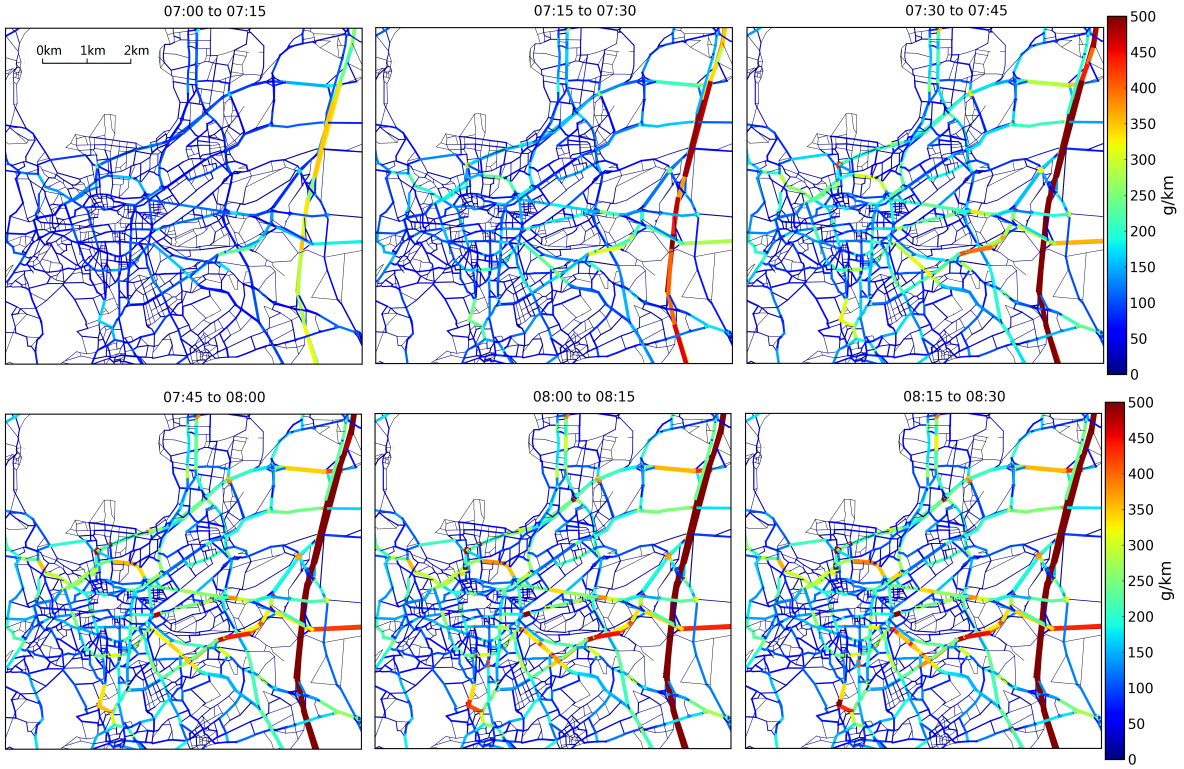


Figure 6.1: Emission maps of NOx (in g/km/15min) in the City of Clermont-Ferrand from 07:00 to 08:30. Black lines represent links with zero emission (due to zero computed traffic flow). The line width is proportional to the NOx emission on the link. For links with the same origin-destination nodes, the sum of their NOx emissions is calculated and displayed in these maps.

Figure 6.3 shows that there is a strong correlation between the traffic volume and computed emissions. During vacations, weekends and public holidays when there are less traffic, the resulting on-road emissions decrease as well. The computed hot emissions of CO are less than those of NOx. This is due to the large proportion of diesel PCs in France ($> 70\,\%$) and Euro 3-4 PCs ($> 70\,\%$ in total) (shown in Table 6.1 and 6.2). For PCs, the HEFs of CO are smaller than those of NOx, especially for Euro 3 and Euro 4 PCs (shown in Figure 6.4).

6.4 Sensitivity analysis

A qualitative local sensitivity analysis (LSA) with the *one-at-a-time* method is carried out on the modeling chain. In the LSA study, the simulation results in Section 6.3.2 are considered as references. The LSA studies the sensitivity of the computed on-road

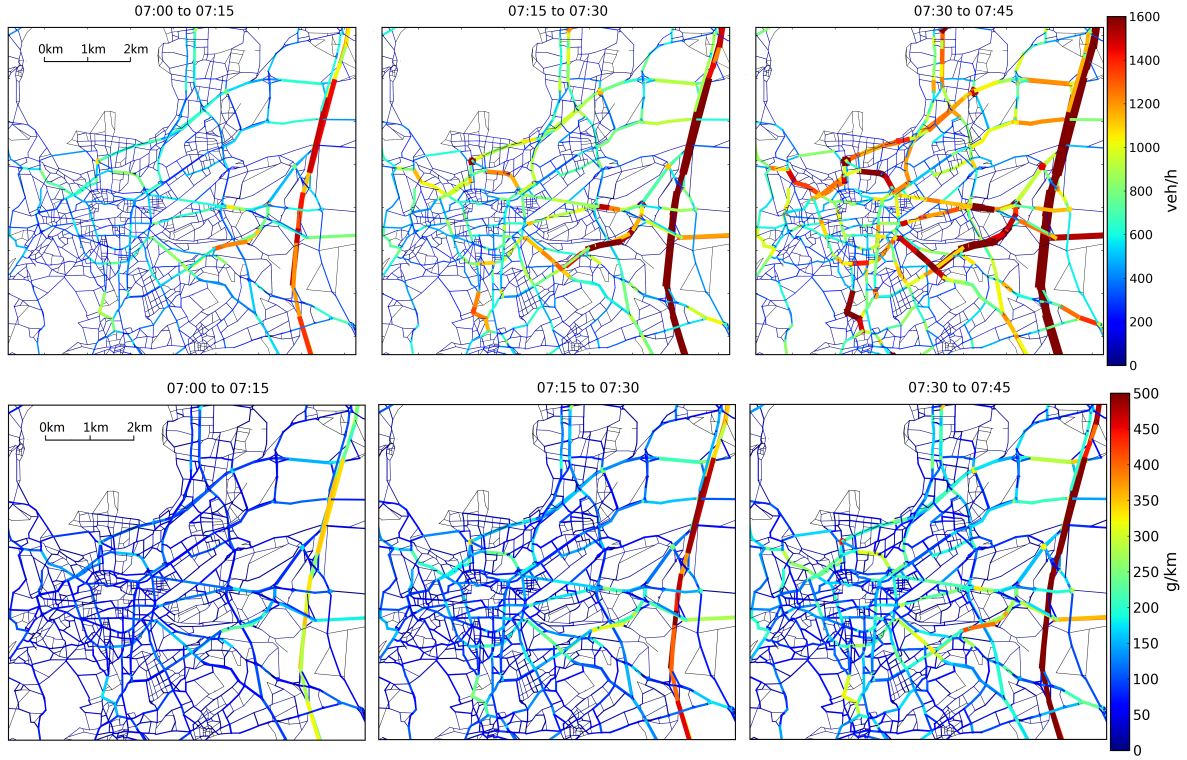


Figure 6.2: Comparison of traffic flow (upper three figures) and emission of NOx (lower three figures) in the city of Clermont-Ferrand from 07:00 to 07:45. Traffic flows (in $\mathrm{veh\,h^{-1}}$) are computed by LADTA model. NOx emissions (in $\mathrm{g\,km^{-1}}$) are computed by the modeling chain. Black lines represent links with zero traffic or emission. The line width is proportional to the traffic flow or to the NOx emission on the link. For the links that share the same nodes but are oriented in opposite directions, the sum of the traffic flows (or of the NOx emissions) from the two directions is calculated and displayed.

traffic emissions of NOx, with respect to the LADTA inputs as well as the inputs of COPERT IV model: (i) total traffic demand, (ii) speed limits of the road network, and (iii) vehicle fleet composition. In addition, with the modeling chain of Meta-LCF-COPERT IV, the model evaluation time for estimating street-resolution emission has been significantly decreased. Then the global sensitivity analysis with Sobol' method is carried out to study the sensitivity of on-road traffic emissions computed by the modeling chain in the whole agglomeration at street resolution, with respect to both (i) the uncertain inputs of LADTA model, and (ii) the vehicle fleet composition inputs for COPERT IV. The reference values for the inputs are those used in the case of Section 6.3.

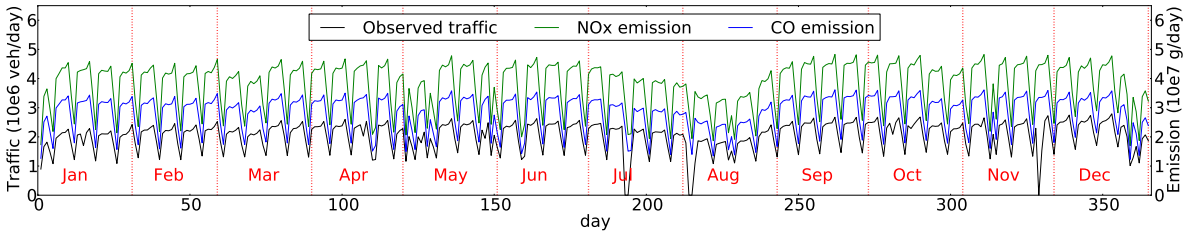


Figure 6.3: Daily observed on-road traffic (black, in 10^6 veh/day), total daily NOx (green) and CO (blue) emissions (in 10^7 g/day) for PCs estimated by Meta-LCF-COPERT IV, during 2014. For each day, the traffic data is the sum of observed vehicle counts of all detectors. The total daily emission is the sum of street-resolution emission computed by the modeling chain Meta-LCF-COPERT IV at all links and all time intervals. The value reaches zero when no data was collected. On these days, it was assumed that the values of emissions were the same as those of the same weekday in the previous week (except that if the day in the previous week is a public holiday, and then the value of the same weekday of the next week was taken).

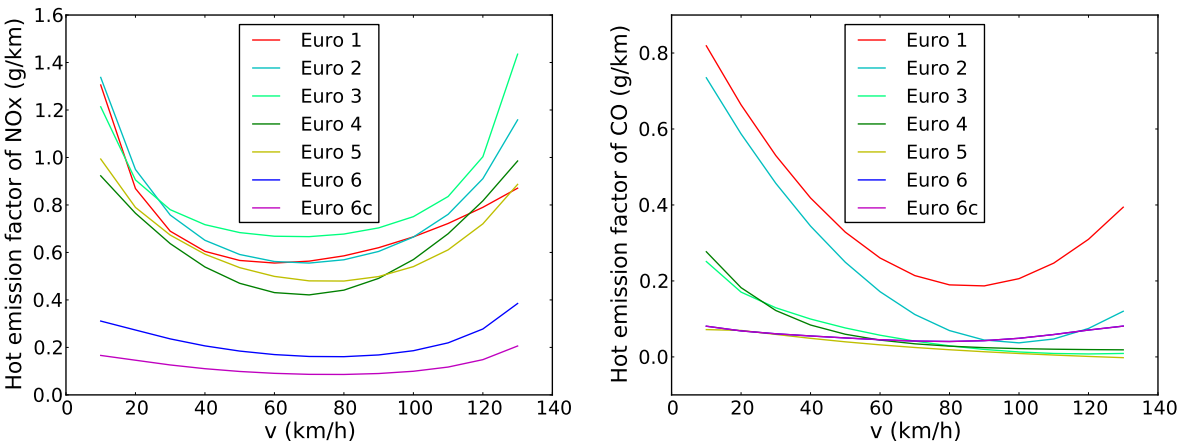


Figure 6.4: Hot emission factors of (a) NOx and (b) CO for diesel passenger cars for different emission standard.

6.4.1 Qualitative sensitivity analysis results

Total traffic demand on the whole network

We change the total demand volume by 50 % to 150 % with respect to the reference demand. The temporal variation of traffic demand remains the same: the temporal profile for a working Tuesday. The data for links' information and vehicle fleet composition remain unchanged. Then the emissions are calculated for each link during each

15 minutes for the working Tuesday with different total traffic demands.

Figure 6.5 shows the spatial distribution of computed traffic flow and NOx emissions during the period from 07:45 to 08:00, with a total traffic demand volume of −50 %, 0 % and 50 % with respect to the reference demand. Results in Figure 6.5 show that NOx emission increases with the increase of travel demand, especially on links whose traffic flow is sensitive to the increase of the total demand.

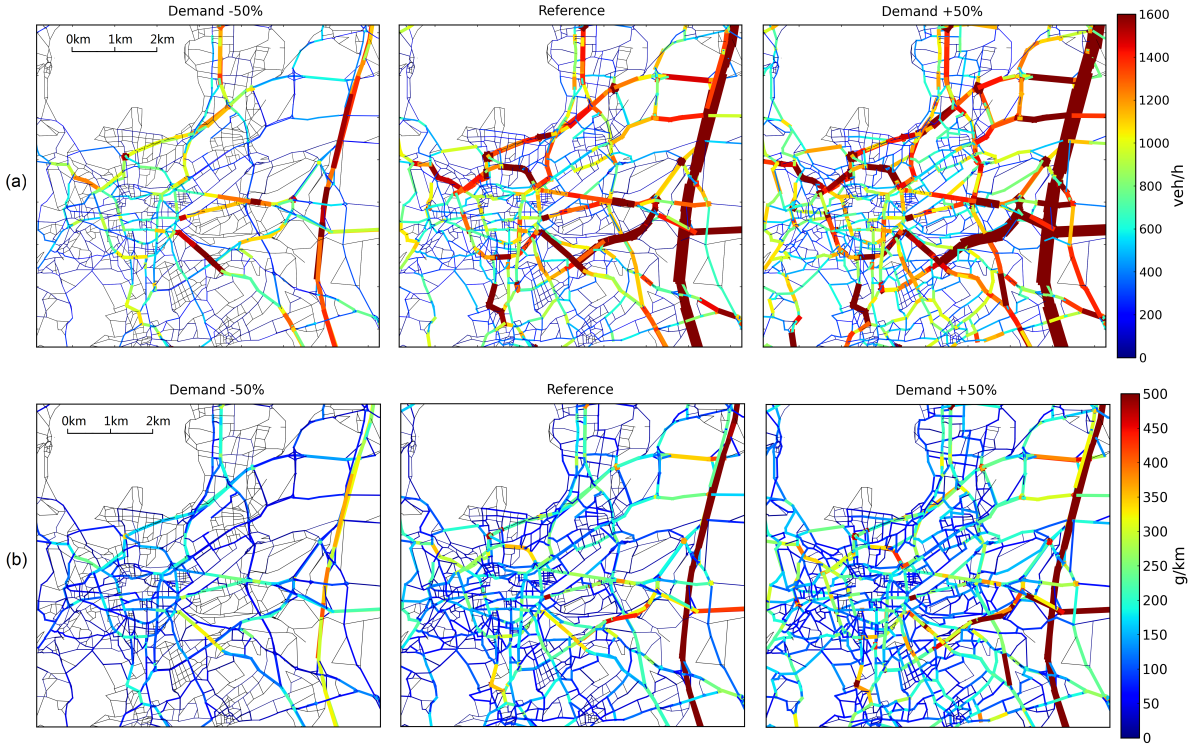


Figure 6.5: The spatial distribution of (a) traffic flows (in veh h^{-1}) and (b) NOx emissions (in g km^{-1}) from 07:45 to 08:00, with the total demand volume of −50 %, 0 % and 50 % of the demand in the reference case. Black lines represent unused links in LADTA model. The line width is proportional to link's emission. For the links that share the same nodes but are oriented in opposite directions, the sum of their NOx emissions is calculated and displayed.

In order to analyze the sensitivity in details, the total values of these indicators on the whole network during 24 h are calculated for each case with different total demand volumes. The sensitivity of total daily NOx emissions with respect to total traffic demand is presented in Figure 6.6.

Figure 6.6 shows that the influence of the total volume to the total daily NOx emission in the whole agglomeration of Clermont-Ferrand is not linear. The total emissions

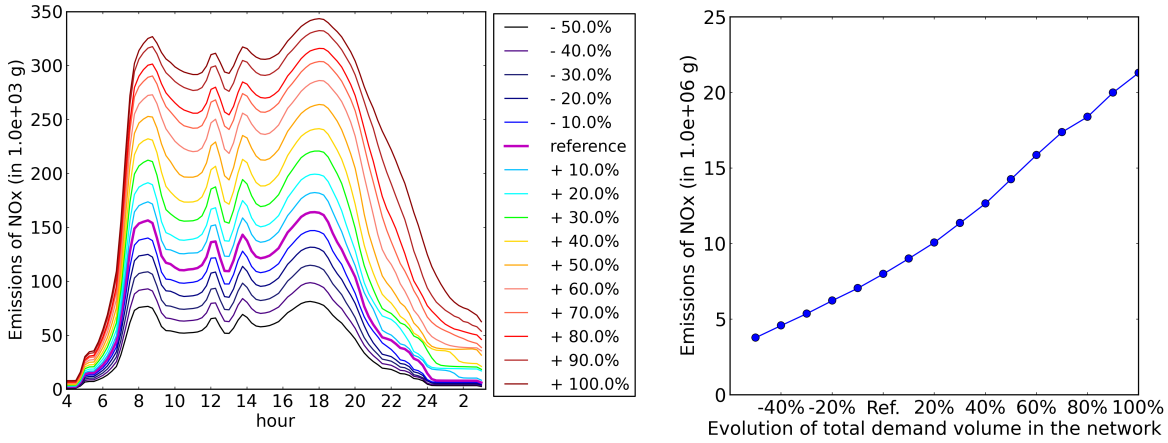


Figure 6.6: Sensitivity of the total emissions of the network to the total demand volume. Left: emission evolution (in 10^3 g) for the whole day, right: evolution of the total vehicle travel time in the whole network (in 10^6 g).

are tripled when we double the total demand volume comparing with the reference case.

Speed limits of the network

This sub-section analyzes the sensitivity of traffic emission to the road's average speed. The maximum authorized speed of each road has been decreased by 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 % and 50 %. It is found that the network total emissions are less sensitive to the speed limitation than to the total traffic demand, as shown in Figure 6.7. Figure 6.7 (right) shows that the variation of the total emissions is not monotonic with the decrease of speed limitation. A minimum is reached for a decrease of 25 % of the speed limits on all links on the network.

In fact, the hot emission factors of NOx change with the average travel speed as shown in Figure 6.8. With fixed traffic demand, the computed emissions follow the same non-monotonic trend as in Figure 6.8, and the variation of the amplitude depends on the vehicle fleet.

Vehicle fleet composition

(1) Effects of vehicle emission Euro standards

In this subsection, we divide passenger cars into two big categories: (i) the "Euro 4 +" vehicles with Euro standards of Euro 4 and higher, (ii) the "Euro 3 –" vehicles with Euro standards from pre-Euro to Euro 3. In the reference simulation, the percentage of "Euro 4 +" vehicles is about 42.3 % for passenger cars. For the sensitivity study, we

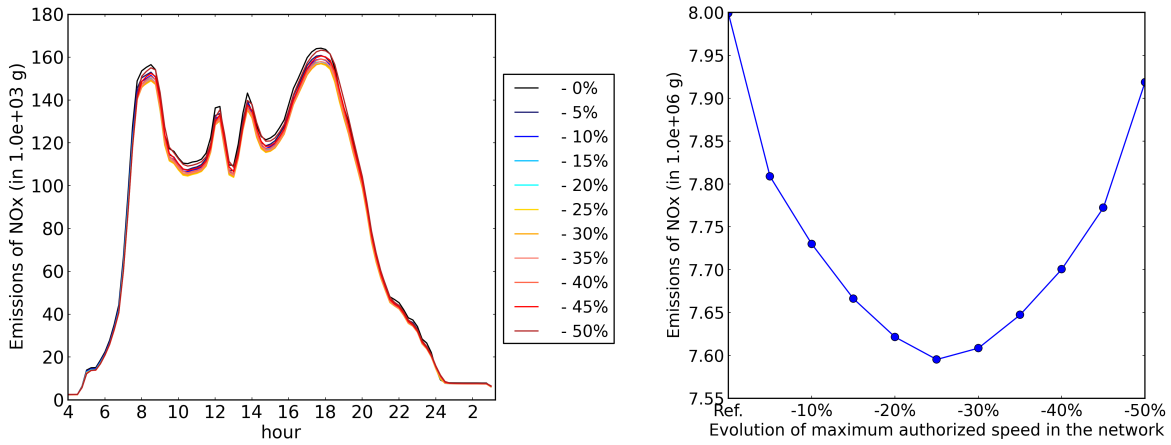


Figure 6.7: Sensitivity of hot emission factor of passenger cars for NOx to the average speed.

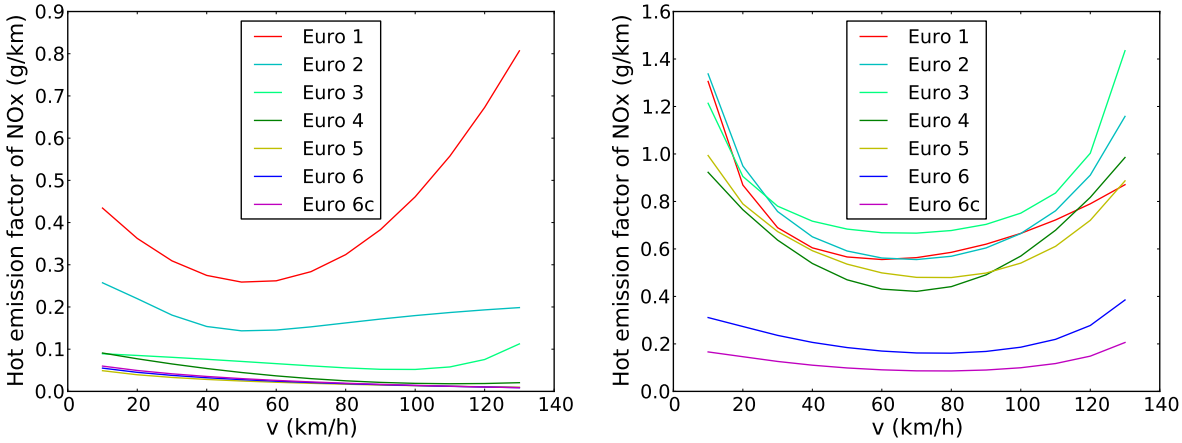


Figure 6.8: Variation of emission factors (in $\mathrm{g\,km^{-1}\,veh^{-1}}$) of NOx for gasoline passenger car (left) and diesel passenger car (right) against the increase of average travel speed.

increase the percentage of "Euro 4 +" cars from 42.3 % to 80 %. Within each category, we keep the same distribution of car standards. Figure 6.9 (left) shows that the total daily emissions of NOx decrease when the proportion of "Euro 4 +" category increases.

(2) Effects of the proportion for gasoline passenger cars and diesel passenger cars

In the reference case, the percentage of diesel passenger cars is 70.6 % and that of gasoline passenger cars is 29.2 %. We increase the proportion of gasoline passenger cars from 30 % to 70 % with steps of 10 %. At the same time, the sum of the percentage of diesel passenger cars and gasoline passenger cars remains 1.0. The results in Fig-

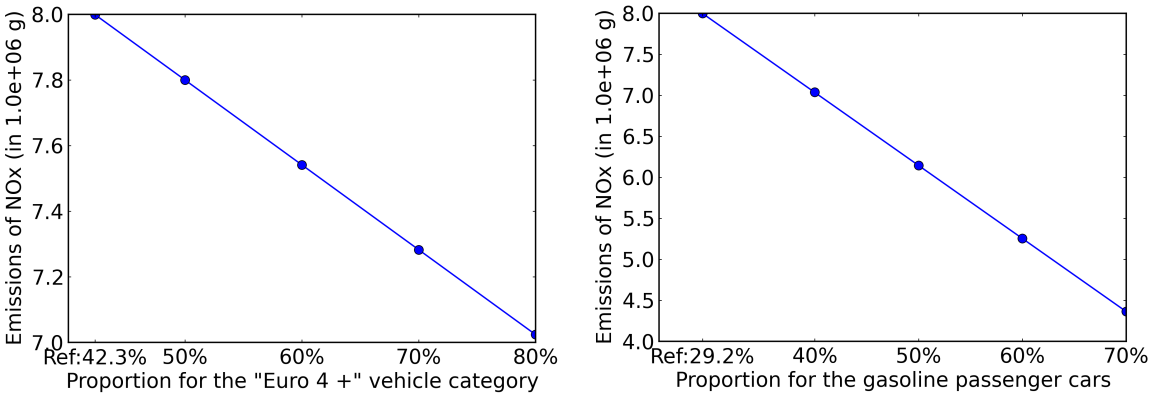


Figure 6.9: Sensitivity of total emissions of NOx to Left: Euro Standard and right: vehicle fleet type.

ure 6.9 (right) show that we can decrease the total NOx emission of 45 % by doubling the percentage of gasoline passenger cars.

6.4.2 Global sensitivity analysis results: Sobol' indices

For computing global sensitivity using Sobol' method presented in Chapter 5, COPERT IV model is coupled with Meta-LCF model built in Chapter 4 for computing traffic flows and average speeds at street resolution. The outputs of the Meta-LCF and the on-road emission modeling chain are the computed flow $\boldsymbol{y}(h_{\text{simu}})$, average speed $\boldsymbol{V}(h_{\text{simu}})$ and emission $\boldsymbol{E}_i$ (in *gram*) at link level during $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$. The emission modeling chain Meta-LCF-COPERT IV is built. According to Equation 6.1, the number of vehicles, the average speeds and vehicle fleet data can all affect the hot emissions computed by the modeling chain. For the global sensitivity analysis (GAS), we focus not only on inputs of COPERT IV model, but also on the 23 traffic assignment (TA) inputs in order to analyze the propagation of uncertainty from the traffic model inputs to the computed emissions. In total, 29 inputs for the whole modeling chain are studied, including 23 inputs of Meta-LCF model, the five inputs presented in Section 6.3.1, and an additional input concerning HDVs. In addition, the GSA study here aims at investigating how the computed emissions *on each link* are influenced by all 29 inputs. For each link, the computed emissions can be considered as scalar output of the modeling chain with 29 inputs: $E_a = g(p_1, p_2, \ldots, p_i, \ldots, p_{29})$. With respect to the same inputs, the sensitivity of the computed output on different links may be different. The first-order and total-effect sensitivity indices ($S_{i,a}$ and $S_{Ti,a}$) for traffic flows, speeds and emissions are therefore computed at all links. Table 6.3 presents the inputs

$\boldsymbol{p} = (p_1, p_2, \ldots, p_i), i \in \{1, 2, \ldots, 23\}$ of Meta-LCF and their variation intervals.

Table 6.3: The variation intervals of Meta-LCF model inputs

Input	Temporal profile	Capacity	Speed	Demands in O-D matrix	Transposed O-D matrix
Symbol	$(P(h))_{h\in\mathcal{H}_{\text{atomic}}}$	$\lambda_{\text{big}}, \lambda_{\text{small}}$	$\mu_{\text{high}}, \mu_{\text{low}}$	$\delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15}$	η
Number	13	2	2	5	1
Variation	$[0.00, 1.5]$	$[0.70, 1.30]$	$[0.80, 1.20]$, $[0.70, 1.10]$	$[0.25, 1.50]$	1 or 0

Table 6.4 presents six inputs of vehicle fleet parameters and their variation intervals. Besides the five parameters defined in Section 6.3.1, we added an input to represent the share of Heavy Duty Vehicles (HDVs) on a group of roads passing through Clermont-Ferrand and near an industrial logistic center (shown in Figure 6.10). Since the traffic flows modeled by Meta-LCF represent only PCs, extra HDV are added on certain links, with the ratio σ_{HDV} between HDVs/PCs varying from 0 % to 30 %. In COPERT IV model, there are many parameters for computing HEFs of HDVs. In our case study, only one type of HDV was added on given links: 100 % charged articulated diesel track of 28 −32 t, with emission standard of Euro IV. It is assumed that the slope is 0 % over all roads. The outputs of an *atomic* simulation with Meta-LCF-COPERT IV are the emissions of NOx and CO at link level of the whole agglomeration, during the time interval of $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$.

Table 6.4: The vehicle fleet inputs for COPERT IV model and their variations in the Monte Carlo simulations for GSA

Input	θ_{gaso}	$\gamma_{\text{gaso\_1.4}}$	$\varepsilon_{\text{diesel\_2.0}}$	$\zeta_{\text{gaso\_euro4}}$	$\varphi_{\text{diesel\_euro4}}$	σ_{HDV}
Variation (%)	$[10, 100]$	$[10, 100]$	$[10, 100]$	$[10, 100]$	$[10, 100]$	$[0, 30]$

The GSA results are represented by plotting sensitivity indices in two kinds of figures: boxplots and maps. Firstly, the boxplots allow us to compare the influence of all input factors among each other, throughout the whole network. Each point in a boxplot corresponds to $S_{a,i}$ or $S_{T_{a,i}}$ on a link a for an input factor p_i. The box contains the indices between the 1^{st} quartile (Q_1) and the 3^{rd} quartile (Q_3). The interquartile range (IQR) is Q_3 - Q_1. The red line corresponds to the median value of $S_{a,i}$ or $S_{T_{a,i}}$. The Whisker parameter is set to 3.0: a value $< Q_1 - 3.0 \times IQR$ or $> Q_3 + 3.0 \times IQR$ will be considered as *out of range* and plotted as points. Secondly, since the sensitivity indices are calculated at link level for each factor p_i, a map of all $S_{a,i}$ for the whole network

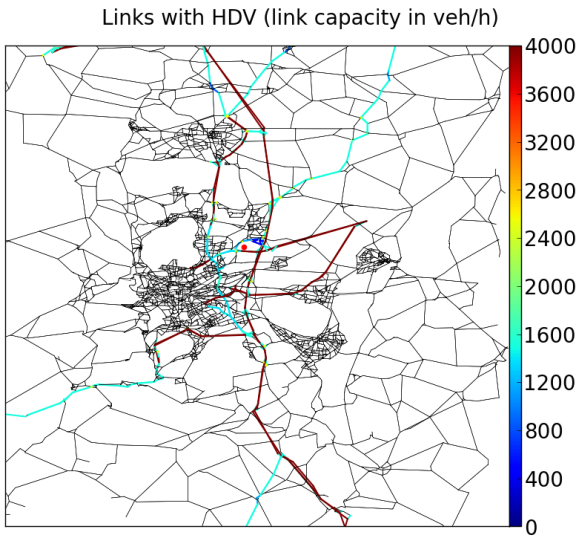


Figure 6.10: Map of links with HDVs, and link capacity map of the whole agglomeration of Clermont-Ferrand. The red point in the map is a big industrial logistic center. Black lines represent roads without HDVs. Lines with colors represent roads with HDVs. The color and width of links represent the link capacity (in veh h^{-1}).

of Clermont-Ferrand is displayed. The maps help us to visualize how the input factors can influence the computed emissions depending on the link spatial locations.

NOx emissions

Figure 6.11 presents $S_{i,a}$ and $S_{Ti,a}$ of the computed NOx emissions at all links, with respect to the total 29 inputs in Table 6.3 and Table 6.4 of modeling chain Meta-LCF-COPERT IV. Classical Monte-Carlo simulations of the modeling chain were applied by taking each input factor p_i randomly in its variation interval. The total sampling size is $n = 80000$.

Figure 6.11 shows that the computed NOx emissions are sensitive to all the factors that are influential to the computed traffic flows mentioned in Chapter 5 (see Figure 5.2). Though the average speed is a parameter of the COPERT IV model, the computed NOx emission is not significantly sensitive to it. The variation of travel speeds does not contribute much to the variance of NOx emissions on most of the streets. Concerning vehicle fleet composition, Figure 6.11 shows that the NOx emissions are not very sensitive to the emission standard technology. This might due to the assumption that no PCs of Euro 6 standard technology is taken into account, while according to HEFs of NOx (Figure 6.4), the HEFs decrease significantly for Euro 6 diesel PCs compared with other emission standard classes. Figure 6.11 also shows that the NOx emissions are

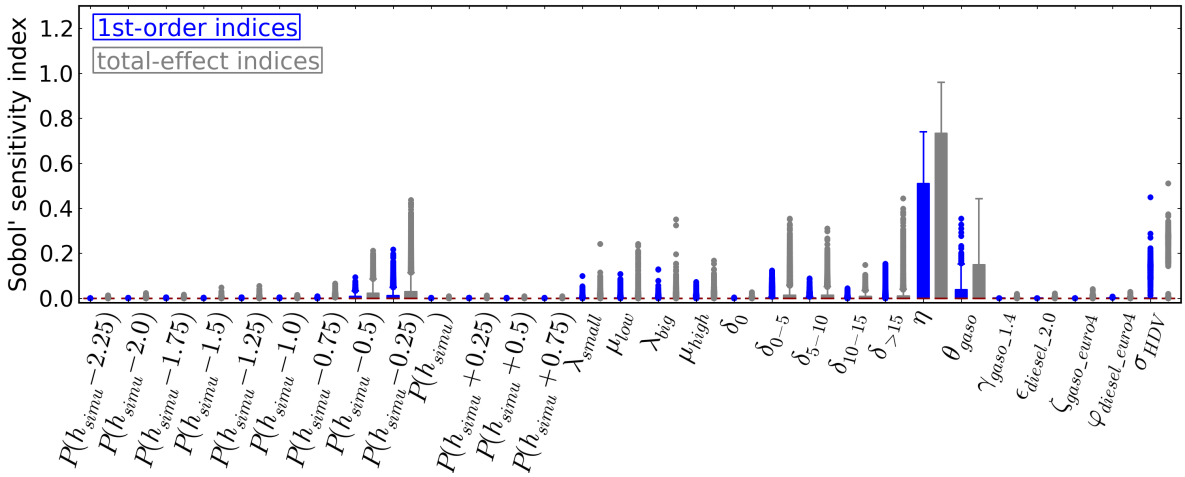


Figure 6.11: First-order and total-effect Sobol' sensitivity indices of the 29 inputs of the modeling chain Meta-LCF-COPERT IV for the computed NOx emissions

very sensitive to the proportion of gasoline cars (θ_{gaso}), and to the HDV share (σ_{HDV}). On links with HDVs, the computed emissions are still significantly more sensitive to the variation of HDVs than other vehicle fleet inputs of the modeling chain. This is all the more noteworthy that σ_{HDV} varies only on a few roads and only up to 30 %. The spatial distribution of $S_{a,i}$ is presented in Figure 6.12. It shows that on links without HDVs, the computed NOx emissions are very sensitive to the proportion of gasoline PCs. While on links with HDVs, even if the variation of σ_{HDV} are small, the uncertainty of σ_{HDV} dominates the variance of the NOx emissions computed by the modeling chain.

CO emissions

We also carried out GSA to the computed CO emissions with respect to the 29 inputs of the modeling chain of Meta-LCF-COPERT IV. Figure 6.13 presents the first-order and total-effect sensitivities of the computed emissions (in *gram*) at link level, with respect to 12 factors that are the most influential factors to the computed flows and speeds (Chapter 5), and to the 6 vehicle fleet factors in Table 6.4 for COPERT IV model.

Figure 6.13 shows that the computed CO emissions are also sensitive to all the factors that are influential for the computed traffic flows mentioned in Chapter 5: traffic demand direction, traffic demand during the previous 30 minutes and traffic demand between short distance O-D pair. Concerning vehicle fleet inputs however, only θ_{gaso} has first-order effect on the computed CO emissions. Unlike in the case for NOx emissions, the total-effect indices in Figure 6.13 show that the HDV share (σ_{HDV}), within the designed

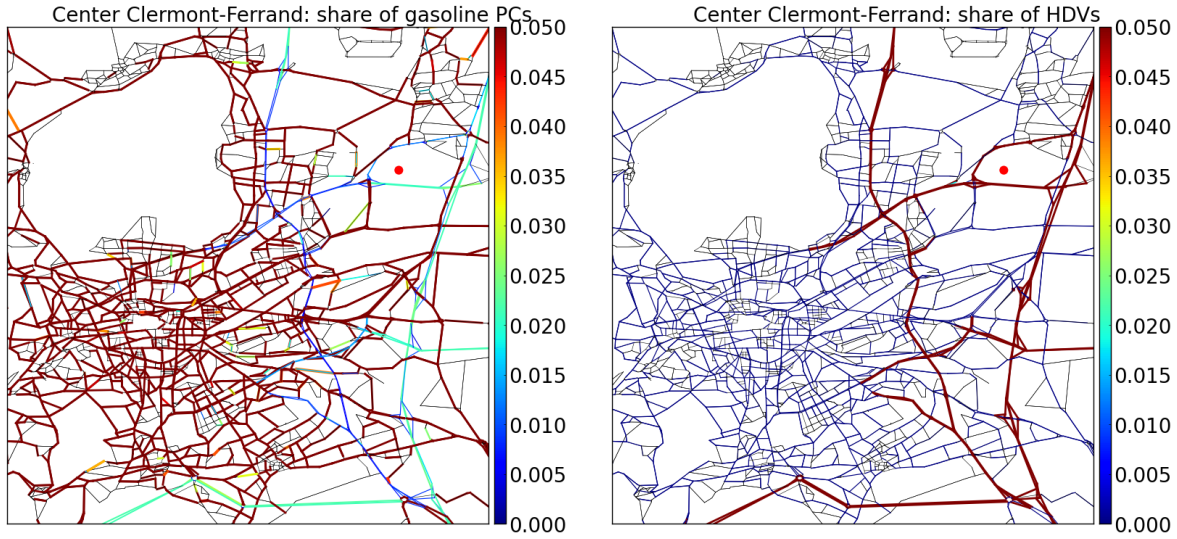


Figure 6.12: Spatial distribution of the 1st order sensitivity indices of the computed NOx emissions with respect to the share of gasoline passenger cars θ_{gaso} (left) and heavy duty vehicles σ_{HDV} (right). The red point in the map represents a big industrial logistic center.

uncertainty range of 0 to 30%, is not an influential factor to computed CO emissions. However, contrary to the emission of NOx, the hot emissions of CO are influenced by some high-order interactions involving the emission standards of PCs. The median values of the total-effect indices for $\zeta_{\text{gaso\_Euro4}}$ and $\varphi_{\text{diesel\_Euro4}}$ are not zero, and are bigger than those of NOx case in Figure 6.11.

Conclusions

In this chapter, we have presented a simulation chain for emission estimations combined with a dynamic traffic assignment model LADTA, and the COPERT IV model. The case study was carried out for a typical working day during non-vacation period for the agglomeration of Clermont-Ferrand. The emission calculations are based on traffic flow results of LADTA for each link during every 15 minutes for the day. Results show that high emissions are found on roads with heavy traffic, especially during peak hours. We have already seen in Chapter 3 that traffic flow changes significantly within a quarter of an hour. The use of a dynamic traffic assignment (DTA) model allows us to estimate emissions down to the same time resolution. Moreover, since a metamodel of the LADTA model applied to the agglomeration of Clermont-Ferrand has already been

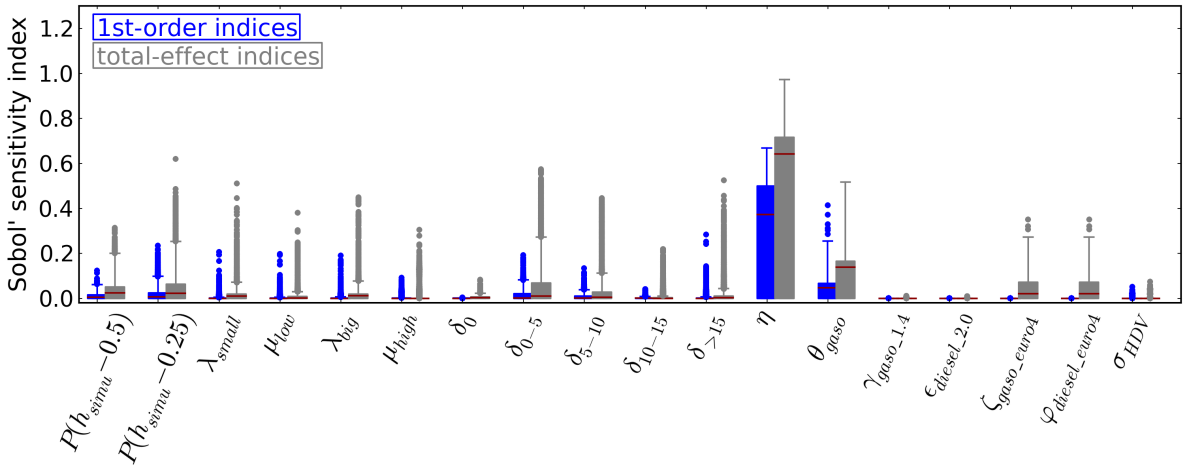


Figure 6.13: First-order and total-effect Sobol' sensitivity index of some inputs of modeling chain of Meta-LCF-COPERT IV, for the computed traffic emissions of CO

built, a modeling chain combing the metamodel of traffic model and the COPERT IV model was also built in this chapter. At last, we studied the sensitivity of the emissions to LADTA model and COPERT inputs.

According to the qualitative sensitivity analysis results, total emissions increase with the total demand volume onto the network. However, the variation is not linear. The variation of the link speed has less influence on the network total vehicle travel time and total emissions. As we decreased the links' speed limits, the total emissions firstly decreased and then increased. The minimum value has been reached when the speed decreased by about 25 %. This is mainly due to the influence of vehicle travel speed on the hot emission factors of NOx.

The global sensitivity analysis (GSA) is carried out to the modeling chain combining the metamodel of LADTA applied to Clermont-Ferrand and the COPERT IV model: Meta-LCF-COPERT IV. Hot exhaust emissions are computed by multiplying hot emission factors (HEFs) and the number of vehicles at street level. The hot emission factors mainly depend on vehicle speed and technology. HEFs were modeled by COPERT IV in our case study. The output emissions are then affected by inputs of the modeling chain via traffic flows and HEFs. The results of GSA show that the computed NOx emissions are sensitive to all the inputs that influence the traffic flows. The traffic speeds do not have significant influence to the NOx emissions computed by Meta-LCF-COPERT. In addition, the computed NOx emissions are very sensitive to some of the vehicle fleet inputs: (i) the share of HDVs and (ii) the percentage of gasoline cars on road. Other fleet parameters such as the emission standards and engine capacity, are non-influential

factors to the on-road NOx emissions simulation. Nevertheless, these sensitivity results are different for different pollutants. For the hot emissions of CO, only the gasoline PCs share has a first-order effect to the emissions computed by the modeling chain. The share of HDVs is not very influential to the computed CO emissions, compared with the case study for NOx emissions. Moreover, the hot emissions of CO are more affected by high-order effects of emission standard factors (the share of Euro 4 and higher PCs in vehicle fleet composition) than in the case of NOx.

Chapter 7

Uncertainty quantification for dynamic traffic assignment and on-road emission simulations at metropolitan scale

---

Summary

The uncertain inputs of the dynamic traffic assignment (DTA) simulation with LADTA are considered as random variables with uniform probability distributions. We then use the Monte Carlo approach to generate two ensembles of simulations: (i) an ensemble of DTA simulations with the metamodel of LADTA, and (ii) an ensemble of emission simulations, for the agglomeration of Clermont-Ferrand at street resolution and for all time intervals of 15 minutes during November 2014. The objective of uncertainty quantification is ideally to derive an estimation of the probability distribution of the computed outputs. For DTA simulations, the ensemble is evaluated with the help of observations of traffic flows, measured by loop detectors on the network of Clermont-Ferrand at all time steps during November 2014. Rank histogram, reliability diagram and statistical scores are chosen criteria to evaluate whether or not the ensemble can well represent the uncertainty of the DTA simulations. Then the uncertainty quantification is carried out based on the ensembles. Standard deviation, relative standard deviation, interquartile range (IQR) and $5^{th} - 95^{th}$ percentile range are used to measure the uncertainty of the outputs. For traffic flow prediction, high absolute uncertainty (measured by the standard deviation) is found on streets with heavy traffic. In addition, high IQR and $5^{th} - 95^{th}$ percentile range are found on roads with heavy traffic volume, too. However, the high relative uncertainty (measured by the relative standard deviation) is mostly found on streets where the simulated traffic flow is low. For on-road traffic emissions, an ensemble of NOx emissions is built based on the ensemble of DTA simulations, using the modeling chain that couples

LADTA and COPERT IV at street resolution. We then analyze the uncertainty of computed emissions due to the uncertainty in inputs of the dynamic traffic assignment. Results show that the emission uncertainty is highly correlated with (i) the computed traffic flow on the network and (ii) the uncertainty of traffic flows. In addition, the uncertainty tends to be amplified from the computed traffic flow to the computed emission, especially on streets with heavy duty vehicles.

Résumé

On considère que les entrées incertaines de l'affectation dynamique du trafic (ADT) avec le modèle LADTA sont des variables aléatoires avec distributions uniformes. Nous utilisons ensuite l'approche Monte Carlo pour générer deux ensembles de simulations: (i) un ensemble de simulations ADT avec le méta-modèle de LADTA, et (ii) un ensemble de simulations d'émissions, pour l'agglomération de Clermont-Ferrand à la résolution de rue, pour tous les intervalles de temps de 15 minutes en novembre 2014. L'objectif de la quantification d'incertitude est idéalement d'obtenir les distributions de probabilité des sorties. Pour les simulations d'affectation du trafic, l'ensemble est évalué à l'aide de données de comptage du trafic, collectées sur le réseau de Clermont-Ferrand pendant la même période que la simulation. Le diagramme de rang, le diagramme de fiabilité et des scores statistiques sont les critères utilisés pour évaluer si l'ensemble représente correctement l'incertitude des sorties. Ensuite, la quantification d'incertitude est effectuée grâce aux sorties des deux ensembles. L'écart type, l'écart type relatif, l'écart interquartile (EI) et l'écart entre les 5e et 95e centiles sont deux mesures choisies pour décrire l'incertitude des sorties. Pour l'affectation du trafic, une forte incertitude absolue (mesurée par l'écart type) se trouve principalement dans les rues où le débit du trafic simulé est grand. De grands EI et écarts entre les 5e et 95e centiles se trouvent également dans les rues avec forts débits de trafic. Cependant, une forte incertitude relative (mesurée par l'écart type relatif) se trouve principalement dans les rues où le débit du trafic simulé est faible. Pour les émissions dues au trafic routier, un ensemble d'émissions de NOx est construit à partir des sorties (débits et vitesses moyennes) de l'ensemble d'affectations du trafic. Nous analysons ensuite l'incertitude des émissions simulées due aux entrées incertaines de l'ADT. Les résultats montrent que l'incertitude des émissions est fortement corrélée avec (i) le débit du trafic simulé sur le réseau et (ii) l'incertitude du débit. De plus, l'incertitude a tendance à s'amplifier depuis le trafic simulé jusqu'aux émissions en sortie, en particulier dans les rues où les poids lourds sont présents.

Contents

7.1 Introduction

The uncertainty quantification aims at quantifying the uncertainty of outputs resulting from numerical models. For modeling atmospheric pollutant emissions of on-road traffic, there are many uncertainty sources in inputs, in the model parameterizations, and in numerical approximations. In our work, we only study the uncertainty due to model inputs for estimating on-road traffic emissions. Before a detailed discussion, we define here more precisely what is *uncertainty* and *uncertainty quantification*. Unlike the deterministic simulation with certain known inputs, the inputs and outputs are both considered as random variables U and X. Let us denote the model as $X = f(U)$ and f is a numerical model for estimating traffic flows or on-road traffic emissions. A *deterministic* simulation can then be denoted as $x = f(u)$ where x and u are *realizations* of random variables X and U. For simplification, we assume that the output variable X follows a normal distribution with mean $\bar{X}$ and standard deviation σ:

$$X \sim \mathcal{N}(\bar{X}, \sigma^2). \tag{7.1}$$

The *true* value of X is an observed value o. If we consider that there is uncertainty in the observation, the observation can also be considered as a random variable O. However, here we assume that the observation is perfect and we ignore the error lying in observations. Therefore, in deterministic simulation, we can evaluate the model f with the *error* e. The error is defined as the distance between a realization of X and the observation o:

$$e = d(x, o), \tag{7.2}$$

where $d(\cdot, \cdot)$ is a measure of distance (Mean Squared discrepancy for example). Another criterion to evaluate the performance or the robustness of the numerical model is the *a priori uncertainty*: $\sigma_X = \sigma$. It is defined only from the *a priori* probability distribution of X without taking into account the observations. There is also *a posteriori uncertainty* which estimated given the observation o: $\sigma_{X|o}$. In our study, we only focus on the *a priori uncertainty* and the term "uncertainty" in this chapter refers to *a priori uncertainty*. Therefore, the objective of uncertainty quantification is to determine the probability distribution of X and to quantify σ. To do so, we need to use a probabilistic approach and introduce uncertainties in inputs U, in order to obtain the probabilistic distribution of output X computed by $f(U)$. This requires a large number of model evaluations and is often computationally costly.

If the *error* can already evaluate the performance of the model f with deterministic

simulations, why do we make use of probabilistic approach to quantify the uncertainty which is computationally costly? In fact, if we consider that f is a traffic assignment model at metropolitan scale, the output X is the traffic flow (or volume) assigned on all the roads of the network. However, only some roads are equipped with loop detectors. There is no available observations on all the roads and errors cannot be computed for the whole network. Instead, the measure of uncertainty does not depend on the observation so that uncertainty estimation can be carried out on all the roads of a network. In addition, the deterministic simulations do not take into account the uncertainty in inputs. In many simulations, the inputs are themselves obtained by other numerical models. For example, in a dynamic traffic assignment (DTA) model, one of the main inputs is an Original-Destination matrix (O-D matrix), representing the traffic demands on the network. This O-D matrix is modeled from household surveys. The road capacities and speed limits are also estimated from observations and we are not sure about their *true* values. Therefore, with probabilistic approach, we are able to give a possible distribution of the inputs U, and then quantify the uncertainty σ in output X by estimating its probability distribution, on all the roads of the network in question.

In this chapter, we firstly quantify the uncertainty of the DTA model applied to the agglomeration of Clermont-Ferrand. Since the probabilistic approach requires a large number of model evaluations, the metamodel Meta-LCF built in Chapter 4 replaces the original LADTA model for uncertainty quantification. In this case, f is the Meta-LCF model $\widehat{\mathcal{M}}$ and the input variable U is $\boldsymbol{p}$. The output X corresponds to $\boldsymbol{y}$ in Chapter 4 and it is time-varying with dimension $D = 19628$. Then we estimate the uncertainty of computed on-road traffic emissions. This time we considered the model f as the whole modeling chain Meta-LCF-COPERT built in Chapter 6. The output X is the estimated on-road traffic emissions on all roads of the network. To estimate the uncertainty of traffic flow and emissions, two ensembles are generated: an ensemble of DTA simulations using Meta-LCF model, and an ensemble of on-road emission estimations.

This chapter is organized as follows. We firstly generate a large ensemble of DTA simulations using Meta-LCF model, with $\boldsymbol{p}$ of known probability distribution as input. The ensemble generation is presented in Section 7.2. Secondly, we use loop detector measurements to evaluate the generated ensemble in Section 7.3. Then, we use the same input distributions to generate the ensemble of on-road emissions by using the modeling chain Meta-LCF-COPERT. At last, the uncertainty quantification of DTA simulations and on-road traffic emissions is presented in Section 7.4.

7.2 Ensemble simulations for dynamic traffic assignment and on-road emission estimations

In this section, we use the Monte Carlo approach to generate an ensemble of DTA simulations with the metamodel of LADTA applied to the agglomeration of Clermont-Ferrand (Meta-LCF model). We consider each input of Meta-LCF as an independent random variable with uniform distribution. The boundaries of the distributions for the inputs are given based on the default inputs of a deterministic simulation, during the whole month of November 2014 in the agglomeration. In this section, we rebuilt a Meta-LCF model using the same methods as in Chapter 4 but with larger input variation intervals. Therefore, we firstly briefly recall the Meta-LCF model and the inputs for carrying out the deterministic simulation for the November 2014 in the agglomeration of Clermont-Ferrand. Then we present the simulation results of this extended Meta-LCF model. An ensemble of DTA simulations with this Meta-LCF model it then built based on Monte Carlo approach. Without specifications, all the results shown in this Chapter are obtained based on this extended Meta-LCF model with larger input variations, including the deterministic simulation results in this Section 7.2, the built ensemble and the corresponding evaluation (Section 7.3), and the uncertainty results (Section 7.4).

7.2.1 Default inputs and deterministic simulation

Metamodel of DTA simulations in Clermont-Ferrand and inputs for deterministic simulation

We recall the Meta-LCF model for the agglomeration of Clermont-Ferrand in Equation 7.3. It computes traffic flows $\boldsymbol{y}$ in veh h^{-1} at street resolution during a time step of $[h_{simu}, h_{simu} + \Delta h]$, $\Delta h = 0.25$ (h in hour) on a day d. The road network in the agglomeration of Clermont-Ferrand is modeled by oriented graph and the roads are represented by *links*.

$$\begin{aligned}\boldsymbol{y}_d(h_{simu}) &= \widehat{\mathcal{M}}(\xi_d P_d(h)_{h\in I_{\text{atomic}}}, \delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15}, \lambda_{big}, \lambda_{small}, \mu_{high}, \mu_{low}, \eta) \\ &= \widehat{\mathcal{M}}(\boldsymbol{p}),\end{aligned} \tag{7.3}$$

where $I_{\text{atomic}} = \{h_1, h_2, \ldots, h_{13}\} = \{h_{simu}-2.25\,\mathrm{h}, h_{simu}-2.0\,\mathrm{h}, \ldots, h_{simu}, \ldots, h_{simu}+0.75\,\mathrm{h}\}$. LADTA requires time-dependent traffic demand as input and computes time-varying traffic flow. We assume that the dynamic demand and the computed flow are

constant during an interval of 15 minutes. We also assume that the traffic flow computed during $[h_{simu}, h_{simu}+\Delta h]$ does not depend on the traffic demand during the period before $h_{simu}-2.25\,\text{h}$ and after $h_{simu}+1.0\,\text{h}$. $P_d(h)_{h\in I_{\text{atomic}}}$ are the *temporal variation coefficient* and its dimension is 13. They are ratios between the traffic demand during $[h, h+\Delta h]$ and during the evening peak of a given simulation day d. ξ_d is the *total demand coefficient* for adjusting the total traffic demand according to the day d. δ_0, δ_{0-5}, δ_{5-10}, δ_{10-15} and $\delta_{>15}$ are five *evening peak coefficients* to represent spatial uncertainty of the demand according to the Origin-Destination distance. $\lambda_{big}, \lambda_{small}, \mu_{high}, \mu_{low}$ represent the uncertainties on road capacities and speed limits. $\eta = 1.0$ if $h_{simu} <= 12$ and $\eta = 0.0$ otherwise.

For the deterministic simulation with Meta-LCF model in Clermont-Ferrand, the capacities and speed limits of the network are supposed to be unchanged: $\lambda_{big} = \lambda_{small} = \mu_{high} = \mu_{low} = 1.0$. Each $P_d(h)$ of the 13 $(P_d(h))_{h\in I_{\text{atomic}}} = (P(h_i))_{i=1,\dots,13}$ can be obtained as follows. We assume that the temporal variation of the traffic demand is the same as the temporal variation of the spatially-averaged traffic flows measured by the loop detectors on the network. This temporal variation on the day d is called *temporal profile* and denoted as $\mathcal{W}_d(h), h = 0.0, 0.25, 0.30, \dots, 23.75$ (in hour). The time step is always Δh and the dimension of $\mathcal{W}_d(h)$ is 96. At any simulation time h_{simu}, $P_d(h_{simu})$ can then be obtained by:

$$\begin{aligned} P_d(h_{simu}) &= \mathcal{W}_d(h_{simu})/(\frac{1}{4}\times\sum_h(\mathcal{W}_d(h)_{h\in\{17.0,17.25,17.50,17.75\}}) \\ &= \mathcal{W}_d(h_{simu})/\mathcal{W}_d^{peak}, \end{aligned} \tag{7.4}$$

where $\mathcal{W}_d^{peak}$ is the average traffic during evening peak hour on day d. For the deterministic simulation, the coefficients $P_d(h)$ depend and only depend on the *temporal profile* on day d. The *evening peak coefficients* are assumed to be 1.0 and there is no spatial difference in the O-D matrix: $\delta_0 = \delta_{0-5} = \delta_{5-10} = \delta_{10-15} = \delta_{>15} = 1.0$.

ξ_d in equation 7.3 can also be obtained from observed temporal profiles. It is obtained as follows. We firstly choose a reference day, the 20^{th} November 2014. The temporal profile of the reference day is $\mathcal{W}_{ref}(h)$, obtained by averaging the measured traffic flows on the network of all detectors during each interval of 15 minutes. Let $\mathcal{W}_{ref}^{peak}$ denote the average flow during evening peak hour on the reference day. Then for any day d, $\xi_d = \mathcal{W}_d^{peak}/\mathcal{W}_{ref}^{peak}$. As a result, all the coefficients regarding traffic demands can be calculated based on *temporal profiles*, obtained from historical traffic flow data measured by loop detectors on the network.

In order to carry out DTA simulations with Meta-LCF models during a long period,

we used 16 *typical profiles* $\mathcal{W}_{type}(h)$ of the following 16 "typical days": 7 days from Monday to Sunday during normal period, 7 weekdays during school vacation period, 1 public holiday during normal period, and 1 public holiday during school vacation period. Each $\mathcal{W}_{type}(h)$ is obtained by averaging the daily profiles of the same *day type*, during the period from September 2013 to August 2014:

$$\mathcal{W}_{type}(h) = \frac{1}{N_{type}} \sum_{d}^{N_{type}} \mathcal{W}_d(h), \tag{7.5}$$

for all the days d of the same day type, and N_{type} is the total number of days of the same type of day, from September 2013 to August 2014. The latter period is the *learning period* in our case study. The inputs are obtained from measurements during the learning period. The simulation and model evaluation are carried out with measured traffic data out of this learning period. With the learned *typical profiles*, the deterministic simulation is carried out during the period of November 2014. The observed traffic data during this period are used in the evaluation of the deterministic simulation and the ensemble evaluation. The different types of days during the simulation month are presented in Table 7.1.

Table 7.1: Type of days during November 2014

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Normal period	4	3	4	4	4	4	4
School vacation	0	0	0	0	0	0	1
Public holiday, normal	0	1	0	0	0	0	0
Public holiday, vacation	0	0	0	0	0	1	0

Metamodel with larger input variation intervals

As mentioned in Section 7.1, the uncertainty quantification aims at determining the probabilistic distribution of the computed traffic flows. To do so, we use Monte Carlo simulation of size n. This requires a large number of model evaluations. That is the reason why we built the metamodel in Chapter 4 in the first place, and then use it for probabilistic DTA simulation. However, the metamodel itself is built based on emulation of training points and training values, obtained from DTA simulations in Clermont-Ferrand with the original LADTA model. Therefore, the variation intervals for sampling the training points when building the Meta-LCF model actually delimit the boundaries of Meta-LCF inputs when generating the ensemble. In other words, the

boundaries of the uniform distributions, given to input variables for carrying out Monte Carlo simulations, cannot exceed the variation intervals given to the inputs when the Meta-LCF model was built. Before generating the ensemble presented in the following sections of this chapter, we have actually built an ensemble of simulations based on the Meta-LCF model in Chapter 4 in the first place. We will see in the Section 7.3.3 that this first ensemble does not satisfy the criteria for a *reliable* ensemble, when compared with the loop detector measurements. Therefore, in this chapter, we built an extended Meta-LCF with the metamodeling approach as shown in Chapter 4, but with larger input variation intervals for sampling the training sets. The sample size for building the extended Meta-LCF model is 4000, for each of the cases carrying out DTA simulation with LADTA before 12:00 and after 12:00. The variation intervals for the 22 inputs of this Meta-LCF model are presented in Table 7.2. The transposition parameter η is not an uncertain input since it has certain value of 1 or 0, if the Meta-LCF is used during the morning before 12:00 or during the afternoon. The same rule is used when carrying out the deterministic DTA simulation with the extended Meta-LCF model. Note that we considered the *total demand coefficient* ξ as scalar and did not give it a probabilistic distribution. It depends on the *type of day* for carrying out LCF model. In fact, the uncertainty in the total demand can already be parameterized by fives coefficients that represent the uncertainty in the O-D matrix. We gave large variation intervals to these five coefficients $(\delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15})$ in order to take into account (i) the spatial uncertainty lying in the O-D matrix with respect to different O-D pair categories, and (ii) the uncertainty in the total traffic demand from day to day.

Table 7.2: The inputs of original LCF model and their variations for building training sets to build the Meta-LCF model. Larger input variation intervals are presented in bold.

Input	Temporal profile	Capacity	Speed	Demands in O-D matrix
Symbol	$P(h_1), \ldots, P(h_{13})$	$\lambda_{\text{big}}, \lambda_{\text{small}}$	$\mu_{\text{high}}, \mu_{\text{low}}$	$\delta_0, \delta_{0-5}, \delta_{5-10}, \delta_{10-15}, \delta_{>15}$
Number	13	2	2	5
Initial variation	$[0.00, 1.5]$	$[0.70, 1.30]$	$[0.80, 1.20]$, $[0.70, 1.10]$	$[0.25, 1.50]$
Larger variation	**[0.00, 2.0]**	**[0.50, 1.50]**	**[0.50, 1.50]**	**[0.00, 1.80]**

Deterministic simulation results

The computed flows computed by the extended Meta-LCF model are compared with the loop detector measurements. The scores are also compared with the scores of the

Meta-LCF model built in Chapter 4. See Table 7.3.

Table 7.3: Scores of the one-month simulation during November 2014 of the traffic flows computed (i) by the complete model in Chapter 4 and (ii) by the metamodel with different input variation intervals when sampling the learning points for building Meta-LCF models. Computed flows are compared with loop detector measurements. (*The global scores are calculated for spatio-temporal flows at all detectors and time steps. The temporal scores are calculated for spatially-averaged flows for all time intervals. The spatial scores are calculated based on temporally-averaged flows on all detectors.*)

Model	Original model	Metamodel	
		Meta-LCF with larger possible input variations	Meta-LCF in Chapter 4
Global bias (veh h^{-1} $detector^{-1}$)	**18**	**5**	11
Global RMSE (veh h^{-1} $detector^{-1}$)	**154**	**142**	142
Global NRMSE (%)	**84.13**	**77.91**	77.77
Global Correlation	**0.69**	**0.70**	0.71
Temporal bias (veh h^{-1} $detector^{-1}$)	**18**	**5**	11
Temporal RMSE (veh h^{-1} $detector^{-1}$)	**25**	**30**	23
Temporal NRMSE (%)	**14.05**	**16.86**	12.50
Temporal Correlation	**0.99**	**0.97**	0.99
Spatial bias (veh h^{-1} $detector^{-1}$)	**20**	**5**	13
Spatial RMSE (veh h^{-1} $detector^{-1}$)	**116**	**105**	108
Spatial NRMSE (%)	**64.02**	**58.16**	59.89
Spatial Correlation	**0.58**	**0.59**	0.59

The deterministic simulation results show that the extended Meta-LCF model has similar performance as the one built in Chapter 4. The comparison of the spatially-averaged flows computed by these two metamodels is presented in Figure 7.1 and Figure 7.2. Results show that the metamodel built with larger input variation intervals is also able to predict the spatially-averaged flows on the network when compared with the loop detector measurements.

Figure 7.3 shows the comparison of temporally-averaged flows computed by the two metamodels. They both have limits to predict the spatial distribution of traffic on the network, when comparing the computed flows with loop detectors. Figure 7.2 and Figure 7.3 show that bigger differences are found between the extended metamodel and the original model when comparing the metamodel built in Chapter 4. However, the two

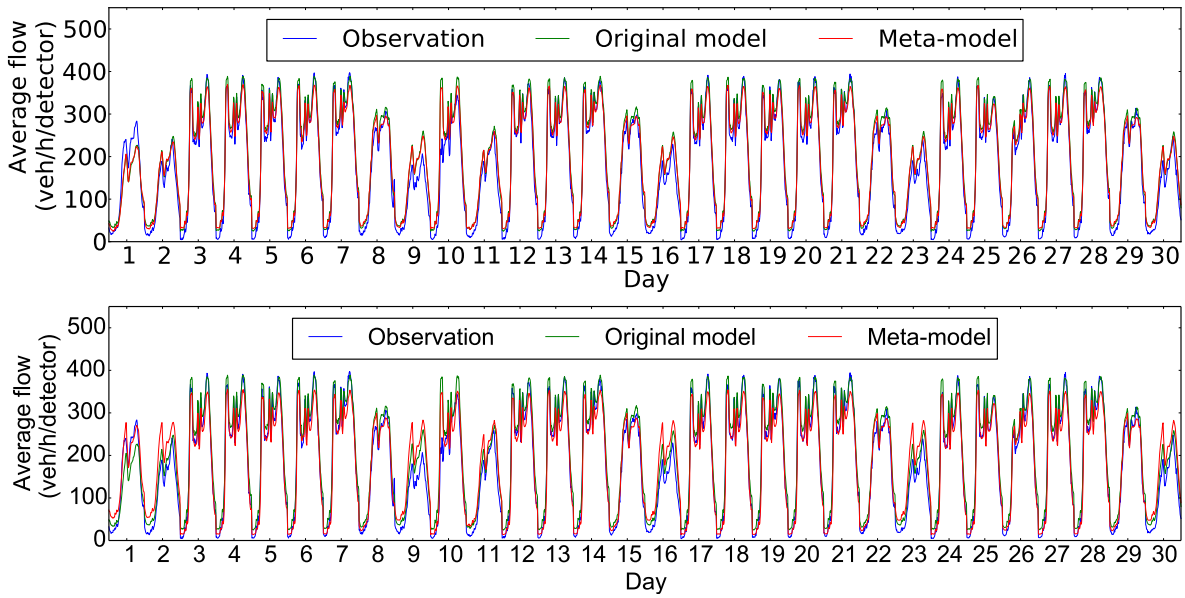


Figure 7.1: Temporal variation of the spatially-averaged flows (in veh h^{-1}) of observation (blue), flows computed by the original model (green) and flows computed by the metamodel (red), with a time step of 15 min during the November 2014 for the agglomeration of Clermont-Ferrand. The upper figure presents the results of the extended Meta-LCF model in this chapter, and the lower figure presents the results of the Meta-LCF built in Chapter 4.

metamodels have approximately similar performance when compared with loop detector measurements.

7.2.2 Ensemble simulations based on Monte Carlo approach

We considered that the inputs of the Meta-LCF model are random variables with known probability distributions. We then used the Monte Carlo approach with sample size $n = 100$ to generate an ensemble of n DTA simulations with Meta-LCF model. The uncertain input data (cf Table 7.2) are perturbed independently, n times, i.e., once for each member of the Monte Carlo simulations. Then each of the n simulations is a unique run of the Meta-LCF with one of the n perturbed inputs. From the n-member ensemble, we obtain a discrete probability density function of the traffic flows and speeds.

The ensemble outputs are traffic flows during the whole month of November 2014 at street resolution at all time intervals of 15 minutes, for the agglomeration of Clermont-Ferrand. The uncertainty added to the input $\boldsymbol{p} = (p_i)_{i=1,\dots,22}$ in Table 7.2 of Meta-LCF model are supposed to be uniformly distributed in order to explore the whole

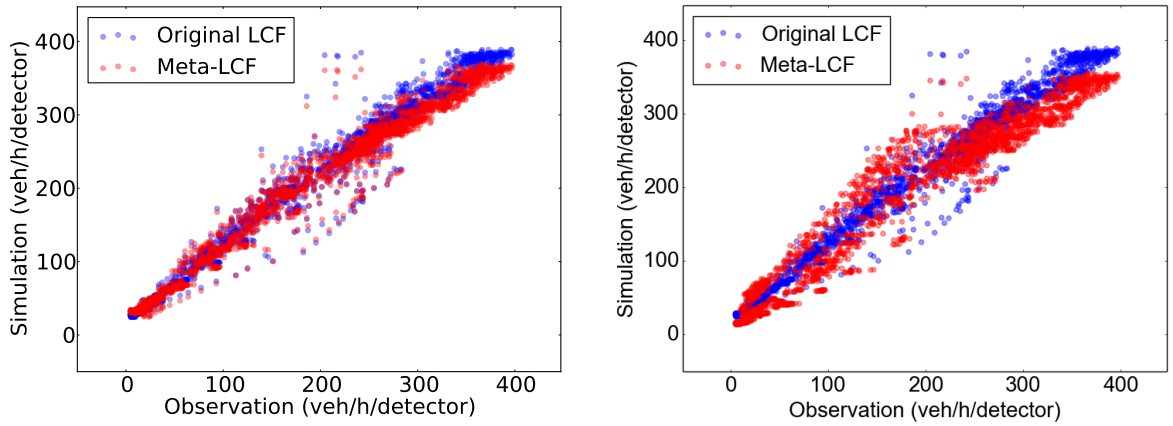


Figure 7.2: Scatter plots of spatially-averaged flows. Left: dispersion diagram of spatially-averaged flows computed by models (blue for the original LCF model, and red for the meta-model firstly built in Chapter 4), against measured flows at each time step. Right: dispersion diagram of spatially-averaged flows computed by models (blue for the original LCF model, red for the meta-model built in this chapter), against measured flows at each time step.

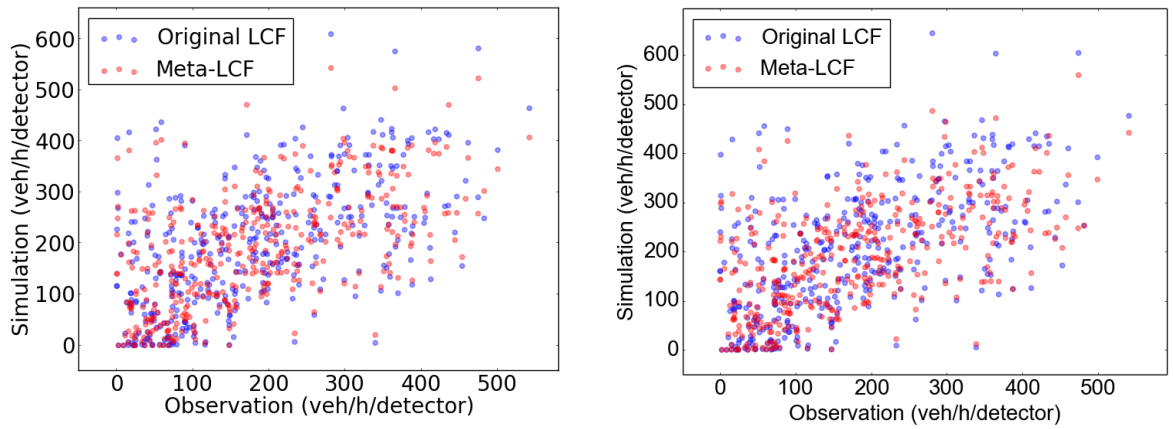


Figure 7.3: Scatter plots of temporally-averaged flows. Left: dispersion diagram of temporally-averaged flows computed by models (blue for the original model, and red for the meta-model built in Chapter 4), against measured flows at each detector. Right: dispersion diagram of temporally-averaged flows computed by models (blue for the original model, red for the meta-model built in this chapter), against measured flows at each detector.

input space. For instance, the uncertainty of p_i follows the uniform distribution on the interval $[a_i, b_i]$. Therefore, in the Monte Carlo simulations, p_i is randomly sampled from $[p_i^* \times a_i, p_i^* \times b_i]$, with $i = 1, 2 \ldots, 22$ and p_i^* is the value of p_i in the deterministic simulation.

The first ensemble is generated with the Meta-LCF model from Chapter 4. The variation intervals for the 22 inputs of Meta-LCF model are presented in Table 7.4. Note that the 23^{th} element in $\boldsymbol{p}$ of the Meta-LCF model is the O-D matrix transposition coefficient η. For generating the ensemble of DTA simulations during a given period, η is a known input, depending on the simulation time (before or after 12:00) of the day. It is not perturbed in our Monte Carlo simulations.

Table 7.4: The variation intervals of Meta-LCF inputs for generating ensemble by Monte Carlo simulations. The Meta-LCF used here is the one built in Chapter 4.

Input	Temporal profile		Link capacity				Speed limits				Demands in O-D matrix	
Symbol	$P(h_1), \ldots, P(h_{13})$		λ_{big}		λ_{small}		μ_{high}		μ_{low}		$\delta_0, \delta_{0-5}, \delta_{5-10}$ $\delta_{10-15}, \delta_{>15}$	
Interval boundaries	$a_i (i = 1, \ldots, 13$	$b_i (i = 1, \ldots, 13$	a_{14}	b_{14}	a_{15}	b_{15}	a_{16}	b_{16}	a_{17}	b_{17}	$a_i (i = 1, \ldots, 22$	$b_i (i = 1, \ldots, 22$
Variations	0.5	1.5	0.7	1.3	0.7	1.3	0.8	1.2	0.7	1.1	0.25	1.5

We will see in the Section 7.3.3 that the first ensemble does not prove to be a *reliable* ensemble, when compared with the loop detector measurements during the whole month of November 2014. Therefore, a second ensemble is generated with larger uniform distributions, using the Meta-LCF model presented in Section 7.2.1. The final boundaries of input variables are presented in Table 7.5.

Table 7.5: The variation intervals of Meta-LCF inputs for generating ensemble Monte Carlo simulations

Input	Temporal profile		Link capacity				Speed limits				O-D matrix	
Symbol	$P(h_1), \ldots, P(h_{13})$		λ_{big}		λ_{small}		μ_{high}		μ_{low}		$\delta_0, \delta_{0-5}, \delta_{5-10}$ $\delta_{10-15}, \delta_{>15}$	
Interval boundaries	$a_i (i = 1, \ldots, 13$	$b_i (i = 1, \ldots, 13$	a_{14}	b_{14}	a_{15}	b_{15}	a_{16}	b_{16}	a_{17}	b_{17}	$a_i (i = 1, \ldots, 22$	$b_i (i = 1, \ldots, 22$
Variations	0.0	2.0	0.5	1.5	0.5	1.5	0.5	1.5	0.5	1.5	0.0	1.8
Uncertainty	$\pm 100\,\%$		$\pm 50\,\%$		$\pm 50\,\%$		$\pm 50\,\%$		$\pm 50\,\%$		$[-100\,\%, +80\,\%]$	

For the whole month of November 2014, there are in total $T_{size} = 30 \times 96 = 2880$ time

steps. The values of the 22 inputs of Meta-LCF model for the deterministic simulation during the simulation period are denoted as $p^*_{i,t}$, with $i = 1, \ldots, 22$ and $t = 1, 2, \ldots, T_{size}$. With the given boundaries, the i^{th} input variables are randomly sampled from the interval $[p^*_{i,t} \times a_{j,i}, p^*_{i,t} \times b_{j,i}]$ with $j = 1, 2, \ldots, n$ for the Monte Carlo simulation of size n. The j^{th} sampled input for carrying out DTA simulation with Meta-LCF is $\boldsymbol{p}_j$. Therefore, there are in total n computed traffic flows $\boldsymbol{x}_{j,t} \in \mathbb{R}^{19628}$ for the agglomeration of Clermont-Ferrand. The output of the j^{th} member of the ensemble is computed as $\boldsymbol{x}_{j,t} = \widehat{\mathcal{M}}(\boldsymbol{p}_j)$. The computed traffic flow of the j^{th} member of the ensemble at link a and time t is $x_{j,a,t}$. Let $X_{a,t}$ denote random output of the Meta-LCF model at link a at time t. Let f and U denote the Meta-LCF model and its input variable. Then on the link a and time t, $X_{a,t} = f(U)$. The objective for uncertainty quantification is to determine the probability distribution of $X_{a,t}$ and calculate its standard deviation, interquartile range and $5^{th} - 95^{th}$ percentile range from the ensemble $(x_{j,a,t})_{j=1,\ldots,n}$.

7.3 Ensemble evaluation

Ensemble simulations are widely used in meteorology, air quality simulations, etc. In this work, an ensemble of DTA simulations with Meta-LCF model is built based on Monte Carlo approach. Before quantifying the output uncertainty, it is necessary to evaluate the performance of the ensemble, in order to verify whether it is able to well represent the uncertainty or not. To do so, traffic flow observations are used to evaluate the generated ensemble. This section firstly presents criteria for evaluating an ensemble with the help of observations. Then we applied these criteria to the ensemble of traffic flow simulations in the agglomeration of Clermont-Ferrand, during November 2014.

7.3.1 Scores for ensemble evaluation

There are three desirable properties for a good ensemble system: reliability, sharpness and resolution. The reliability indicates the accuracy of prediction of an event. An ensemble with high reliability is so that the probabilistic forecast by the ensemble is (approximately) the same as the observed frequency of this event. The sharpness indicates the variability of the ensemble forecasts. The resolution indicates the ability of the ensemble to make distinct predictions for different events. For uncertain quantification using ensemble simulations, the most important criteria for the generated ensemble is its reliability. A *reliable* ensemble can give us *reliable* uncertainty estimations of the outputs of the simulation system. The other two criteria are more frequently used to evaluate whether an ensemble is able to provide good probabilistic forecasts for the

occurrence of an event. Therefore, in the following part of this chapter, we focus only on the *reliability* criterion to evaluate our ensemble of traffic flow simulations.

To evaluate the *reliability* of the ensemble generated in Section 7.2, we use the rank histogram, the reliability diagram and statistical scores for the outputs of the ensemble. We only have traffic flow observations on the network of Clermont-Ferrand. Since the reliability of an ensemble relies on the comparison between the ensemble forecast outputs and observations, only the ensemble of DTA simulations is evaluated and improved in this section.

Rank histogram

A rank histogram measures the reliability of an ensemble. The rank histogram [Anderson, 1996; Hamill and Colucci, 1997; Talagrand, 1999] is used to compare two random variables: (i) the variable predicted by the ensemble of simulations, and (ii) the observations of the target variable. Let $\boldsymbol{X} = X_1, X_2, \ldots, X_{n-1}, X_n$ be a sequence of n independent and identically distributed random variables. X_j are sorted in increasing order so that $X_1 \leq X_2 \leq \ldots \leq X_{n-1} \leq X_n$. Let Y be the target random variable to be compared with. If all X_j and Y has the same probability distribution, it can be proved that

$$E_X[P_Y(Y \leq X_j)] = \frac{j}{n+1}, \text{ and} \tag{7.6a}$$

$$E_X[P_Y(X_{j-1} < Y \leq X_j)] = \frac{1}{n+1}, \tag{7.6b}$$

where $E_X[.]$ denotes the expectation related to X, P_Y denotes the probability associated with Y. For the ensemble of traffic flow simulations, $\{X_1, X_2, \ldots, X_{j-1}, X_j, \ldots, X_n\}$ are traffic flow outputs of the n-member ensemble at detector i and time t, sorted by increasing order. y is a realization of the random variable Y: traffic flow measured by the loop detector at the same i and t. If the generated ensemble is *reliable*, the predictions $X = (X_1, \ldots, X_j, \ldots, X_n)$ of the ensemble should has the same probability distribution as the observation variable Y. Therefore, y should satisfy the conditions where (i) the probability of $y \leq X_j$ equals to $\frac{j}{(n+1)}$ and (ii) the probability of $X_{j-1} < y \leq X_j$ equals to $\frac{1}{(n+1)}$. The rank histogram is computed by counting the ranks of all the observations at all detectors and times. It gives a quick examination of some qualities of the ensemble. A perfect rank histogram is flat, while a U-shape rank histogram shows a lack of variability in the ensemble. Consistent biases in the ensemble forecast will show up as a sloped rank histogram. If there are more observations in lower ranks ($= 0$), the ensemble

forecast presents an overestimation compared with the observations. If there are more observations in higher ranks ($\geq n$), it means that the ensemble forecast presents an underestimation when compared with the observations.

A quantitative indicator can be calculated in order to evaluate the flatness of a rank histogram. Let M_{ob} be the total number of observations. Let r_j be the number of observations of rank j in the rank histogram and $j \in [0, n]$. For a reliable system, the expectation of r_j should be $\gamma = M_{ob}/(n+1)$. The quantity

$$\Delta = \sum_{j=0}^{n} (r_j - \gamma)^2 \tag{7.7}$$

measures the deviation of the histogram from flatness. The ratio $\delta = \frac{\Delta}{n\gamma}$ is a measure of the reliability of an ensemble prediction system for a scalar variable [Candille and Talagrand, 2005]. A value of δ that is significantly larger than 1.0 indicates that the ensemble prediction is not reliable. A value of δ significantly less than 1.0 indicates that the realizations of the prediction process are not independent.

Reliability diagram

The reliability diagram is used to summarize and evaluate probabilistic forecasts. It plots the observed relative frequency of an event, against the predicted probability by the ensemble. Y is a variable that is equal to one when an event does happen, or equal to zero when the event does not happen. The variable Y is called the *verification*. Let $Y_i, i = 1, \ldots, M_{ob}$ be a dataset of verifications. For each i, the forecast value from the ensemble is X_i and it is between 0.0 and 1.0. It represents the forecast probability that the corresponding verification Y_i equals to one (i.e, the event does happen). The reliability diagram assesses whether X_i is reliable or not. A probabilistic forecast is considered to be *reliable* if the event actually happens with an observed relative frequency consistent with the forecast value. The forecast values of X_i are collected into a number of representative bins. The definition of *reliability* is that: the relative frequency of the event $Y_i = 1$, when computed over all i for which X_i falls into a small interval, must be equal to the mean of X_i over that interval [Bröcker and Smith, 2007].

Reliability diagram plots the observed relative frequencies against the forecast values. Now we present how the diagram is computed. Firstly, we partition $[0, 1]$ into B sub-intervals: $\mathcal{B}_1, \ldots, \mathcal{B}_b, \ldots, \mathcal{B}_B$, $b = 1, \ldots, B$. Next, for each i, we search in which interval $\mathcal{B}_b$ the forecast value X_i falls. For each sub-interval $\mathcal{B}_b$, let I_b be the set of all indices i for which X_i falls into $\mathcal{B}_b$: $I_b = \{i | X_i \in \mathcal{B}_b\}$. Then the corresponding *observed relative frequency* f_b is the number of times the event happened ($Y_i = 1$), given that

$X_i \in \mathcal{B}_b$, divided by the total number of forecast values $X_i \in \mathcal{B}_b$ (i.e. $card(I_b)$): $f_b = (\sum_{i \in I_b} Y_i)/card(I_b)$. Finally, we compute the average of the forecast values over the bin $\mathcal{B}_b$ and denote it as $r_b = (\sum_{i \in I_b} X_i)/card(I_b)$. The reliability diagram is then obtained by plotting f_b against r_b for all bins $\mathcal{B}_b$. For a *reliable* forecast, the reliability diagram should follow the diagonal (i.e., $f_b = r_b$) for any $b = 1, 2, \ldots, B$.

Statistical scores

The distributions of the model outputs are to be determined and this is the objective of uncertainty quantification. The discrete probability distributions of model outputs are predicted by the ensemble. Regardless of the underlying output distribution, statistical scores based on percentiles can be used to evaluate whether two random variables have approximately the same probability distribution. For example, about 50 % of the predicted values should be in the interquartile range (IQR), which equals to the difference between 75th and 25th percentiles of the variable. About 90 % of the predicted values should be in the 90 % confidence interval (denoted as $CI90$), which corresponds to interval between the 5th and the 95th percentiles of the variable. Therefore, if the variable X predicted by the ensemble has more or less the same distribution as the true variable Y, we will have: (i) $P(Y \in IQR_X) \simeq 0.5$, and (ii) $P(Y \in CI90_X) \simeq 0.90$. In our case study, the traffic flows are computed at link level and at time resolution of 15 minutes. If an ensemble can well predict the variability and uncertainty of a simulation system, the statistical scores of the traffic flow predicted by the ensemble should be consistent with the observed frequency.

In practice, assuming that there are M_{ob} observation points. The ensemble predicts n values at the observation point i, with $i = 1, 2, \ldots, M_{ob}$. The observed value at point i is denoted as y_i. The variable computed by the ensemble is denoted as $x_{i,j}$ with $i = 1, 2, \ldots, M_{ob}$ and $j = 1, 2, \ldots, n$. Let us denote I_{IQR} and I_{CI90} as the number of the observations where $y_i \in IQR_i$ and $y_i \in CI90_i$, respectively. We can say that the ensemble prediction is reliable if (i) $\frac{I_{IQR}}{M_{ob}} \simeq 0.50$ and (ii) $\frac{I_{CI90}}{M_{ob}} \simeq 0.90$.

7.3.2 Evaluation of the 100-member ensemble of DTA simulations for the agglomeration of Clermont-Ferrand with Meta-LCF model

In this section, we apply the evaluation criteria to the 100-member ensemble of DTA simulations built in Section 7.2.2. There are 469 detectors on the network in the city of Clermont-Ferrand (Figure 7.4). They give the number of vehicles passing a road with

time resolution down to 1 minute. In our study, the observation data are aggregated to time intervals of 15 minutes. The simulation period is November 2014. Since the traffic flows computed by the Meta-LCF model are at street resolution while the loop detector measurements are at lane resolution, we divide the simulated link-level traffic flow by the number of lanes, in order to convert link-level flow into lane resolution traffic flow. Therefore, the *total* scores compare the simulation and observations for all detectors and at all time intervals of 15 minutes during the whole month. The *temporal* scores are calculated based on spatially-averaged traffic flows: for each time interval $[h, h + \Delta h]$, spatially-averaged flows are obtained by taking the mean value of all the detectors during the same interval, and we then compute statistical scores. The *spatial* scores are calculated based on temporally-average traffic flows: for each detector, we calculated the average flow over the whole simulation period, and we then computed statistical scores.

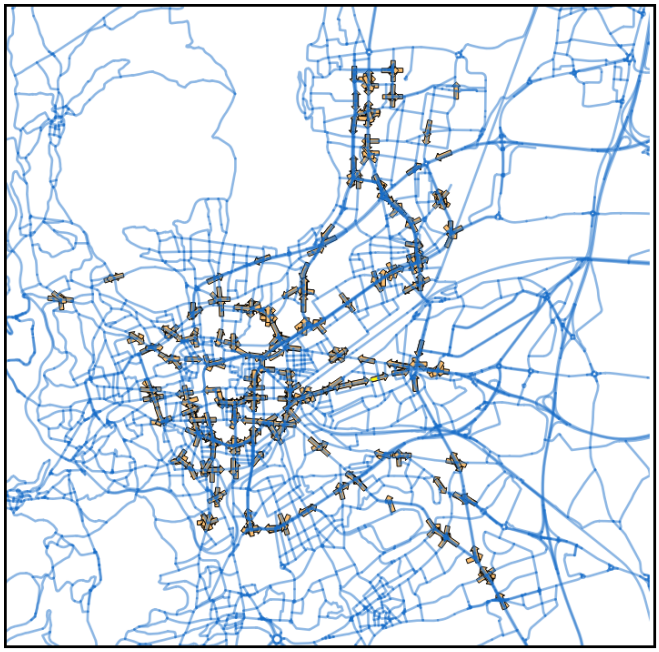


Figure 7.4: Locations and directions of loop detectors in the City of Clermont-Ferrand. The detectors are at lane resolution. The directions of the arrows in the figure indicate the directions of traffic flows that the detectors can measure.

Rank histogram

Figure 7.5 shows the rank histogram of the 100-member ensemble of DTA simulations with Meta-LCF model. The observations used for the evaluation are traffic flows at all detectors and at all time intervals during November 2014. There are in total $M_{ob} =$

$469 \times 30 \times 96$ observations. Results show that there are many observations in the last rank: the traffic flow predictions from the ensemble show an underestimation for some detectors and/or during some time intervals. 8.25% of the M_{ob} observations are higher than the upper envelope of the ensemble.

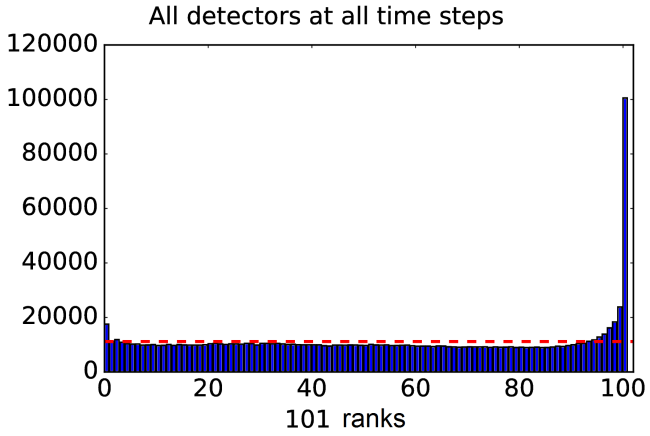


Figure 7.5: Rank histogram of the 100-member ensemble on the network of Clermont-Ferrand for traffic flows at all detectors and at all time intervals. The red horizontal dashed line corresponds to the ideal value for a flat rank histogram with respect to the number of members. The large number of observations on the right (in the last rank) means that many observations are higher than the upper envelope of the ensemble of DTA simulations with Meta-LCF model. 8.25% of the $469 \times 30 \times 96$ observations are higher than the upper envelope of the ensemble.

Now we want to analyze the locations (detectors) and time intervals where there is an underestimation by the ensemble of DTA simulations. Let $o_{k,t}$ be the traffic flow measured by the loop detector at detector k and time t. Let $s_{j,k,t}$ be the traffic flow of the j^{th} ensemble member, computed by the Meta-LCF model. In total, for the detector k, there are at most $T_{size} = 96 \times 30$ observations during the simulation period of November 2014. The mean of the predicted traffic flow at k and time t is obtained by: $\mu_{k,t} = \frac{1}{n} \sum_{j=1}^{n} s_{j,k,t}$. We denote $\mathcal{K}^{last}$ and $\mathcal{T}^{last}$ the set of positions and times, where the ensemble underestimates the traffic flows. At these times and detectors, the observations are in the n^{th} rank in the rank histogram. These observations are higher than the upper envelope of the ensemble of DTA simulations. At detector k and time t, we calculate the error between (i) the mean of the traffic flows simulation by the ensemble and (ii) the observations: $e_{k,t} = \mu_{k,t} - o_{k,t}$ with $k \in \mathcal{K}^{last}$ and $t \in \mathcal{T}^{last}$. Then we calculate the bias between the observation and the mean of the traffic flows simulated by the ensemble. The *temporal* bias is the temporally-averaged error, calculated at detector

k: $\beta_k^{temporal} = \frac{1}{card(\mathcal{T}^{last})} \sum_t e_{k,t}$ with $k \in \mathcal{K}^{last}$ and $t \in \mathcal{T}^{lase}$. The *spatial* bias is the spatially-averaged error, calculated at time t: $\beta_t^{spatial} = \frac{1}{card(\mathcal{K}^{last})} \sum_k e_{k,t}$. Figure 7.6 shows the distribution of spatial and temporal bias. It shows that we have bigger spatial bias than temporal bias. These results are compatible with the results of the deterministic simulation: we can better predict the spatially-average traffic flows on the network, but there are still errors in predicting route choices of network users.

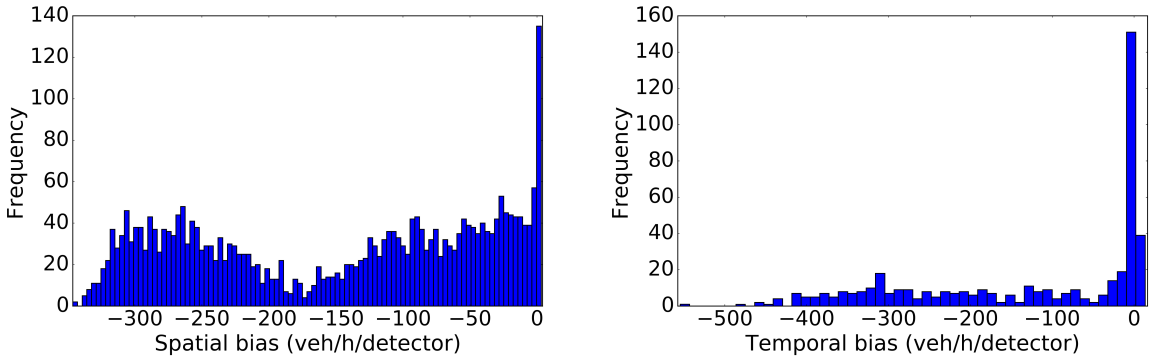


Figure 7.6: The distribution of the spatial bias at all times (left) and the temporal bias at all detectors (right) for the points where the ensemble underestimates the traffic flows when compared with loop detector measurements.

Figure 7.7 shows the locations and the associated errors of the last-rank observations in the road map of Clermont-Ferrand at six time intervals of 15 minutes during different time periods, on the 20^{th} November 2014. The figures show locations where the ensemble underestimates the traffic flow at lane resolution when compared with loop detector measurement.

Results in Figure 7.7 show that big errors are found during periods where the traffic on the network is heavy. For example, on the 20^{th} November 2014, we have the maximum number of observations above the upper envelope of the ensemble during 17:30 - 17:45, when there are 17.70 % of the 469 detectors at which the observed traffic flows are higher than the upper envelope of the ensemble. Results show that most of the underestimations of the ensemble prediction are located in the southern part of the ring road of Clermont-Ferrand city. Results also show that there is a systematic underestimation for the same detectors. The errors might due to the following reasons. (i) The representativeness errors are ignored. Traffic flow at lane resolution is obtained by dividing traffic flows at link resolution by the number of lanes. On some links with more than one lane, there may be an underestimation for the traffic flow predicted on the right-side detector, or an overestimation for the traffic forecast on the left-side detector. (ii) Uncertainty

lying in the O-D matrix. Only the travel purpose of *to/from work from/to home* is given in the current O-D matrix. Other categories of travel purpose (leisure, personal, social, etc.) are not taken into account. However, the traffic flow measured by the loop detectors take into account all kinds of traffic on the network. In addition, since the O-D matrix is obtained from household travel survey, there are errors in modeling O-D matrix from the traffic survey, too. (iii) Assumptions for route choices criteria. In LADTA model, we assume that users choose the least cost route from the origin to destination, with full information of the traffic situation of the network. This is certainly not be exactly the same case in reality. (iv) The uncertainty in estimating the link capacity. In fact, the traffic flow computed by LADTA model cannot exceed the link capacity. Therefore, if a link has three lanes and its capacity is 1200 veh h^{-1}, the capacity at lane resolution is 400 veh h^{-1}. The computed traffic flow cannot exceed 400 veh h^{-1} at lane resolution. However, some measured traffic flows are much higher than the capacities at lane resolution. On these detectors, big errors are found between (i) the traffic flow computed by the ensemble and (ii) the loop detector measurements, especially during peak periods when there are heavy traffic on the network. Figure 7.8 (left) shows the difference between the capacity and the loop detector measurements at lane resolution during 8:00 to 8:15 on the 20^{th} November 2017. Figure 7.8 (right) presents the errors between the computed flow and measurements on detectors where the observations are in the last rank of the rank histogram (see Figure 7.5). Figure 7.8 shows that about half of the observations in the last rank are higher than the capacity at lane resolution, during the corresponding period.

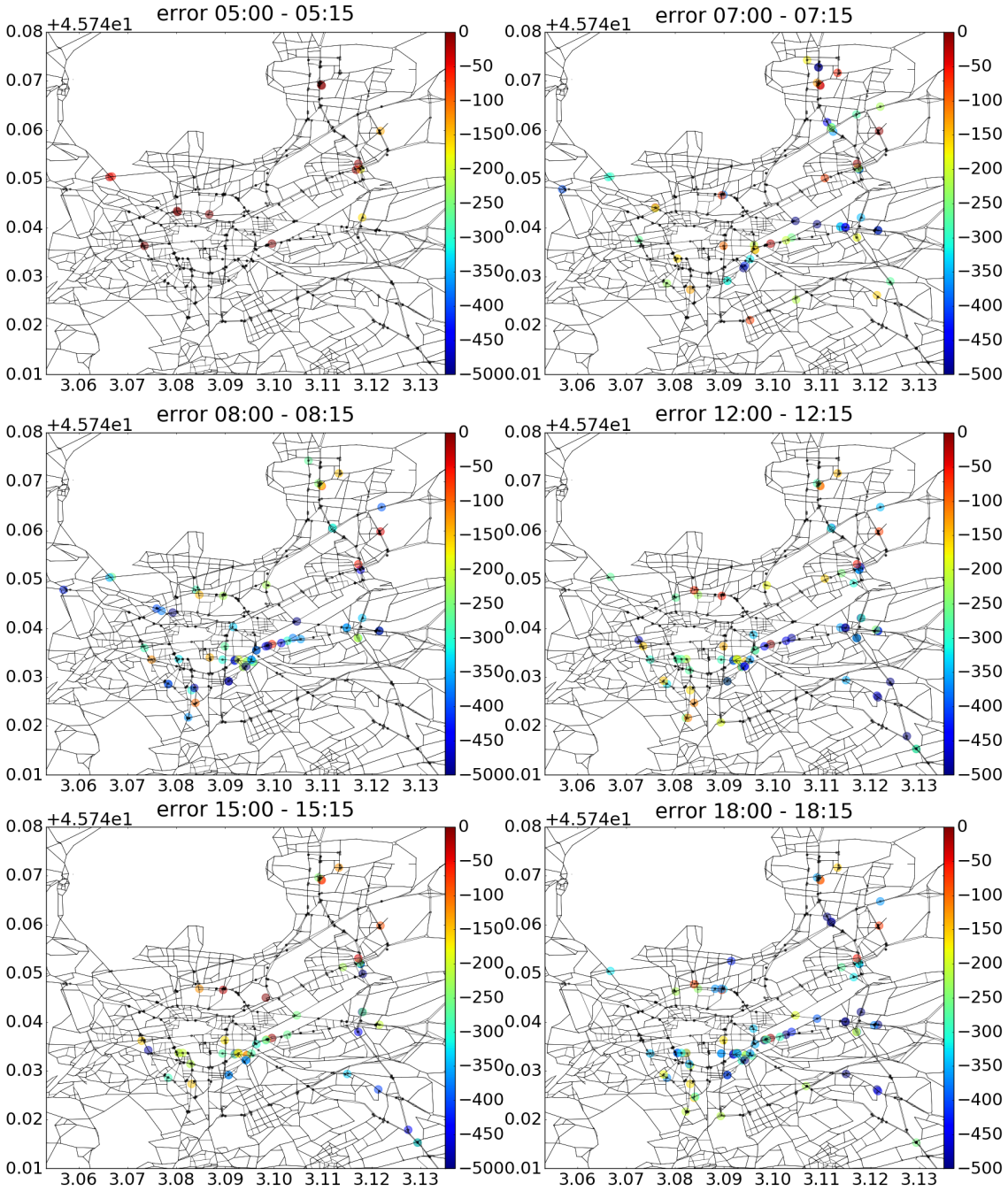


Figure 7.7: Errors of the traffic flow predicted by the ensemble, compared with the observations at which the measured traffic flows are higher than the upper envelope of the ensemble, on the 20^{th} November 2014 during different time periods. The points with colors represent detectors at which the measured flows are above the upper envelope of the ensemble in Figure 7.5. The different colors represent the difference between the mean flow predicted by the ensemble and the observed flow. The black dots represent detectors where the envelope of the ensemble prediction can cover the measured traffic flows.

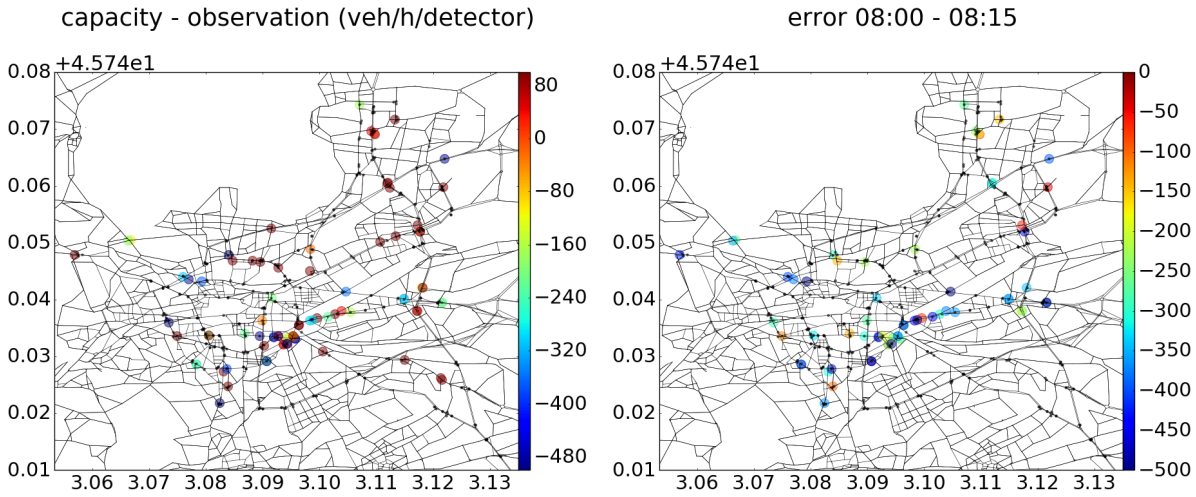


Figure 7.8: Difference between the capacity and loop detector measurements (left), and the difference between the computed traffic flows and corresponding loop detector measurements (right) at lane resolution (detector resolution) during 8:00 - 8:15 on the 20^{th} November 2014, for the locations where the measured traffic flows are in the last rank of the rank histogram in Figure 7.5. The black dots represent detectors where the envelop of the ensemble prediction can cover the measured traffic flows.

The rank histogram can be improved by better representing the input uncertainty, especially better determining the inputs associated to spatial uncertainties. For example, we could add more spatial uncertainty into the link capacities, speed limits, and O-D matrix in order to increase the spatial uncertainty of traffic flows computed by the ensemble.

Reliability diagram

The reliability diagram is used to evaluate whether an ensemble can well predict an event. For example, in our application of predicting traffic flows, an event can be whether traffic flow on a road exceeds a certain threshold. Transportation planners are interested in the service level of service (LOS) of a road network [TRB, 2000]. The density is used to define LOS, detailed in Table 7.6.

There are in total A to F levels of service for freeways. LOS-A describes completely clear traffic conditions and the average speed is close to the free-flow speed. LOS-B also indicates a nearly free flow condition. LOS-C describes a situation where vehicles interfere significantly with one another. LOS-D marks the border of unstable regimes. The average travel speed are greatly restricted by the congestion, and medium disturbances

Table 7.6: Traffic density criteria for basic freeway segments for LOS. The traffic density describes the number of vehicles passing through a segment of unit distance (km).

Level of Service (LOS)	Density range ($\mathrm{veh\,km^{-1}\,lane^{-1}}$)
A	$[0-7[$
B	$[7-11[$
C	$[11-16[$
D	$[16-22[$
E	$[22-28[$
F	>28

can cause a transition to LOS-F. LOS-E indicates that the capacity is reached and it marks unstable regimes. Since the capacity is reached, the vehicles follow each other very closely. Any perturbation can lead to LOS-F. At last, the LOS-F indicates that the road is in a constant traffic jam . For determining the LOS of freeway segment, criteria is given based on density of a freeway segment by the definition of LOS. With given free-flow speed and speed-volume curve of fundamental diagram, the density threshold can be converted to volume-to-capacity ratio (V/C). The V/C criterion is presented in Table 7.7 for freeways for which the free-flow speed is $90\,\mathrm{km\,h^{-1}}$.

Table 7.7: LOS volume-capacity ratio criteria for basic freeway segments for which the free-flow speed is $90\,\mathrm{km\,h^{-1}}$.

	Level of service				
	A	B	C	D	E
Volume-capacity-ratio (V/C)	0.28	0.44	0.64	0.87	1.00
Density threshold ($\mathrm{veh\,h^{-1}}$)	7	11	16	22	28

In our case study, we take the V/C criteria as an example to show whether the ensemble of DTA simulations can give reliable prediction of LOS, for all the roads on whole network of the agglomeration. Here we give a brief introduction of how to plot reliability diagrams in our case study using the LOS criteria V/C. There are M_{ob} observation points for measuring traffic flows (in $\mathrm{veh\,h^{-1}\,lane^{-1}}$) at all detectors and all times: $M_{ob} = N_{det} \times T_{size} = 469 \times (30 \times 96)$. The indices of computed variables and observations are denoted by $i = 1, 2, \ldots, M_{ob}$. The observed traffic flow at i is denoted as o_i and the traffic flows computed by the ensemble at i is denoted as $s_{j,i}$ with $j = 1, \ldots, n$, and n is the size of Monte Carlo simulation. We choose the $v/c = 0.44$ as an example to illustrate how to draw a reliability diagram. The binomial variable Y_i equals to 1 when

the traffic flows $y_i > 0.44C_i$, where C_i is the lane capacity at the observation point i. In our case study, the forecast probability at i is obtained as $X_i = \frac{card(\{j|s_{j,i}>0.44C_i\})}{n}$. The *verification* is given by loop detector measurements: $Y_i = 1$ if $o_i > 0.44C_i$ and $Y_i = 0$ otherwise. We then define the *bins* $\mathcal{B}_b \in \{]0.0, 0.1],]0.1, 0.2], \ldots,]0.9, 1.0]\}, b = 1, 2, \ldots, 10$. $I_b = \{i|X_i \in \mathcal{B}_b\}$. The *observed relative frequency* for the bin b is computed by

$$f_b = \frac{\sum_{i \in I_b} Y_i}{card(I_b)} = \frac{card(\{i \in I_b | o_i > 0.44C_i\})}{card(I_b)}. \tag{7.8}$$

The average of the forecast values over bin $\mathcal{B}_b$ is computed by

$$r_b = \frac{\sum_{i \in I_b} X_i}{card(I_b)}. \tag{7.9}$$

The reliability diagram then plots f_b against r_b with $b = 1, \ldots, 10$, for the given event *"traffic flows being higher than 0.44 of the capacity"*.

Similarly, we use the V/C criterion to draw reliability diagrams for each LOS criteria of Table 7.7. Four events are defined as $v/c > 0.28$, $v/c > 0.44$, $v/c > 0.64$, and $v/c > 0.87$. The Figure 7.9 shows the four resulting reliability diagrams.

Reliability diagrams in Figure 7.9 show that with the given input uncertainty presented in Table 7.5, the ensemble prediction performs better for predicting events with V/C threshold lower than 28 %. The ensemble underestimates the probabilities for lane-traffic flow exceeding 44 %. This result is consistent with the results for the rank histogram, showing that the ensemble prediction has a negative bias. The reason of this underestimation might mainly due to the reasons we proposed in Section 7.3.2. The reliability diagrams can be improved by better determining the boundaries a and b of the uniform distributions for input uncertainties. It can also be improved by introducing other uncertainty sources such as those in the observations, and by adding more uncertain inputs in the parameterization when building the metamodel of the LADTA.

Statistical scores

Now we want to see whether the traffic flows predicted by the ensemble show good statistical characteristics when compared with loop detector measurements. The ensemble size n is 100. The total number of observations is M_{ob}. At detector k and time t, $o_{k,t}$ is the loop detector measurement and the ensemble gives $n = 100$ predicted traffic flows $s_{j,k,t}$ with $j \in 1, 2, \ldots, n$. $\boldsymbol{s}_{k,t} = (s_{j,k,t})_j$ is a vector of dimension n. Let $IQR_{k,t}$ denote the interquartile interval of $\boldsymbol{s}_{k,t}$, where $Q1_{k,t}$ and $Q3_{k,t}$ are the traffic flow values of the first and third quartiles of $\boldsymbol{s}_{k,t}$ respectively. Let $CI90_{k,t}$ denote the inter-percentile between the 5^{th} and 95^{th} values of the ensemble forecasts at detector k and time t. In

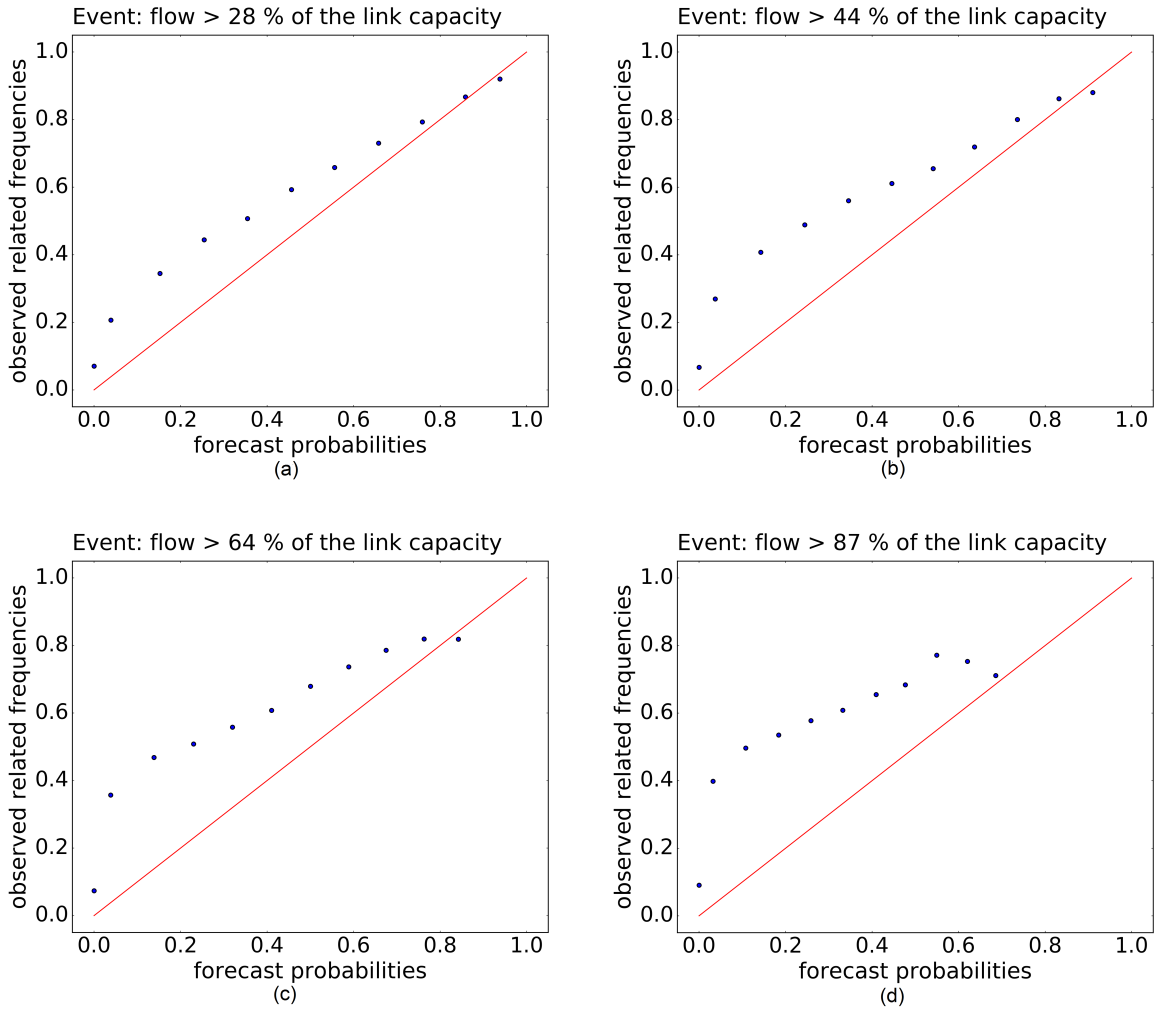


Figure 7.9: Reliability diagrams of the ensemble simulation for volume-capacity-ratio (V/C) at lane resolution. The V/C threshold is 0.28 (a), 0.44 (b), 0.64 (c) and 0.87 (d). The corresponding events are "the level of service being lower than LOS-A" (a), "the level of service being lower than LOS-B" (b), "the level of service being lower than LOS-C" (c), and "the level of service being lower than LOS-D" (d). The red line corresponds to a perfect reliability diagram.

order to evaluate the performance of the ensemble prediction of our case study, the two statistical scores presented in Section 7.3.1 can be defined as follows.

– (i) The probability of $o_{k,t} \in IQR_{k,t}$: $P(Q1_{k,t} < o_{k,t} \leq Q3_{k,t}) \simeq 0.5$. It can be approximated by calculating the occurrence frequency of observations that satisfy the condition among all the M_{ob} observations. In practice, it comes to count the number of observations points (k,t) where $o_{k,t} \in IQR_{k,t}$ and divide it by M_{ob}.

– (ii) The probability of $o_{k,t} \in CI90_{k,t}$: $P(o_{k,t} \in CI90_{k,t}) \simeq 0.9$. It can be approximated by calculating the occurrence frequency of observations that satisfy the condition among all the M_{ob} observations. In practice, it means counting the number of observations points (k,t) where $o_{k,t} \in CI90_{k,t}$ and dividing the total number by M_{ob}.

The statistical scores of the ensemble of traffic flow simulations are presented in the 3^{rd} column in Table 7.8. Results show that the statistical scores of the ensemble prediction do not perfectly meet the statistical requirements. The output uncertainty is not dispersed enough to statistically cover the loop detector measurement during the one-month simulation period of November 2014.

Table 7.8: Statistical scores of the ensemble simulations compared with loop detector measurements at all detectors and times

Criterion	Target	Ensemble score	Temporal score	Spatial score
P_{IQR}	50 %	44.50 %	96.98 %	50.12 %
P_{CI90}	90 %	77.72 %	100 %	73.89 %

Now we use the same statistical criteria to evaluate the temporal performance and the spatial performance of the ensemble. The temporal performance is evaluated by computing statistical scores for spatially-averaged traffic flows during each of the 15-minute interval. The spatial performance is evaluated by computing statistical scores for temporally-averaged flows for each detector. There are $N_{det} = 469$ detectors and in total $T_{size} = 96 \times 30$ time intervals in our case study. The two kinds of averaged flows and the corresponding mean (μ) are presented in Table 7.9. The temporal and spatial scores of the ensemble in our case study are presented the 4^{th} and 5^{th} column in Table 7.8.

Table 7.9: Spatially-averaged and temporally-averaged flows and the formula for the statistical scores

-	Spatially-averaged flow		Temporally-averaged flow	
-	Measurement	Simulation	Measurement	Simulation
Variable	y_t $= \frac{1}{N_{det}} \sum_{k=1}^{N_{det}} o_{k,t}$	$\boldsymbol{x}_t = (x_{j,t})_{j=1,\dots,n}$ $= (\frac{1}{N_{det}} \sum_{k=1}^{N_{det}} s_{j,k,t})_j$	y_k $= \frac{1}{T_{size}} \sum_{t=1}^{T_{size}} o_{k,t}$	$\boldsymbol{x}_k = (x_{j,k})_{j=1,\dots,n}$ $= (\frac{1}{T_{size}} \sum_{t=1}^{T_{size}} s_{j,k,t})_j$
μ	-	$\mu_t = \frac{1}{n} \sum_{j=1}^{n} x_{j,t}$	-	$\mu_k = \frac{1}{n} \sum_{j=1}^{n-1} x_{j,k}$
Criteria	Occurence frequency for $y_t \in IQR_{\boldsymbol{x}_t}$ Occurence frequency for $y_t \in CI90_{\boldsymbol{x}_t}$		Occurence frequency for $y_k \in IQR_{\boldsymbol{x}_k}$ Occurence frequency for $y_k \in CI90_{\boldsymbol{x}_k}$	

The temporal scores (in the 4^{th} column in Table 7.8) show that the ensemble overestimates the uncertainty in spatially-averaged traffic flows. The interquartile envelope computed by the ensemble can already cover almost 100 % of the spatially-averaged observations. Figure 7.10 illustrates the spatially-averaged flows during all time intervals of one month. We can see that the observations almost all fall into the interquartile range produced by the ensemble.

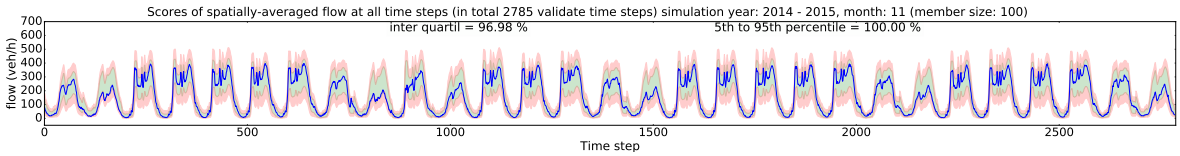


Figure 7.10: Interquartile score for the ensemble prediction of spatially-averaged traffic flows for all time intervals. The loop detector observations are represented by blue line, averaged from observations of all detectors for each 15 minutes time interval during November 2014. The green color band delimits the values between the 25^{th} and 75^{th} percentiles of the spatially-averaged flows computed by the ensemble. The red color band delimits the values between the 5^{th} and 95^{th} percentiles of the spatially-averaged flows computed by the ensemble.

However, the spatial scores (in the 5^{th} column in Table 7.8) show that we cannot perfectly predict the spatial uncertainty of the traffic flows. Even though the criteria of interquartile occurrence frequency is approximately equal to the target frequency 50 %, there are still some locations (detectors) where the temporally-averaged flows cannot be well predicted by the ensemble. There are less than 90 % observations that fall in the 90 %-confidence interval. The statistical scores of temporally-average flows are presented in Figure 7.11. We can observe that more than 10 % of observations (blue points) are not included in the 90 %-confidence range estimated by the ensemble. In particular, results in Figure 7.11 show that the lack of uncertainty spread is mostly found for detectors where the mean value of the flow predicted by the ensemble is low. Most of the detectors whose measured temporally-averaged flows are not covered in the envelope of 90 %-confidence interval are those with low simulated temporally-average flows ($< 100\,\mathrm{veh\,h^{-1}\,lane^{-1}}$). In other words, the lack of spread appears mostly on detectors where there are underestimations of the mean traffic flow predicted by the ensemble, when compared with measurements. In the previous subsection 7.3.2, we have already analyzed some observations points where underestimation might happen, as illustrated by the rank histogram (see Figure 7.5) and error maps (see Figure 7.7).

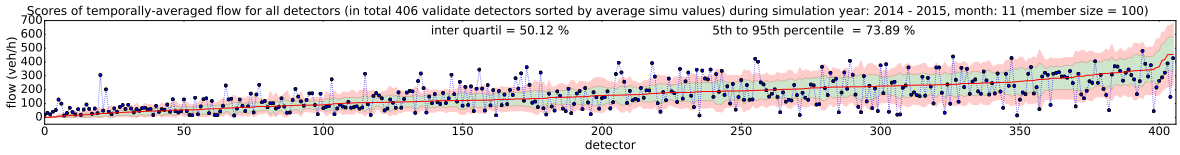


Figure 7.11: Statistical scores for the ensemble prediction of temporally-averaged traffic flows for all detectors. The loop detector observations are represented by blue points, averaged from observations of all 15 minute intervals for each detector during November 2014. The green color band delimits the values between the 25^{th} and 75^{th} percentiles of the temporally-averaged flows computed by the ensemble. The red color band delimits the values between the 5^{th} and 95^{th} percentiles of the temporally-averaged flows computed by the ensemble. The detectors are sorted from left to the right with respect to the mean of temporally-averaged flows estimated by the ensemble (plotted by the red line).

We have already used large ranges of traffic demand uncertainty. On the one hand, the total traffic demand for computing traffic flow during an interval of 15 minutes can be scaled from 0 % to 200 %, compared with the demand inputs in the deterministic simulation. On the other hand, the demands in static O-D matrix are categorized by 5 groups according to the distance between O-D pairs. Traffic demand in each of the five categories can vary from 0 % to 180 % (see Table 7.5). This the reason why we overestimate the uncertainty of the spatially-averaged traffic flows, since the latter reflect the total traffic volumes on the network. However, there are still slight underestimations of the spatial uncertainty. This might due to the lack of spatial uncertainties given in inputs, for link capacities, speed limits and for O-D matrix. This lack of spatial input uncertainty might result in poor variations in route choice strategies in DTA simulations. It might also be due to the limits of principal component analysis and interpolation when building the metamodel of LADTA applied to Clermont-Ferrand.

Influence of the input variation intervals of the Meta-LCF used for generating ensemble

In fact, since the Monte Carlo simulation for uncertainty quantification is based on the metamodel Meta-LCF, the maximum and minimum boundaries of the uniform distribution given to inputs are limited by the boundaries of input variations previously given for building the metamodel. The Meta-LCF firstly built in Chapter 4 has smaller boundaries (Table 7.4). Before building the new Meta-LCF model with larger input variation intervals, we have also generated an ensemble with the Meta-LCF model built

in Chapter 4, and we have analyzed the scores of this ensemble as well. Here we present some scores of the ensemble generated by the first Meta-LCF model. The statistical scores for this first ensemble are presented in Table 7.10.

Table 7.10: Statistical scores of the ensemble simulations compared with loop detector measurements at all detectors and times. The ensemble is generated with smaller uncertainties in inputs (see Table 7.4).

Criterion	Target	Ensemble score	Temporal score	Spatial score
P_{IQR}	50 %	34.30 %	84.17 %	34.05 %
P_{CI90}	90 %	63.10 %	96.59 %	56.16 %

The rank histogram, statistical scores for spatially-averaged flows and temporally-averaged flows are presented in Figure 7.12, Figure 7.13 and Figure 7.14.

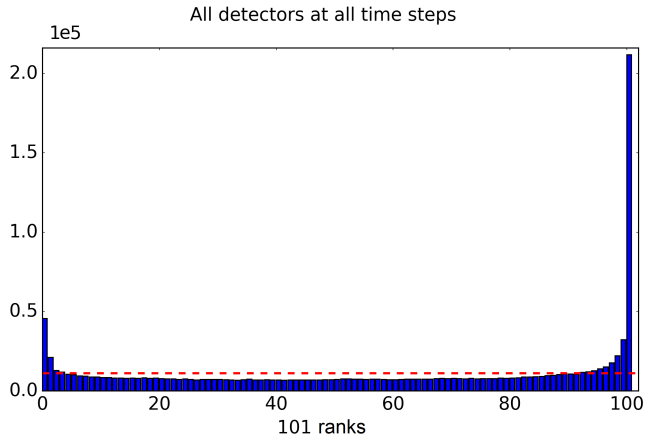


Figure 7.12: Rank histogram of the 100-member ensemble on the network of Clermont-Ferrand for traffic flows at all detectors at all time intervals of 15 minutes. The ensemble is built with the first Meta-LCF model built in Chapter 4. The higher bar on the left (first rank) means that there are observation values lower than the minimum values predicted by the ensemble. The higher bar on the right (the last rank) means that the observations higher than the envelope of the ensemble of DTA simulations with Meta-LCF model.

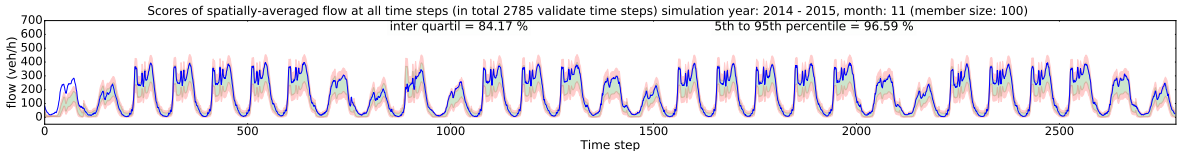


Figure 7.13: Statistical scores for the ensemble prediction of spatially-averaged traffic flows for time intervals. The ensemble is built with the first Meta-LCF model built in Chapter 4. The loop detector observations are represented by blue line, averaged from observations of all detectors for each 15 minutes time interval during November 2014. The green color band delimits the values between the 25^{th} and 75^{th} percentiles of the spatially-averaged flows computed by the ensemble. The red color band delimits the values between the 5^{th} and 95^{th} percentiles of the spatially-averaged flows computed by the ensemble.

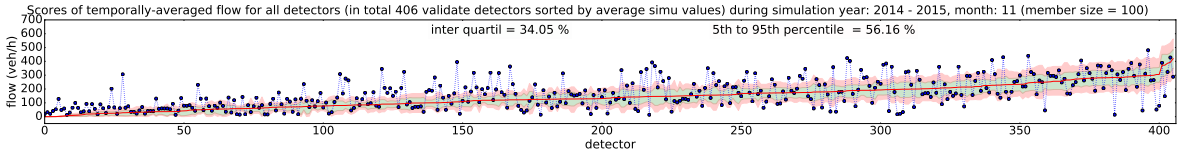


Figure 7.14: Statistical scores for the ensemble prediction of temporally-averaged traffic flows for all detectors. The ensemble is built with the first Meta-LCF model built in Chapter 4. The loop detector observations are represented by blue points, averaged from observations of all 15-minute intervals for each detector during November 2014. The green color band delimits the values between the 25^{th} and 75^{th} percentiles of the temporally-averaged flows computed by the ensemble. The red color band delimits the values between the 5^{th} and 95^{th} percentiles of the temporally-averaged flows computed by the ensemble. The detectors are sorted from left to the right with respect to the expectation of temporally-averaged flows estimated by the ensemble.

The obtained statistical scores from the ensemble generated by the Meta-LCF model built in Chapter 4 have not satisfied the statistical criteria. These results illustrate that the performance of the ensemble prediction is influenced of the inputs boundaries when building the Meta-LCF model. With the limits of the metamodel built in Chapter 4, we cannot increase the boundaries of inputs when generating the simulation ensemble. This is the reason why we rebuilt a second Meta-LCF model with larger input variation intervals, and use it to generate our prediction ensemble presented in this chapter.

Conclusions on the ensemble evaluation

The *total* statistical scores of the ensemble almost satisfy the statistical criteria. If we focus on the *temporal* scores and *spatial* scores, ensemble overestimates the temporal uncertainty of the spatially-average traffic flows. However, this overestimation of the temporal uncertainty still cannot offset the underestimation of spatial uncertainty, estimated by the ensemble of DTA simulations. There are still some negative bias when comparing the simulated traffic flows with loop detector measurements. The results in this subsection give information where the ensemble underestimates the uncertainty of computed traffic flows. The underestimation might be due to (i) the model assumptions; (ii) the lack of spatial uncertainty in traffic demands and in link parameters and (iii) potentially substantial measurement errors and representativeness errors when using the observations to evaluate the ensemble performance. In fact, in the evaluation, the traffic flows measured by inductive-loop detectors are considered to be *perfect*. The measurement error has not been taken into account when we compared the simulated traffic flows with measurements. In addition, since the traffic flows computed by the DTA model are at link-level while the observations are at lane resolution, representativeness errors are introduced and they are not taken into consideration when evaluating the ensembles. The statistical scores show that the underestimation of uncertainty mostly occurs at detectors where the mean value predicted by the ensemble underestimates the traffic flows, when compared with measured traffic flows at the same observation points. The presented approach gives criteria and a tractable framework to evaluate the generated ensemble, as well as the performance of the uncertainty quantification. Even though the whole uncertainty quantification framework is presented with our case study of DTA simulations in the agglomeration of Clermont-Ferrand, it can also be applied to other traffic assignment models for other cities by using the same approaches.

7.3.3 Improvement of the 100-member ensemble of DTA simulations for the agglomeration of Clermont-Ferrand with Meta-LCF model

It is shown that the performance of the ensemble can be improved by widening the boundaries of the uniform distribution given to Meta-LCF inputs. In this section, we use the same Meta-LCF model as in Section 7.2.2 with the largest input variations for generating the ensemble. However, instead of using the uniform distribution determined in Table 7.5, we define an objective function to search the *optimal* boundaries of a_i and b_i for the uniform distributions, in order to get the improved evaluation criteria presented

in Section 7.3.1. In other words, we focus on searching the distribution in inputs, in order to improve the reliability of the generated ensemble. Since it is not possible to find the analytical relation between the objective function and the input uncertainty distribution, a numerical optimization treats the relation as a black box and we use large sample of simulations to search for the improved uncertainty on the inputs.

The procedure for finding the *optimal* input uncertainty distribution is presented as follows.

Step 1. The uncertainty of an input p_i is modeled by the uniform distribution on $[a_i, b_i]$. We sample a_i and b_i independently from two non-overlapping intervals. We use Latin Hypercube Sampling (LHS) and the sample size for sampling a_i and b_i is $N_{opti} = 1000$.

Step 2. With the boundaries $[a_i^{(N)}, b_i^{(N)}], N = 1, 2, \ldots, N_{opti}$, N_{opti} ensembles are generated and the size of each ensemble is $n = 100$, the same as in Section 7.3.2.

Step 3. The target ensemble evaluation criteria is computed for each of the N_{opti} ensembles, and the value of some evaluation function is computed.

Step 4. We select the boundaries $([a_i^{N*}, b_i^{N*}])_{i=1,\ldots,K=23}$ that give the best value of the evaluation function.

To represent the temporal variations of traffic demand, the deterministic simulation makes use of $(P^*(h_i))_{i=1,\ldots,13}$ and for the ensemble simulation, each of $(P(h_i))_{i=1,\ldots,13}$ is sampled from the interval $[P^*(h_i) \times a_i, P^*(h_i) \times b_i]$. According to the global sensitivity analysis in Chapter 5, the traffic flows computed by Meta-LCF are only sensitive to the traffic demand during the previous 30 minutes before the simulation time. Therefore, instead of defining 13 pairs of boundaries a and b for the uncertainty of the 13 *temporal variation coefficient* $(P(h_i))_{i=1,\ldots,13}$, only one parameter α_1 is defined to represent the boundaries of $(P(h_i))_{i=1,\ldots,13}$ uncertainty. The minimum and maximum boundaries of the uncertainty of $(P(h_i))_{i=1,\ldots,13}$ are then $(a_i)_{i=1,\ldots,13} = 1.0 - \alpha_1$ and $(b_i)_{i=1,\ldots,13} = 1.0 + \alpha_1$. In the step 1 for sampling a_i and b_i, we need to make sure that the LHS will not fall outside the input intervals with which we built the metamodel. Therefore, α_1 varies in $[0.0, 1.0]$, so that $[P^*(h_i) \times a_i, P^*(h_i) \times b_i]$ do not exceed the intervals of $(P(h_i))_{i=1,\ldots,13}$ in Table 7.5. For the LHS, the variation ranges for minimum and maximum boundaries a_i and b_i in Table 7.5 are presented in Table 7.11.

Table 7.11: Variation ranges for the boundaries a_i and b_i.

Meta-LCF input	Boundaries for uniform distributions	Range for LHS of boundaries	Notes
$P(h)$	$a_i, i = 1, \ldots, 13$	$[1.0 - \alpha_1, 1.0]$	Sample α_1 from $[0.0, 1.0]$
	$b_i, i = 1, \ldots, 13$	$[1.0, 1.0 + \alpha_1]$	
λ_{big} λ_{small} μ_{high} μ_{low}	$a_i, i = 14, \ldots, 17$	$[0.5, 1.0]$	Sample all a_i and b_i randomly and independently
	$b_i, i = 14, \ldots, 17$	$[1.0, , 1.5]$	
δ_0 δ_{0-5} δ_{5-10} δ_{10-15} $\delta_{>15}$	$a_i, i = 18, \ldots, 22$	$[0.0, 1.0]$	
	$b_i, i = 18, \ldots, 22$	$[1.0, , 1.8]$	

The step 2 and step 3 are carried out in parallel. The process is carried out independently for different ensemble evaluation criteria to be optimized. Here we present the optimization process for reliability diagram as the criterion to be optimized.

For improving the reliability diagram, the objective is to minimize the difference between f_b and r_b calculated by Equation 7.8 and Equation 7.9. Our objective function is set as $\mathcal{C} = \frac{1}{B}\sum_{b=1}^{B}(f_b - r_b)^2$ with $b = 1, \ldots, B = 10$. Here we define the event of interest with $v/c = 0.64$ as the threshold of volume/capacity ratio. The optimization objective is then to minimize $\mathcal{C}$ for the event that the traffic flow is higher than 0.64 % of the link capacity. Table 7.12 shows (i) the uncertainty given to inputs with maximum boundaries defined in Table 7.5 for the original ensemble of DTA simulations in Clermont-Ferrand, and (ii) the uncertainty of inputs that results in *optimal* reliability diagram for the given event.

The obtained *improved* reliability diagram is presented in Figure 7.15 (left) and the corresponding rank histogram is presented in Figure 7.15 (right).

Compared with the reliability diagram in Figure 7.9 (c), the diagram in Figure 7.15 (left) is improved and closer to the diagonal line. In fact, with maximum possible uncertainty in inputs, the original ensemble of DTA simulations shows an underestimation for predicting the event that the traffic flows are bigger than 64 % of the link capacity. After optimization, the uniform distributions given to the *temporal coefficient* $(P(h_i))_{i=1,\ldots,13}$ and the link capacity parameter (λ_{big}) for links with big capacities $(\geq 900\,\mathrm{veh\,h^{-1}})$ remain the same. However, for other inputs, the uncertainty is described with an in-

Table 7.12: Uncertainty given to inputs of Meta-LCF model for the original ensemble simulation with maximum possible uncertainty ranges (2^{nd} column) and for the ensemble simulation giving *optimal* reliability diagram (3^{rd} column).

Meta-LCF input	Maximum uncertainty	Uncertainty corresponding Optimal to *optimal* reliability diagram
$P(h)$	$[-100\,\%, +100\,\%]$	$[-100\,\%, +100\,\%]$
λ_{big}	$[-50\,\%, +50\,\%]$	$[-50\,\%, +50\,\%]$
λ_{small}	$[-50\,\%, +50\,\%]$	$[0.0\,\%, +50\,\%]$
μ_{high}	$[-50\,\%, +50\,\%]$	$[0.0\,\%, +50\,\%]$
μ_{low}	$[-50\,\%, +50\,\%]$	$[-1\,\%, +18\,\%]$
δ_0	$[-100\,\%, +80\,\%]$	$[0.0\,\%, +48.55\,\%]$
δ_{0-5}	$[-100\,\%, +80\,\%]$	$[0.0\,\%, +80\,\%]$
δ_{5-10}	$[-100\,\%, +80\,\%]$	$[0.0\,\%, +80\,\%]$
δ_{10-15}	$[-100\,\%, +80\,\%]$	$[0.0\,\%, +80\,\%]$
$\delta_{>15}$	$[-100\,\%, +80\,\%]$	$[0.0\,\%, +35\,\%]$

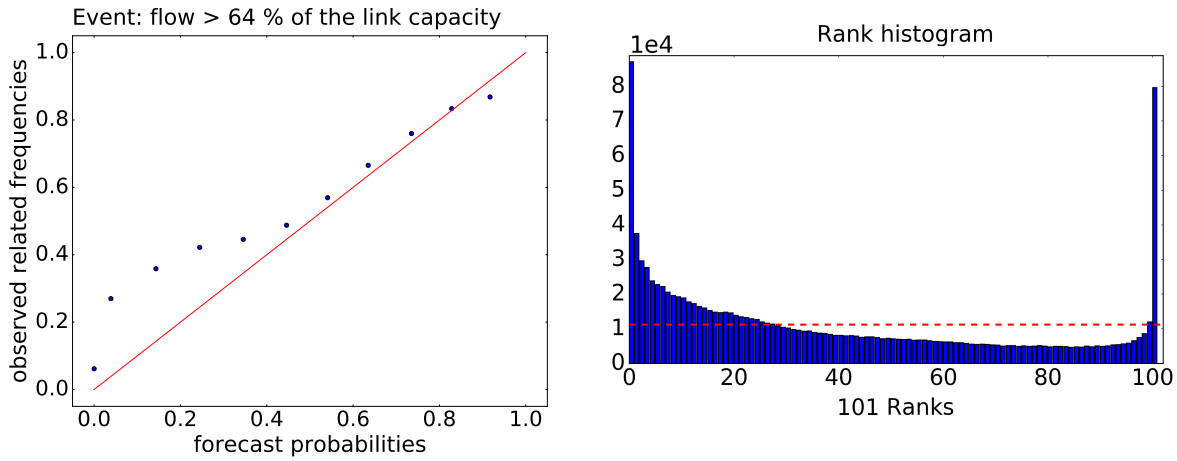


Figure 7.15: The improved reliability diagram (left) of the ensemble simulation generated by using the corresponding inputs uncertainty (in the 3^{rd} column of Table 7.12), and the corresponding rank histogram (right) of the same ensemble. The V/C threshold is 0.64.

creased left boundary of the uniform distribution. Most of the right-side boundaries remain the same, except for μ_{low} (uncertainty corresponding to low speed limits), δ_0 (uncertainty corresponding to traffic demands in the same zone) and $\delta_{>15}$ (uncertainty

corresponding to traffic demand between O-D pairs between which the distance is larger than 15 km). This adjustment of input uncertainty distribution decreases the number of simulations resulting in low traffic flows in the ensemble after optimization. Thus it increases the probability of high computed traffic flows predicted by the ensemble and leads to an improved reliability diagram. However, improving one of the ensemble evaluation criteria cannot make sure that other criteria can be improved at the same time. Figure 7.15 (right) presents the rank histogram of the ensemble after optimization. There is still a right side bar. In other words, there are still too many observations that are higher than the upper envelope of the ensemble. The number of observations in the last rank has not decreased when compared with the original rank histogram in Figure 7.5. Instead, since the some of left boundaries a_i are higher after optimization, the ensemble cannot predict low traffic flows as well as the ensemble in Section 7.3.2. As consequence, more observations are in the lower ranks. Moreover, if we compare the spatially-averaged traffic flows predicted by the two ensembles with loop detector measurements, the mean of the ensemble resulting from the optimization presents overestimation of the spatially-average flows during the one-month period, as shown in Figure 7.16. Figure 7.15 (right) and Figure 7.16 show that the optimization of reliability diagram for a given event may worsen the rank histogram and the performance of the ensemble mean.

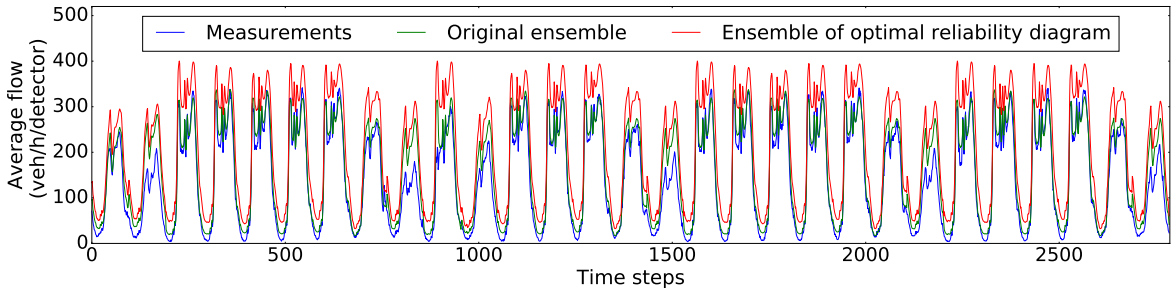


Figure 7.16: The spatially-averaged traffic flows computed with the original ensemble (green line), by the ensemble resulting from the optimization of reliability diagram (red line) and the spatially-averaged traffic flows measured by loop detectors. At each time step, the spatially-averaged flow is calculated by averaging the traffic flow over all 469 detectors on the network of Clermont-Ferrand.

We also carried out the optimization of the rank histogram. The results show that for optimizing the rank histogram, we should take the maximum intervals given in Table 7.5. Note that the optimization approach is computationally costly. For example, it took about two weeks for optimizing the reliability diagram in our case study for predicting

street-resolution traffic flows at metropolitan scale during one month. In addition, as shown in Figure 7.15 (right) and Figure 7.16, improving the ensemble according to one given criterion might not bring improvement on uncertainty quantification, and might even worsen other performance evaluation criteria of the ensemble. Other approaches could be investigated. For example, we can choose a sub-ensemble of the ensemble so that the criteria of the sub-ensemble are optimized. Then the selected sub-ensemble can be used to carry out uncertainty estimation. This approach has proved useful in the field of air quality forecasting [Garaud and Mallet, 2011].

7.4 Uncertainty quantification with the ensemble simulations

We use the ensemble with the maximum uncertainty as defined by Table 7.5 in Section 7.2.2, in order to quantify the uncertainty of output traffic flows and on-road traffic emissions at link resolution of the whole agglomeration of Clermont-Ferrand. Each member of the n-member ensemble computes traffic flow and speed at street resolution and for all intervals of 15 minutes. The traffic flows and speeds computed by each member are then coupled with the emission model built in Chapter 6 in order to compute the corresponding emissions. Then an n-member ensemble of emissions is built. We can measure the uncertainty of the emissions, resulting from the uncertainty in traffic flows.

7.4.1 Uncertainty in dynamic traffic assignment simulations

Uncertainty of computed traffic flows

At time t and link a in the agglomeration of Clermont-Ferrand, the uncertainty of the traffic flow computed by Meta-LCF model can be represented by the standard deviation $\sigma_{a,t}$. It is obtained by the ensemble of DTA simulations. σ is time-dependent because the traffic flow computed by the LADTA model is time-dependent. The mean of the traffic flows computed by the ensemble is denoted as $\mu_{a,t}$. The size of the ensemble is n. In our case study, $n = 100$ and total number of time steps during November 2014 is $T_{size} = 30 \times 96 = 2880$. Since the value of $\sigma_{a,t}$ depends on the value of $\mu_{a,t}$, another criterion is also used to describe the spread of traffic flows: the relative uncertainty, measured by the *relative standard deviation (RSD)*. At link a and time t, $\mu_{a,t}$, $\sigma_{a,t}$ and $RSD_{a,t}$ are computed as follows:

$$\mu_{a,t} = \frac{1}{n}\sum_{j=1}^{n} x_{j,a,t}, \tag{7.10a}$$

$$\sigma_{a,t} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_{j,a,t} - \mu_{a,t})^2}, \tag{7.10b}$$

$$RSD_{a,t} = \frac{\sigma_{a,t}}{\mu_{a,t}}, \tag{7.10c}$$

where $x_{j,a,t}$ is the computed traffic flow of the j^{th} member of the ensemble at time t and link a. The monthly averages μ_a, σ_a and RSD_a for road a during one month are:

$$\mu_a = \frac{1}{T_{size}}\sum_{j=1}^{n} \mu_{a,t}, \tag{7.11a}$$

$$\sigma_a = \frac{1}{T_{size}}\sum_{t=1}^{T_{size}} \sigma_{a,t}, \tag{7.11b}$$

$$RSD_a = \frac{1}{T_{size}}\sum_{t=1}^{T_{size}} RSD_{a,t}. \tag{7.11c}$$

A map of the estimated uncertainty is then displayed at street resolution. With Monte Carlo simulations, the uncertainties given to the inputs propagate to the computed traffic flows throughout the whole network of the agglomeration. Figure 7.17 shows the monthly-averaged uncertainty map over the Clermont-Ferrand city (left) and the whole agglomeration (right).

Uncertainty maps in Figure 7.17 show that high traffic flow uncertainty appears on highways and streets with high capacity. This might be due to the fact that monthly-averaged traffic flows assigned on these roads are also higher on these roads so that the absolute standard deviations are high. However, if we measure the uncertainty with the *relative standard variation (RSD)*, we can see that traffic flows on roads with big capacities are less uncertain than the traffic flows computed on roads with less traffic. The temporally-averaged RSD map is displayed in Figure 7.18. $RSD > 1.0$ means very high relative uncertainty of computed traffic flow. Figure 7.18 show that even though the value of σ_a is higher on highways and roads with high assigned traffic flows, the RSD on these roads remain less than 1.0. High relative uncertainty is found on links with low monthly-averaged traffic flow μ_a. The high values of RSD ($> 100\,\%$) for computed traffic flows are found on links with $\mu_a < 60\,\mathrm{veh\,h^{-1}}$.

Figure 7.19 presents the relative frequency distributions of (i) the interquartile range (denoted as I^{flow}_{IQR}) and (ii) $5^{th} - 95^{th}$ percentile range (denoted as I^{flow}_{CI90}) of traffic flows

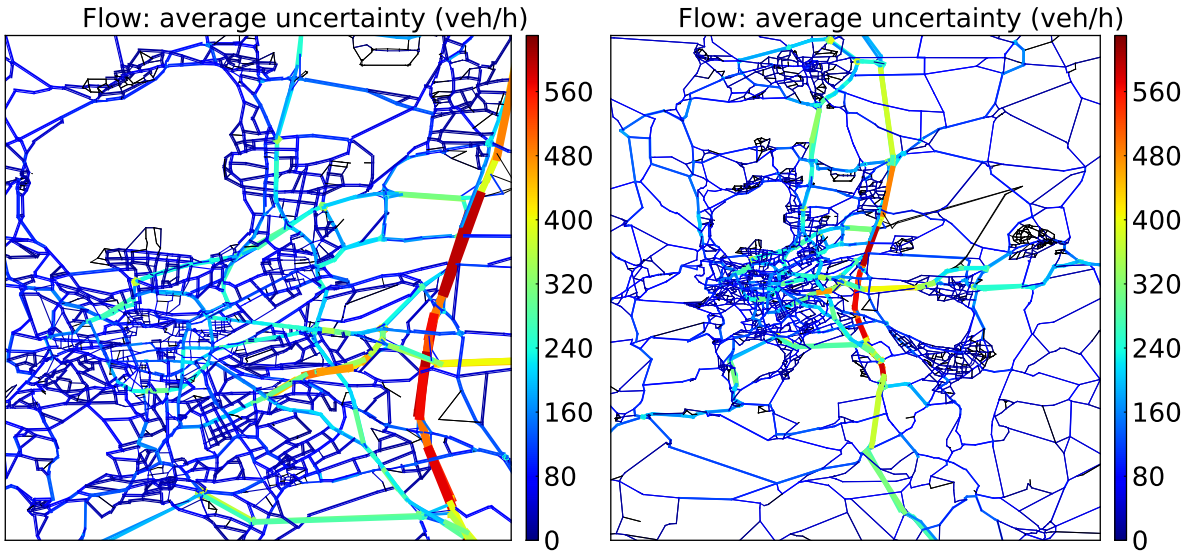


Figure 7.17: Monthly average of traffic flow uncertainty (σ_a) estimated by the DTA simulation ensemble during November 2014 over Clermont-Ferrand city (left) and the whole agglomeration (right) (in veh h^{-1}).

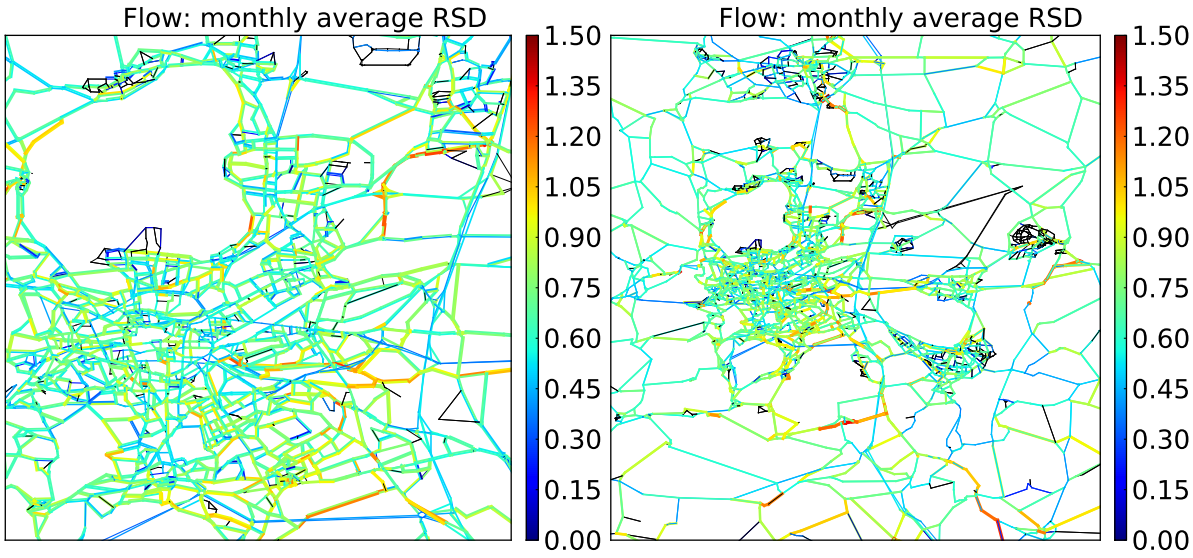


Figure 7.18: Monthly average of *relative standard deviation (RSD)* of traffic flow computed by the DTA simulation ensemble for November 2014 over Clermont-Ferrand city (left) and the whole agglomeration (right).

over the whole network. These two distributions are obtained based on traffic flows computed by the ensemble at all the links of the network and all time intervals of

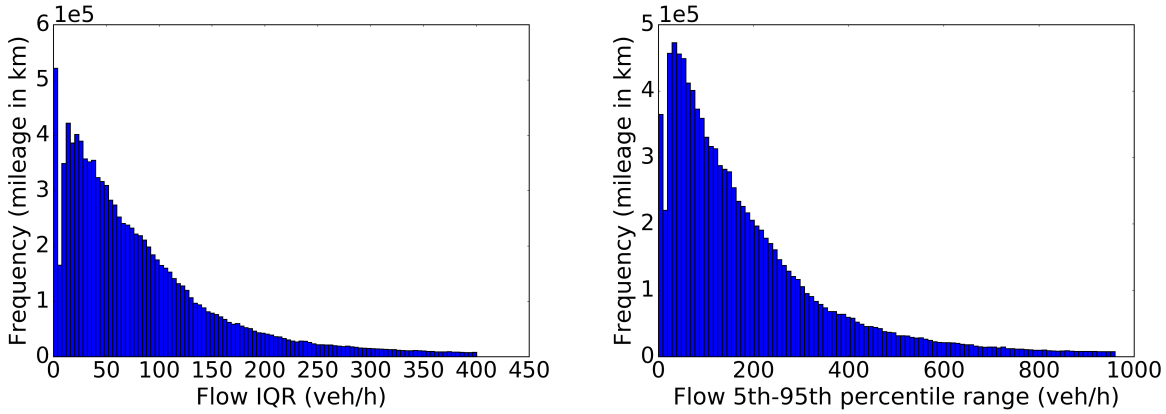


Figure 7.19: Relative frequency distribution of (i) the interquartile range (left) and (ii) $5^{th} - 95^{th}$ percentile range (right) of traffic flows over the whole network. The values whose occurrence frequencies are less than 0.025 are filtered out.

15 minutes during November 2014. The frequencies are weighted according to the lengths of links over the whole network. We have filtered out the values whose occurrence frequencies were less than 0.025. Figure 7.19 (left) shows that 97.5 % of I_{IQR}^{flow} forecasted by the ensemble are less than 400 veh h^{-1}. In addition, 68 % of I_{IQR}^{flow} are less than 100 veh h^{-1}. Figure 7.19 (right) shows that 97.5 % of the I_{CI90}^{flow} forecasted by the ensemble are less than 960 veh h^{-1}. It shows that 63 % of the forecasted I_{CI90}^{flow} are less than 200 veh h^{-1}. Figure 7.19 shows the results over the whole simulation period of November 2014 for all time intervals of 15 minutes. The relative frequency distributions of the interquartile range and $5^{th} - 95^{th}$ percentile range vary with the time of the day (e.g., during the peak hour and non-peak hour). These distributions also vary with the *type of day* (e.g., working day, Saturday, Sunday or public holiday, etc.). Figure 7.20 and Figure 7.21 show the interquartile ranges and $5^{th} - 95^{th}$ percentile ranges during non-peak periods and peak periods on different kinds of days during November 2014, over the whole network of the agglomeration. The values whose occurrence frequencies are less than 0.025 are filtered out. The frequencies are also normalized so that the integral of the complete distribution is 1.0.

These results show that for the same kind of day, the relative frequencies of high forecasted I_{IQR}^{flow} and I_{CI90}^{flow} are found on peak period. For different kinds of day, results show that higher values of I_{IQR}^{flow} and I_{CI90}^{flow} are frequent on working days than on week-ends or public holidays. On working days, the maximum values of I_{IQR}^{flow} and I_{CI90}^{flow} are also higher during the morning peak of 08:00 - 08:15 than during 00:00 - 00:15. In addition, there is a significant difference of the shown relative frequency distributions

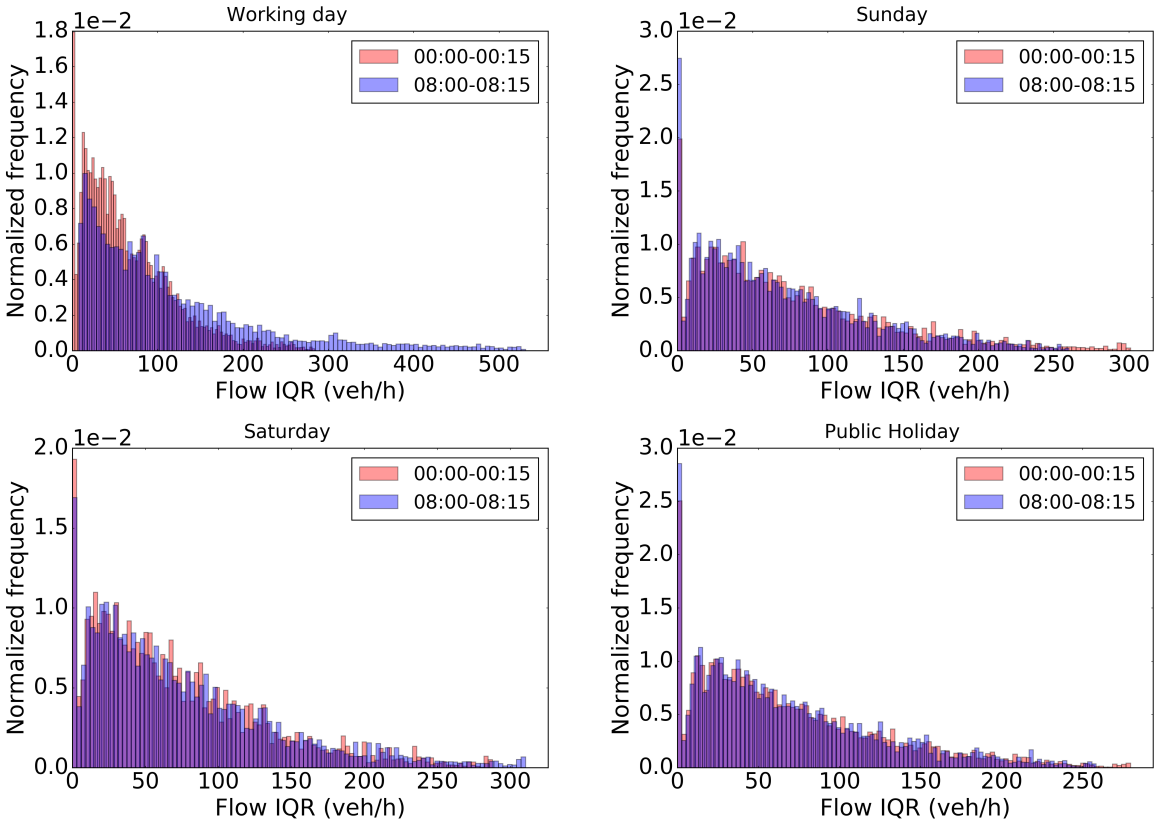


Figure 7.20: Relative frequency distributions of the interquartile ranges of computed traffic flows during two different time intervals (00:00 - 00:15 and 08:00 - 08:15), over different types of day during November 2014: working days, Saturdays, Sundays and public holidays. The values whose occurrence frequencies are less than 0.025 are filtered out. The frequencies are normalized so that the integral is 1.0 for each of the complete distributions.

between peak and non-peak period during working days, whereas this different is not that remarkable during other kinds of day. These results are mainly due to the fact that the traffic flow is higher during 08:00 - 08:15 than during 00:00 - 00:15, and the traffic volume is less heavy on weekends and holidays than on working days. These results are consistent with the results in Figure 7.17, showing that high uncertainty (measured by standard deviation) is found on links with heavy computed traffic flows. In addition, Table 7.13 shows the average I_{IQR}^{flow} and I_{CI90}^{flow} against measured traffic flows, on links with loop detectors in the network of Clermont-Ferrand. Results in Table 7.13 also show that the higher forecasted interquartile ranges and $5^{th} - 95^{th}$ ranges are found on roads with heavy measured traffic flows.

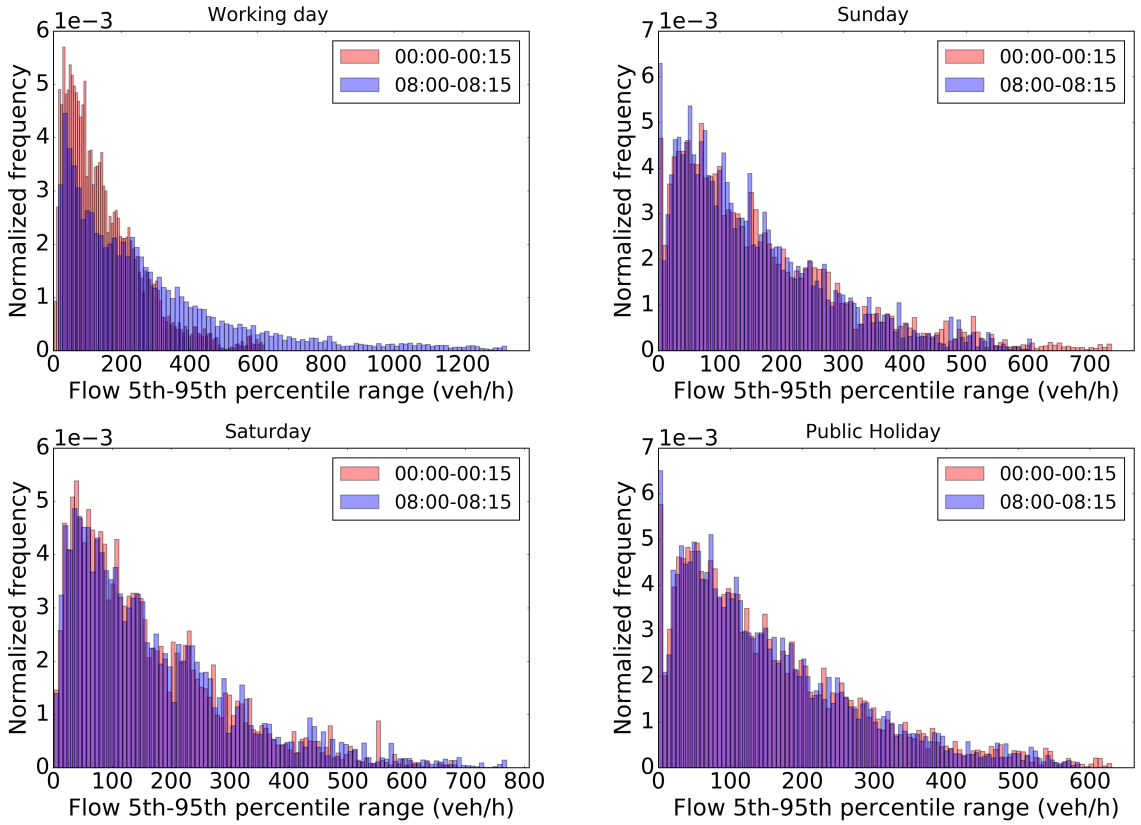


Figure 7.21: Relative frequency distributions of the $5^{th} - 95^{th}$ percentile ranges of computed traffic flows during two different time intervals (00:00 - 00:15 and 08:00 - 08:15), over different types of day during November 2014: working days, Saturdays, Sundays and public holidays. The values whose occurrence frequencies are less than 0.025 are filtered out. The frequencies are normalized so that the integral is 1.0 for each of the complete distributions.

Uncertainty of computed travel speeds

The original LADTA model also computes travel time for each time interval of 15 minutes. When building the metamodel for traffic flows, we also saved a training set of computed travel times at link resolution. The street lengths are constant and known. Then we can get a training set of average travel speed at link-level and at time resolution of each 15 minutes. We used the same metamodeling approach to build a metamodel for the travel speed. We then used the same variables as in Table 7.5 to carry out a Monte Carlo simulation for computing travel speed, during November 2014. Then an ensemble of travel speeds is generated and the uncertainty of computed travel speeds is obtained.

Table 7.13: We partition the observations of all the detectors and for all time intervals of 15 minutes into 20 intervals according to the measured traffic flow range (in veh h^{-1} detector^{-1}, as presented in columns 1 and 5). For the observations points in each interval, we calculate the mean of observed traffic flows (in veh h^{-1} detector^{-1}, presented in columns 2 and 6). For links with detectors whose measured traffic flow fall into an interval, we calculate the mean of interquartile ranges (denoted as IQR in veh h^{-1} detector^{-1}, presented in columns 3 and 7), and the corresponding mean of $5^{th} - 95^{th}$ ranges (denoted as I_{CI90} in veh h^{-1} detector^{-1}, presented in columns 4 and 8) forecasted by the ensemble of DTA simulations on the network of Clermont-Ferrand in November 2014.

Range	Avg. observation	Avg. IQR	Avg. I_{CI90}	Range	Avg. observation	Avg. IQR	Avg. I_{CI90}
0-50	17	66	141	500-550	523	182	412
50-100	72	96	211	550-600	573	185	421
100-150	122	114	254	600-650	623	188	428
150-200	173	128	286	650-700	673	192	439
200-250	223	138	312	700-750	722	201	459
250-300	273	146	333	750-800	772	208	474
300-350	323	155	353	800-850	822	211	483
350-400	373	163	372	850-900	872	215	494
400-450	423	171	389	900-950	922	216	502
450-500	474	177	403	≥ 950	971	221	515

We calculated the temporally-averaged uncertainty of travel speed at link resolution and plotted the uncertainty map of estimated travel speed in Figure 7.22. The RSD of estimated travel speed is presented in Figure 7.23.

Results in Figure 7.22 show that big speed uncertainty is found at crossroads and highways. The results in Figure 7.23 show that the spatial distribution of relative uncertainty is not the same as the spatial distribution of absolute uncertainty. Figure 7.24 compares the traffic flow uncertainty, relative uncertainty of travel speed, and link capacity. In fact, big traffic can be assigned on the upstream link a with big capacity. According to assumptions of LADTA model, if the capacity of the downstream link $a+1$ is smaller, traffic flow from link a is restricted by the link capacity of link $a+1$, and a vertical queue occurs at the exit point of link $a+1$. This will increase the average travel time on the link $a+1$ and the travel speed is therefore affected.

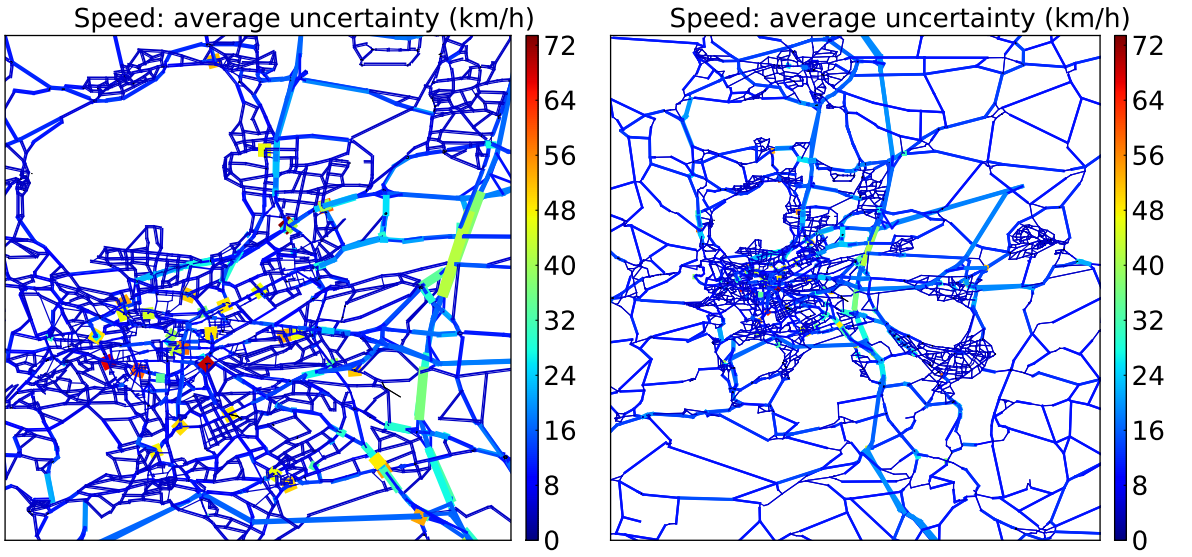


Figure 7.22: Monthly average of travel speed uncertainty from the ensemble of travel speeds for November 2014 over Clermont-Ferrand city (left) and the whole agglomeration (right) (in $\mathrm{km\,h^{-1}}$).

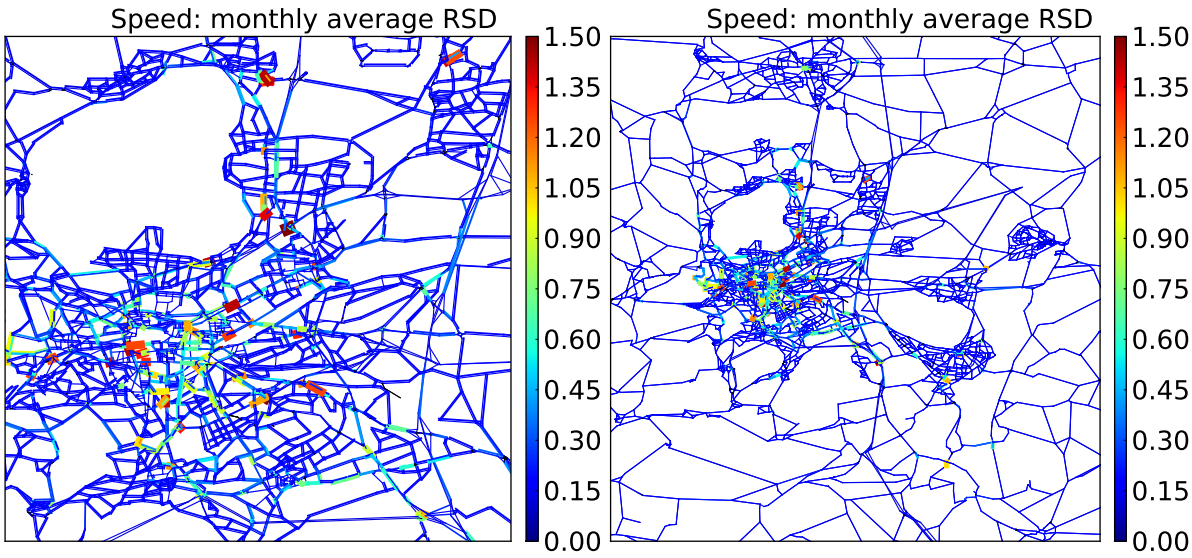


Figure 7.23: Monthly average of *relative standard deviation (RSD)* of average traffic speed computed by the ensemble of travel speeds for November 2014 over Clermont-Ferrand city (left) and the whole agglomeration (right).

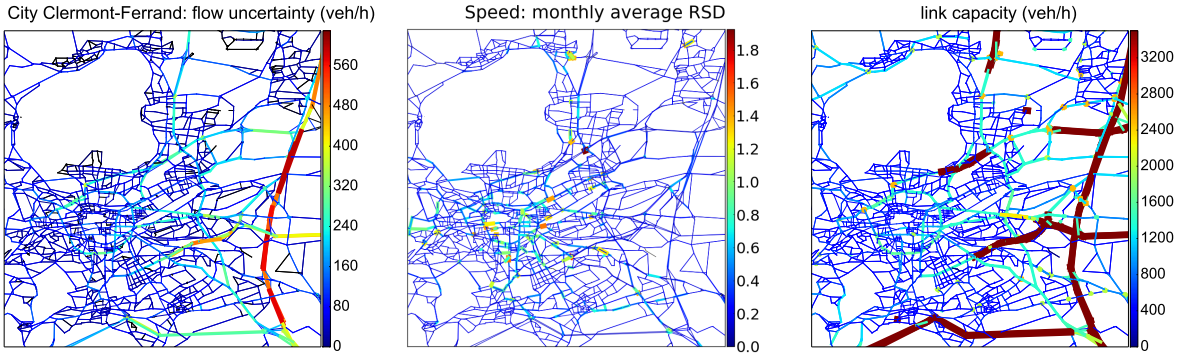


Figure 7.24: Monthly average traffic flow uncertainty (left, in veh h^{-1}), monthly average *relative standard deviation (RSD)* of traffic speed (center) and link capacity (right, in km h^{-1}) over the Clermont-Ferrand city.

7.4.2 Uncertainty of on-road emission simulations

Then we used the modeling chain built in Chapter 6 to generate the ensemble of on-road emissions for the agglomeration of Clermont-Ferrand during November 2014. The ensemble of traffic flows and travel speeds are inputs for the emission ensemble generation. We firstly generated an ensemble without changing the vehicle fleet composition, i.e, taking the values given from Table 6.1 and Table 6.2 in Chapter 6. Here we present only the ensemble of NOx emissions. For each member of the emission ensemble, the emission of NOx is computed at link level and at time resolution of 15 minutes. The resulted NOx emissions are presented in g km^{-1} (per 15 minutes).

We then calculated the temporally-averaged uncertainty of the computed emission at link resolution and plotted the uncertainty map in Figure 7.25. The *RSD* of estimated emissions is presented in Figure 7.26.

Results in Figure 7.25 and Figure 7.26 show that large emission uncertainty is found on links with large flow uncertainty. The uncertainty found in DTA simulation is propagated to the computed emissions through the modeling chain. Detailed results are presented in the following subsection 7.4.3.

Figure 7.27 presents the relative frequency distributions of (i) the interquartile range (denoted as I_{IQR}^{NOx}) and (ii) $5^{th}-95^{th}$ percentile range (denoted as I_{CI90}^{NOx}) of on-road traffic emissions over the whole network. These two distributions are obtained based on the ensemble of emissions at all the links of the network and all time intervals of 15 minutes during November 2014. The frequencies are weighted according to the lengths of links over the whole network. We have filtered out the values whose occurrence frequency is less than 0.025. Figure 7.27 (left) shows that 97.5 % of I_{IQR}^{NOx} forecasted by the ensemble

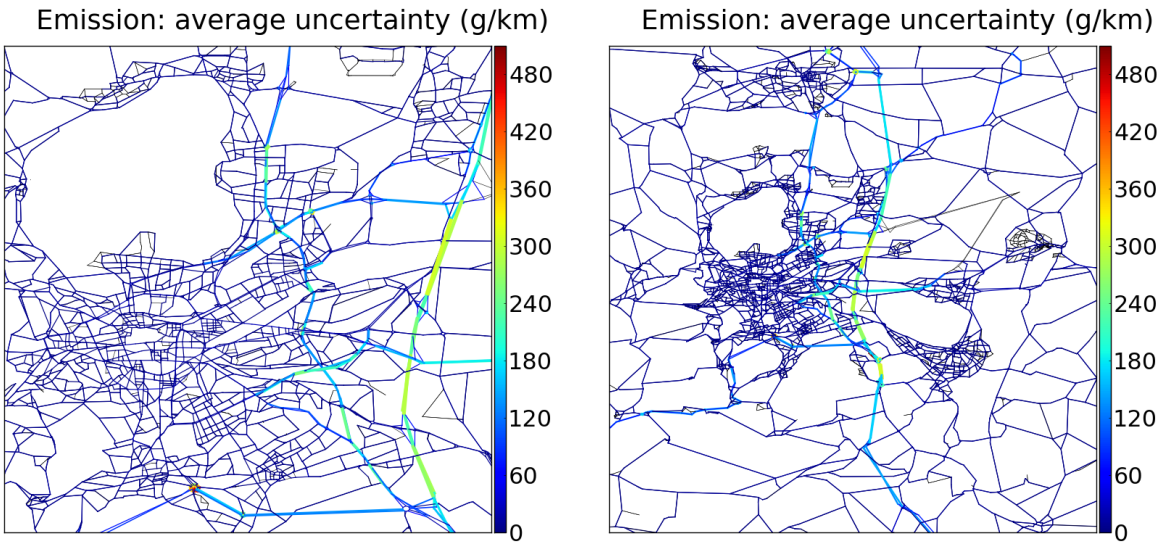


Figure 7.25: Monthly average of emission uncertainty from the ensemble of emission estimations for November 2014 over Clermont-Ferrand city (left) and the whole agglomeration (right) (in $\mathrm{g\,km^{-1}}$).

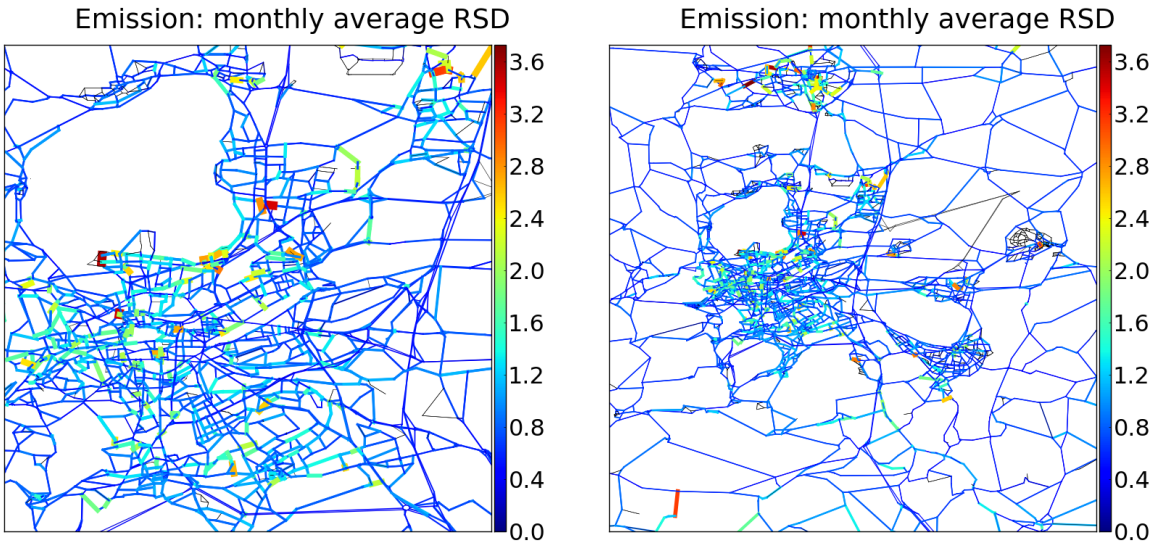


Figure 7.26: Monthly average of *relative standard deviation (RSD)* of on-road traffic emissions computed by the ensemble of emission estimations for November 2014 over Clermont-Ferrand city (left) and the whole agglomeration (right).

of NOx emissions are less than $38\,\mathrm{g\,km^{-1}}$ during an interval of 15 minutes. In addition, 78 % of the forecasted I_{IQR}^{NOx} are less than $10\,\mathrm{g\,km^{-1}}$. Figure 7.27 (right) shows that

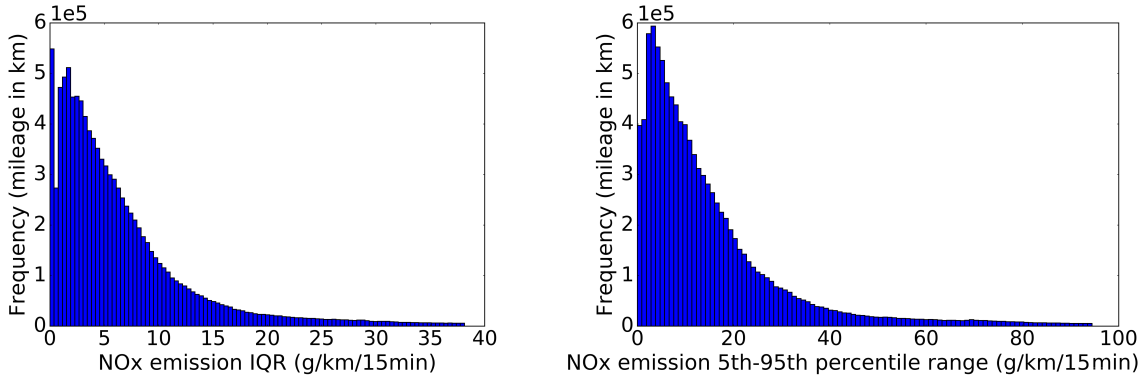


Figure 7.27: Relative frequency distribution of (i) the interquartile range (left) and (ii) $5^{th} - 95^{th}$ percentile range (right) of on road traffic emissions over the whole network. The values whose occurrence frequencies are less than 0.025 are filtered out.

97.5 % of the I_{CI90}^{NOx} forecasted by the NOx emission ensemble are less than $95\,\mathrm{g\,km^{-1}}$ during an interval of 15 minutes. It also shows that 74 % of the forecasted I_{CI90}^{NOx} are less than $20\,\mathrm{g\,km^{-1}}$ over the whole network during an interval of 15 minutes.

Figure 7.28 and Figure 7.29 show the interquartile ranges and $5^{th} - 95^{th}$ percentile ranges during non-peak periods and peak periods on different kinds of day during November 2014, over the whole network of the agglomeration. The values whose occurrence frequencies are less than 0.025 are filtered out. The frequencies are also normalized so that the integral of the complete distribution is 1.0. The results shown in Figure 7.28 and Figure 7.29 are consistent with the results for forecasted traffic flows. Higher I_{IQR}^{NOx} and I_{CI90}^{NOx} are more frequently found on peak periods (e.g., 08:00 - 08:15) of working days, than during non-peak periods (e.g., 00:00 - 00:15) or during non-working days.

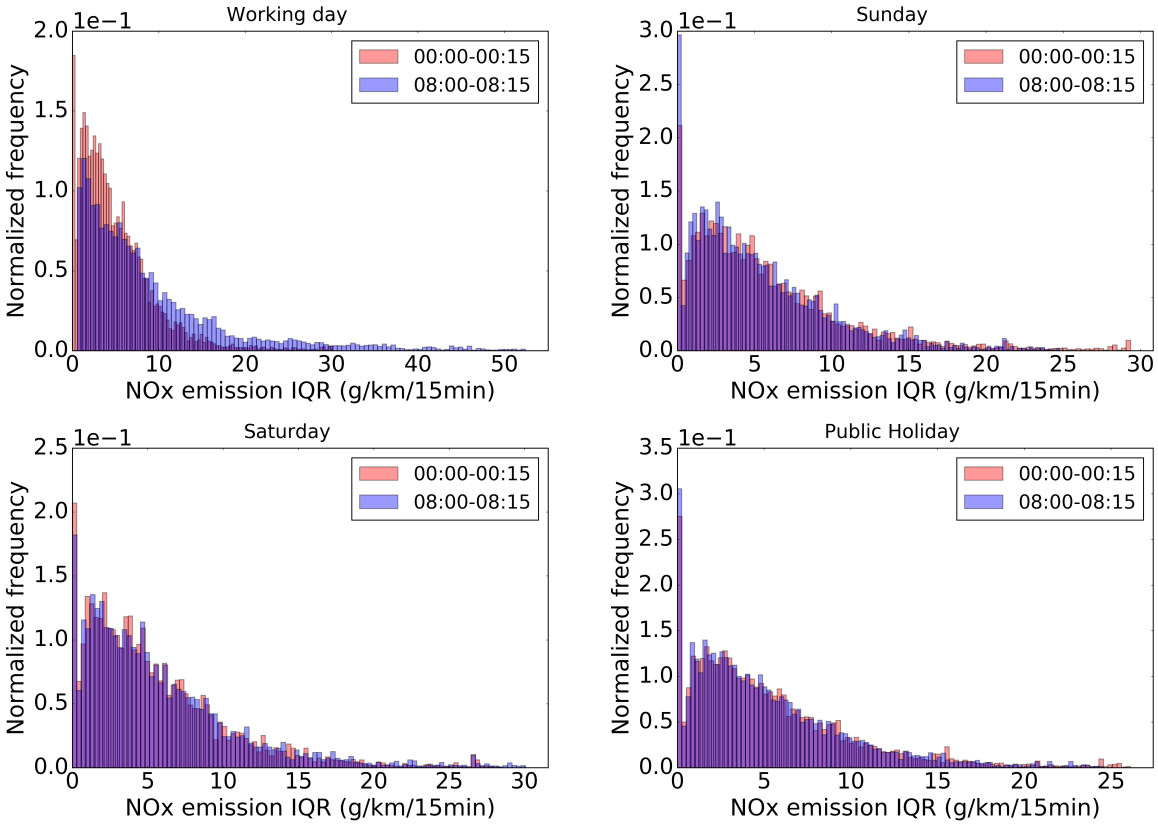


Figure 7.28: Relative frequency distributions of the interquartile ranges of computed NOx emissions during two different time intervals (00:00 - 00:15 and 08:00 - 08:15), over different types of day during November 2014: working days, Saturdays, Sundays and public holidays. The values whose occurrence frequencies are less than 0.025 are filtered out. The frequencies are normalized so that the integral is 1.0 for each of the complete distributions.

7.4.3 Uncertainty propagation from inputs of traffic assignment to on-road emission simulations

We calculated the correlation between the uncertainties of (i) computed traffic flow, (ii) average speed and (iii) traffic emissions. The results are presented in Table 7.14.

The results in Table 7.14 show that the monthly average traffic flow, flow uncertainty and emission uncertainty are highly correlated. This is due to the structure of the modeling chain, combing link-level traffic flow with emission factors computed by COPERT IV model. In fact, on a link with a given vehicle fleet data and travel speed, the on-road traffic emission computed by the modeling chain is proportional to the com-

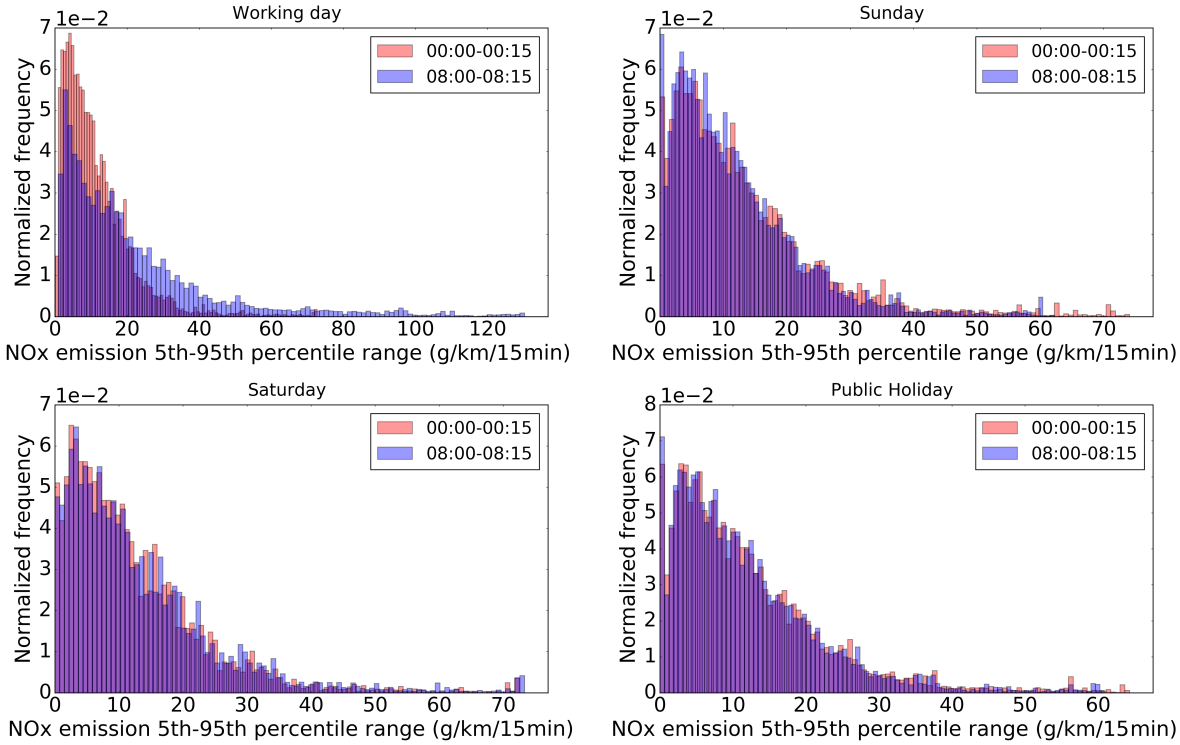


Figure 7.29: Relative frequency distributions of the $5^{th} - 95^{th}$ percentile ranges of computed NOx emissions during two different time intervals (00:00 - 00:15 and 08:00 - 08:15), over different types of day during November 2014: working days, Saturdays, Sundays and public holidays. The values whose occurrence frequencies are less than 0.025 are filtered out. The frequencies are normalized so that the integral is 1.0 for each of the complete distributions.

puted traffic flow on the link. However, the correlation is not perfectly equal to one because the heavy duty vehicles (HDVs) are not uniformly distributed on the network. If the share of HDVs is zero, we can see that the correlation between the traffic flow uncertainty and emission is almost 1.0. Figure 7.30 (up) and Figure 7.30 (down) show the log plots of monthly average traffic flow, monthly average flow uncertainty and emission uncertainty, of the two ensembles of road traffic emissions generated in two cases: (i) when there is no heavy duty vehicle (HDV) on the road network, and (ii) when there are additional HDVs with a volume equal to 30 % of the simulated vehicles on the roads near the logistic center of Clermont-Ferrand (cf. Figure 6.10).

Table 7.14: Correlation between the uncertainties of traffic flow, travel speeds and emissions. "-" denotes the case where the results are from two ensembles of emissions computed from the ensemble of the dynamic traffic assignments in two cases: (i) when there is no heavy duty vehicle (HDV) on the road network, and (ii) when there are additional HDVs with a volume equal to 30 % of the simulated vehicles on the roads near the logistic center of Clermont-Ferrand (cf. Figure 6.10).

	Standard deviation of			
	flow	speed	emission	emission without HDV
flow uncertainty	1.00	0.38	0.71	0.97
speed uncertainty	0.38	1.00	0.24	0.41
emission uncertainty	0.71	0.24	1.00	-
emission uncertainty without HDV	0.97	0.41	-	1.00
average flow	0.96	0.38	0.71	0.95
link capacity	0.03	0.0	0.03	0.03
speed limit	0.35	0.45	0.25	0.31

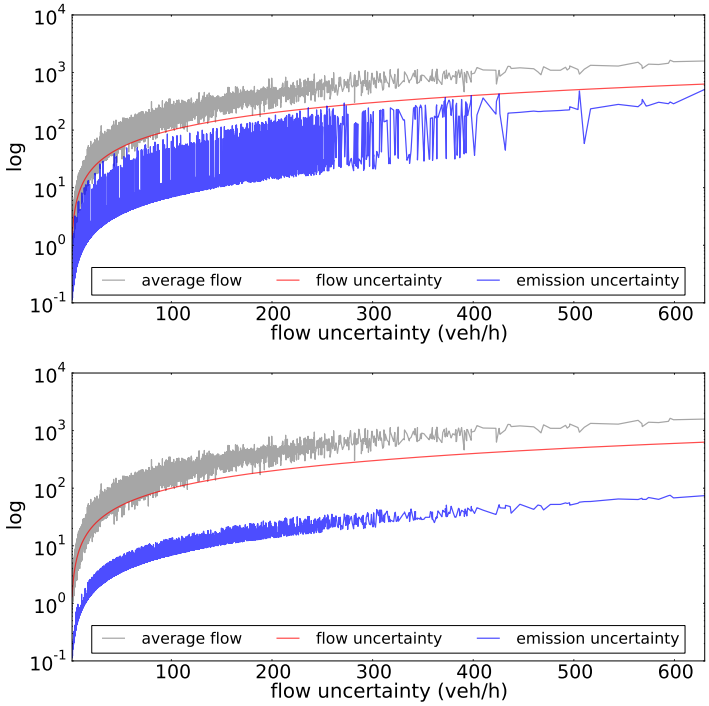


Figure 7.30: Log plots of monthly-averaged traffic flow (in veh h^{-1}), uncertainty (measured by standard deviation) of traffic flow (red line) and emission (blue line) against the flow uncertainty (x axis). The plotted values are sorted by the increasing order of traffic flow uncertainty. The emission uncertainty presented in the upper plot is computed from the ensemble of emissions with the presence of HDV. The uncertainty in lower plot is computed from the ensemble of emissions without HDV on the network.

Figure 7.30 confirms the correlation between the monthly average compute flow, flow uncertainty and emission uncertainty. Figure 7.31 shows the comparison between the relative uncertainties of the monthly average flow and emission. The relative uncertainty of computed emissions are almost always bigger than the relative uncertainty of computed traffic flow. In other words, if the given input uncertainty shown in Table 7.5 leads to 50 % of traffic flow uncertainty on a link a, the resulting traffic emission uncertainty would be not less than 50 %.

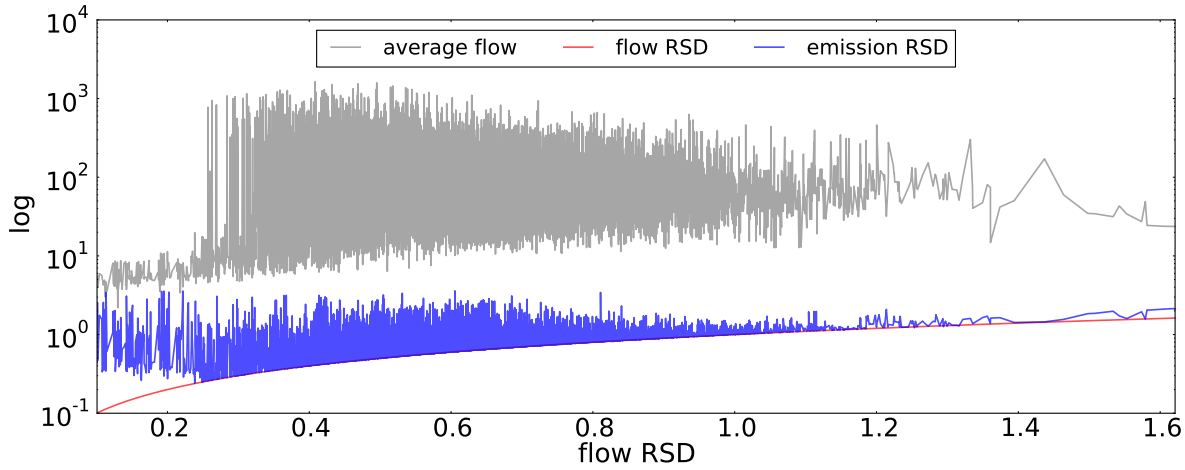


Figure 7.31: Log plots of monthly average traffic flow, flow uncertainty and emission uncertainty against the relative uncertainty of traffic flow. The plotted values are sorted in increasing order of relative uncertainty on traffic flow.

Results in Table 7.14 also show that the correlation between the traffic flow uncertainty and speed uncertainty is not high. This may due to the assumptions of the original LADTA model. On the one hand, we assume that the travel time equals to the free flow travel time when the traffic flow does not exceed the link capacity. If the flow uncertainty is big but the traffic flow does not reach the link capacity, the travel time is constant and equals to the free flow travel time. Then the uncertainty on the average travel speed is low, yet the uncertainty on emissions is big due to the high uncertainty on traffic flow. On the other hand, the traffic flow computed by LADTA model cannot exceed the link capacity. On links with low capacities, the traffic flow might reach the link capacity and bring congestion on the link. The traffic flow does not increase any more but the travel time increases, according to the bottleneck flowing model of LADTA model. Therefore, the uncertainty in the traffic flow might not be high but the uncertainty in computed travel speed might be.

Conclusions and perspectives

In this chapter, we considered that the inputs of the modeling chain are random variables with known probability distributions. We then used the Monte Carlo approach to generate two ensembles of simulations: (i) an ensemble of DTA simulations with the metamodel of LADTA, and (ii) an ensemble of emission simulations, for the agglomeration of Clermont-Ferrand at street resolution and at all time intervals of 15 minutes during November 2014. The objective of uncertainty quantification was then to estimate the probability distribution of the outputs of the generated ensemble, taking into account the given uncertainty lying in inputs. The measure of uncertainty was (i) the standard deviation, and (ii) the relative standard deviation of the output variables predicted by the ensemble

We recall here that the outputs of LADTA applied to Clermont-Ferrand (LCF model) are link-resolution traffic flows for all time intervals of 15 minutes, during November 2014. The uncertainty given to inputs are (i) $\pm 100\,\%$ on the total demand compared to the inputs for the deterministic simulation, (ii) $[-100\,\%, 80\,\%]$ in the O-D matrix for each of the five categories of O-D pairs, and (iii) $\pm 50\,\%$ for link capacities and speed limits. We evaluated the ensemble of DTA simulations with the help of traffic flow observations, measured by loop detectors on the network of Clermont-Ferrand during November 2014. Rank histogram, reliability diagram and statistical scores were chosen criteria to evaluate whether or not the ensemble can well represent the uncertainty of the DTA simulations. Results show that the ensemble of DTA simulations overestimate the temporal uncertainty and slightly underestimate the spatial uncertainty, with the chosen criteria of ensemble evaluation in this section. We also found that there is a systematic negative bias on a group of detectors, especially when the traffic on the network is heavy. These results may be due to (i) the representativeness errors; (ii) errors in estimating link capacities; (iii) the assumptions in the original LADTA model for route choice; (iv) the losses due to the dimension reduction and emulation when building the metamodel, especially on links with low computed traffic flow; (v) the lack of spatial uncertainty in link characteristics and O-D matrix. The performance of the ensemble can be improved by re-adjusting the input distributions in order to optimize the evaluation criteria.

Then we carried out the uncertainty quantification using the built ensemble. For the DTA simulation in the agglomeration of Clermont-Ferrand, the high relative uncertainty ($\geq 100\,\%$) is mostly found on links where the computed traffic flow is low ($\leq 60\,\mathrm{veh\,h^{-1}}$). The relative uncertainty in travel speeds were mostly found on links for which the upstream links have bigger capacity than its own capacity. There is no clear correlation

between the uncertainty in travel speeds and the uncertainty in the traffic flow. For on-road traffic emissions, we built an ensemble of emissions based on the ensemble of DTA simulations. We only took into account the uncertainty of DTA simulation inputs and analyze the uncertainty of computed emissions due to the uncertainty in traffic flows. Results show that the emission uncertainty is highly correlated with the computed traffic flow on the network and its uncertainty. This result is consistent with the results of the global sensitivity analysis in Chapter 5, showing that the emissions are sensitive to traffic flows, but not significantly sensitive to traffic speeds. Results in this chapter also show the relative uncertainty tends to be amplified from the computed traffic flows to the computed traffic emissions, especially on links with heavy duty vehicles.

In future works, an ensemble of emissions can be generated with perturbations in both DTA simulation inputs and vehicle fleet composition inputs in order to study the uncertainty due to vehicle fleet composition. In addition, the performance of the ensemble of DTA simulations can be improved by (i) adding spatial perturbations on some uncertain inputs (in O-D matrix, link capacities and speed limits), (ii) adding spatial uncertain inputs in the parameters of the metamodel. The ensemble can also be calibrated by sampling a sub-ensemble for which the evaluation criteria are optimum.

Conclusions and Perspectives

1. Conclusions

The main objective of this PhD work was to quantify the uncertainty in the simulation of on-road emissions due to the inputs of a dynamic traffic assignment model and the inputs of the emission factor model, in a metropolitan area with street resolution. The main findings of this work are summarized as follows.

- **DUE-based dynamic traffic assignment (DTA) model applied to real-world network with more than 19,000 links and 8,000 nodes for a one-day DTA simulation**

  Chapter 3 gave the results of a DTA simulation with LADTA model, applied to the road network of the agglomeration of Clermont-Ferrand (France). A qualitative sensitivity analysis was also carried out. The model network for the agglomeration has more than 19,000 links and 8,000 nodes. The traffic demand is represented by an Origin-Destination (O-D) matrix with dimension 124×124. A reference simulation with static VISUM model was also carried out. The results computed by LADTA and VISUM were compared at link resolution. Results show that the correlation between the traffic flow computed by the two models was 0.87. The traffic flows computed by two models were also compared with loop detector measurements. Both models have limitations for predicting spatial distribution of traffic flow on the network when compared with traffic flow measurements. These may due to (i) the uncertainties in O-D matrix, (ii) assumptions for route choice criteria, (iii) errors in the modeled network and (iv) the conversion from link-level flows to detector-level flows by dividing the link-resolution flow by the number of lanes. A DTA simulation with LADTA model for a working Tuesday was then carried out for the same agglomeration with the modeled network in 2012. The computed flows were compared with loop detector measurements of all the working Tuesdays from September 2014 to July 2015. Results show that LADTA model can well predict temporal variation of spatially-averaged traffic flow in the whole agglomeration, even for a long period of more than 30 days with a temporal resolution of 15 minutes. The temporal normalized root-mean-square error (NRMSE) was about 6.6 % and the correlation was 0.99 when we compared the computed and measured spatially-averaged traffic flow for all time intervals. Results show that the DTA simulation with LADTA gives more detailed temporal evaluation of the traffic flows and congestions on the network. For example, the LADTA results

during morning peak hour from 07:00 to 08:30 show that the traffic flows on the network vary significantly within a quarter of an hour. In addition, the congestion on the network is represented by the increase of vehicle travel time with the same time resolution. The use of DTA model can better reflect the temporal variation of traffic flow and congestions with a high time resolution than static traffic assignment models. A qualitative sensitivity analysis was then carried out to analyze the sensitivity of computed flows with respect to total traffic demand and link speed limits. Results show that the total vehicle travel time (in veh·h) is very sensitive to the traffic demand, especially during evening peak hour due to the existing traffic on the network before the arrival of the peak hour. The total vehicle travel time is not very sensitive to the speed limits of the network. It increases almost linearly with the decrease of speed limits on the network.

- **Introducing a metamodeling method coupling (i) a model reduction based on principal component analysis and (ii) a statistical emulation, and building a metamodel of a DTA model applied to a metropolitan-scale network**
  Chapter 4 introduced the metamodeling approach coupling a dimensionality reduction and a statistical emulation. A reduced base for the model outputs was firstly built based on principal component analysis (PCA). The outputs of the target model were projected onto this reduced subspace. Then, the relations between the projection coefficients and the inputs of the original model were reproduced by a statistical emulator, based on radial basis functions (RBF). This metamodeling method was applied to the dynamic traffic assignment with LADTA model applied in the agglomeration of Clermont-Ferrand. Firstly, uncertain inputs for LADTA model applied in Clermont-Ferrand were represented by a vector of multiplicative coefficients. They were chosen in order to represent the uncertainties in (i) the temporal variations in the traffic demand, (ii) the spatial uncertainty in the O-D matrix, and (iii) the spatial uncertainty in network parameters (link capacities and speed limits). Secondly, a metamodel of LADTA applied to Clermont-Ferrand was built based on the uncertain inputs, using the metamodeling approach presented in Chapter 4. For a DTA simulation with LADTA before 12:00 (local time) of the day for example, a set of 3003 training points was generated with Latin Hypercube Sampling. The corresponding training values were the traffic flows computed by the original LADTA model using the training points. The performance of the metamodel was evaluated by comparing the outputs of the metamodel with the training values. Results show that even though there were some loses due to the dimension reduction and the interpolation, big errors between the metamodel and the original model were mainly found on links

with low computed traffic flows. Thirdly, the traffic flows computed by the original model and the metamodel were compared with loop detector measurements at all detectors and all time intervals of 15 minutes during November 2014. Scores show that the performance of the metamodel was similar to that of the original model during a one-month simulation, but the computational time for the simulation decreased from 2 days on 110 cores to less than 1 minute on one core. The spatio-temporal correlation between the metamodel and the traffic flow measurements during one month was about 0.7. Note that the metamodeling treats the traffic assignment (TA) model as a black box, and builds reasonable relationships between the TA inputs and outputs for a given road network. The independence of the method to the model type gives us new insights on the application of metamodeling to TA problems on large scale networks during long-term periods, and opens the way to further studies such as global sensitivity analysis and uncertainty quantification. The proposed metamodeling method can be applied to other traffic assignment models for other cities and networks. The resulting metamodel preserves well the main features of the complete model but with very low computational cost.

- **Global sensitivity analysis (GSA) with a variance-based method (Sobol') for the dynamic traffic assignment simulation at metropolitan scale down to street resolution**

  With the metamodel built in Chapter 4, a global sensitivity analysis (GSA) was carried out on the DTA simulation with LADTA model applied in the agglomeration of Clermont-Ferrand. The Sobol' method was chosen for GSA studies. For an input p_i, first-order sensitivity index measures the effect of varying p_i alone, but averaged over variations in other input parameters. The total-effect index measures the contribution of p_i to the output variance, including all variance caused by its interactions with any other input variables. The first and total-effect Sobol' indices were computed for the computed traffic flows and average speeds at street level over the whole agglomeration, with respect to each uncertain input identified in Chapter 4.

  GSA results show that the computed traffic flows are very sensitive to the direction of traffic demands in the O-D matrix. They are also sensitive to the volume of traffic demands: (i) the temporal variation during the previous 0.5 hour before the target simulation time and (ii) the total demands between O-D pairs for which the inter-zone distance is between 0 - 5 km. The computed flows are not very sensitive to the link capacity or speed limit in our case study. Besides, the resulting Sobol' indices also depend on spatial location and characteristics of the links. On the one hand, the computed flows on highways are sensitive (i) to the traffic demands departing

from 0.5 hour before the target simulation period and (ii) to the traffic demands between long-distance O-D zones. On the other hand, the computed flows on links in the city center are sensitive (i) to the traffic demands during the previous 0.25 hour and (ii) the demands between short-distance O-D zones (less than 5 km). For the computed average speeds in the agglomeration, they are sensitive to the direction of traffic demand and the uncertain inputs corresponding to the street speed limits. The computed speeds are not significantly sensitive to other inputs on most of the streets in our case study. However, results show that on some street on which the capacity is low, the computed speeds are sensitive to the traffic demands during the previous 0.5 hour before the target simulation time. This suggests that traffic jams are more likely to happen on these streets.

- **Building a modeling chain coupling a dynamic traffic assignment model with an average-speed-based emission model (COPERT IV) at metropolitan scale with street resolution**

  In Chapter 6, a modeling chain for emission estimations was built by combining the dynamic traffic assignment model LADTA, and the COPERT IV model. The use of a dynamic traffic assignment (DTA) model allows us to estimate emissions with a finer time resolution in order to compute time-varying emissions due to on-road traffic. For example, in our case study in the agglomeration of Clermont-Ferrand, the emission calculations are based on traffic flow results of LADTA for each link during every 15 minutes for a working day. Moreover, with the metamodel built in Chapter 4, a modeling chain combing the traffic metamodel and the COPERT IV model was also built. With this modeling chain, CO and NOx emissions during the whole year of 2014 were computed, at a time resolution of 15 minutes.

  With the built modeling chain, qualitative and quantitative sensitivity analyses were carried out in order to study the sensitivity of the estimated emission with respect to both the inputs of LADTA model and COPERT IV. Qualitative sensitivity analysis results show that total NOx emissions increase with the total demand volume onto the network. They are less sensitive to the variation of network speed limits than to the traffic demand. As we decreased the links' speed limits, the total emissions firstly decreased and then increased. This non-linearity is mainly due to the influence of vehicle travel speed on the hot emission factors of NOx. Quantitative GSA with Sobol' method was also carried out on the modeling chain for computing street-level emission. The results of GSA show that the computed NOx emissions are sensitive to all the inputs that influence the traffic flows. The traffic speeds do not have significant influence to the NOx emissions computed by modeling chain in our case study. In

addition, the computed NOx emissions are very sensitive to some of the vehicle fleet inputs: the share of heavy duty vehicles (HDVs) and the percentage of gasoline cars on road. Other fleet parameters such as the emission standards and engine capacity, are not very influential to the on-road NOx emissions computed by the modeling chain. Nevertheless, these sensitivity results are different for different pollutants. For the hot emissions of CO, only the gasoline car share has a first-order effect on the emissions computed by the modeling chain. The share of HDVs is not as much influential to the computed CO emissions, as they are to NOx emissions. Moreover, the hot emissions of CO are affected by high-order effects of emission standard factors: the share of passenger cars for which the emission standard is Euro 4 and higher. This result is not found in the case of GSA study of NOx emission.

Though this work only presents the application of the built modeling chain to the network of Clermont-Ferrand for CO and NOx emissions, the computation of link-resolution emissions is implemented in the open-source code *Pollemission* Chen and Mallet [2016], for various types of vehicles (e.g., passenger cars, heavy duty vehicles, light utility vehicles, motorcycles, buses, etc.) and pollutants (e.g., CO, NOx, PM, fuel consumption, hydrocarbon, etc.). It can be applied to other cities as well.

- **Introducing a whole framework of uncertainty quantification of the computed traffic flow and its emissions due to inputs of DTA model for a full city scale down to street resolution, and presenting some preliminary results of uncertainty quantification and ensemble calibration**

  In Chapter 7, inputs of the LADTA model applied in Clermont-Ferrand were considered as random variables with uniform distributions. The uncertainties given to inputs are $\pm 100\,\%$ in the temporal variation of traffic demand, $[-100\,\%, 80\,\%]$ in the O-D matrix, and $\pm 50\,\%$ for link capacities and speed limits. A Monte Carlo approach with simple size $n = 100$ was then used to generate two ensembles of simulations: (i) an ensemble of DTA simulations and (ii) an ensemble of emission simulations, for the agglomeration of Clermont-Ferrand at street resolution and at all time intervals of 15 minutes during November 2014. The measure of uncertainty is the standard deviation and relative standard deviation of the traffic flows, average speeds and emissions predicted by the ensembles.

  Before the uncertainty quantification, the ensemble of DTA simulations was evaluated with the help of traffic flows observations, measured by loop detectors on the network of Clermont-Ferrand during November 2014. Rank histogram, reliability diagram and statistical scores were chosen criteria to evaluate whether the ensemble could well represent the uncertainty of the DTA simulations. Results show that the

ensemble of DTA simulations gives too-large uncertainty on the spatially-averaged traffic flows (hence overestimates the temporal uncertainty), but slightly underestimates the uncertainty of the temporally-averaged traffic flows. Results also show that on some detectors, the traffic flows measured by loop detectors are often higher than the upper envelope of the ensemble of DTA simulations, especially when the measured traffic on the network is heavy. These results may be due to (i) not having taken measurements uncertainty into account, (ii) the lack of spatial uncertainty in the parameterization of the uncertain inputs (in the O-D matrix, link capacities and speed limits, highway tolls). The performance of the ensemble can be improved by re-adjusting the input distributions in order to optimize an evaluation criteria.

Concerning uncertainty quantification results for the DTA simulation in the agglomeration of Clermont-Ferrand, big uncertainty on traffic flow was found (i) on links with heavy traffic and (ii) at time intervals and days with heavy traffic (e.g., during peak hours on working days). However, the high relative uncertainty ($\geq 100\,\%$) was mostly found on links where the computed traffic flow is low ($\leq 60\,\mathrm{veh\,h^{-1}}$). The relative uncertainty of computed travel speeds were mostly found on links where there is a decrease of link capacity when compared with upstream links. There is no clear correlation between the uncertainty of travel speeds and that of the traffic flow. For on-road traffic emissions, only uncertainty in DTA inputs were taken into account. Results show that the emission uncertainty is highly correlated with (i) the computed traffic flow on the network and (ii) the uncertainty of traffic flow. This result is consistent with the results of the global sensitivity analysis in Chapter 5, showing that the computed emissions are sensitive to the inputs that are influential to the traffic flows computed by LADTA model. Uncertainty quantification results also show the relative uncertainty tends to be amplified from the computed traffic flows to the computed traffic emissions, especially on links with heavy duty vehicles.

2. Future research perspectives

In view of what has been achieved during this PhD work, extensions and new lines of research for the future can be identified.

DTA simulation at metropolitan scale with LADTA model

More detailed investigation of the traffic flow data measured by loop detectors may help to calibrate the DTA simulation. From the conclusions of this work, more studies to decrease the representativeness error should be carried out. The distribution of traffic

on different lanes on a road can be obtained from the measurements. This distribution can then be applied to the conversion of link-level traffic flow to lane-resolution traffic flow. In addition, instead of comparing the computed and observed traffic flows at lane resolution for all detectors, one can also convert lane-resolution traffic flow to link-level flow so that the comparison can be carried out at link resolution. This would decrease the representativeness error on links where each lane has a detector so that the link-level traffic flow can be obtained by summing up the lane-resolution measured flows. Besides, the flowing model in LADTA is the bottleneck model. The bottleneck delay depends on the time passed in point queues. For a given link in congested condition, the length of queue depends on the modeled link capacity which is an estimated value based on traffic flow data and the fundamental diagram. With the help of traffic flow measured by loop detectors, link capacities can be calibrated before carrying out DTA simulation with LADTA model. This would help us to avoid the case where the measured traffic flows are always higher than the estimated capacity when the traffic is heavy.

DTA model calibration for a very large scale network is a challenging task. The performance of the DTA simulation can also be improved by (i) better modeling route choice models with multi-class users, (ii) better modeling the physical queues with other kind of flowing models in order to take into account the spill-back effect when the network is very congested, and by (iii) better modeling the O-D matrix with calibration based on loop detector measurements, etc.

Modeling chain for estimating the on-road traffic emissions and sensitivity analysis

In the modeling chain built in this PhD work, the cold-start emissions have not been taken into consideration. This might be an uncertainty source of the estimated emissions. However, the formulations for taking into account the cold-start emissions [EEA, 2016] are already implemented in *Pollemission* [Chen and Mallet, 2016]. According to COPERT IV model, the cold-start emission is computed based on hot-start emissions, the ambient temperature and the travel distance for which the vehicles are considered to be in the warm-up phase. At least two inputs can be added into the modeling chain for estimating the cold-start emissions. We would then be able to analyze the sensitivity and uncertainty of the total emission with consideration of the uncertainty of inputs related to cold-start emissions.

In addition, the chosen emission model is the COPERT IV model based on average speed. At urban area with dense traffic and traffic lights, the distribution of the travel speed on a road is not homogeneous. Taking into account the distribution of travel speed

on a street instead of using the average speed could be a solution [Aguiléra and Tordeux, 2014]. The travel speed distributions can be obtained from loop detector observations.

With the built link-resolution modeling chain for estimating road traffic emissions, more inputs can be studied in the global sensitivity analysis (GSA). For example, we can estimate the emissions due to buses on the network and then analyze the sensitivity of these emissions with respect to the bus frequency and fleet composition of buses. In addition, if the DTA simulation is carried out for multi-class users, i.e., multi-class users are added in the O-D matrix in inputs (e.g., light utility vehicles, heavy duty vehicles, motorcycles, etc.), emissions from other user classes can also be estimated. Sensitivity analysis for new categories of users can help us to identify whether the uncertain inputs concerning these users can influence the computed traffic emissions on the network. Based on these eventual results, we can then decide whether or not to add this new vehicle fleet composition into the Monte Carlo simulations to carry out uncertainty quantification.

Metamodeling of the DTA model

Parameterization of *uncertain* inputs

The metamodeling of the LADTA model applied to the agglomeration of Clermont-Ferrand is an essential step for this PhD work. It is built based on *uncertain inputs*. Results of the uncertainty quantification show that the generated ensemble still slightly underestimates the spatial uncertainty, even though we have already overestimated the uncertainty of spatially-averaged traffic flows. There are still more than 10 % of the observation points for which the measured traffic flows are not covered by the $5^{th} - 95^{th}$ percentile range of the generated ensemble of DTA simulations. One of the main reasons is that there is a lack of parameterization in the representation of spatial uncertain inputs. We have only considered 5 uncertain inputs in the O-D matrix for representing spatial uncertainty in traffic demand, and 4 uncertain inputs for the modeled network (for link capacities and speed limits). According to local and global sensitivity analysis, traffic demand is the most influential input for the computed traffic flow on the network. One may add inputs that represent the spatial uncertainty in the O-D matrix in the metamodeling parameterization, in order to bring more spatial uncertainty in the computed traffic flow. Concerning the parameters in the modeled network, sensitivity analysis results in this work show that the computed traffic flow in the agglomeration is not very sensitive to the variation of speed limits or link capacities. However, in the experiment setting of our case study, there are only a few inputs that represent the *spa-*

tial uncertainties in speed limits and link capacities. We have only focused on the road traffic variation by scaling the speed limits of all the links on the network in the qualitative sensitivity analysis. In the quantitative GSA study, links have been divided into two categories with respect to their speed limits, and two other categories with respect to their capacities. The GSA was then carried out by scaling speed limits or capacities of links according to their *categories*. Spatial uncertainty can be added by adding more spatial uncertain inputs in the parameterization associated to the modeled network. For example, we could add more categories of links with respect to link speed limits and capacities. We could then define the corresponding scalar coefficients to multiply to the reference speed limits and capacities of the modeled network. In fact, link travel time is modeled by the sum of (i) free-flow travel time and (ii) the bottleneck delay in point queues in the flowing model of LADTA. With the assumption that travelers choose their *best* routes in the network by minimizing their own travel costs, the spatial distribution of both speed limits and link capacities can influence travelers' route choice strategies, and thus affect the spatial distribution of the traffic flows from the DTA simulation. On the one hand, speed limits can influence the free-flow travel time when the network is not congested. On the other hand, link capacities can influence the lengths of point queues and then influence the delay time spent in queues. In addition, it is shown in Chapter 3 that when we took into account the highway toll, the *value of time* could influence the spatial distribution of traffic flow, too. It is worth considering the *value of time* as an uncertain input and give it a distribution. This would allow us to study the uncertainty of traffic flows due to the *value of time* input.

However, there should be a tradeoff between (i) the dimension of the input and (ii) the representativeness of spatial uncertainty in inputs. The input dimension should be moderate in order to assure the performance of the metamodel, the global sensitivity analysis and the ensemble simulations with Monte Carlo approach [Saltelli et al., 2008].

Operational applications

There are various applications of the metamodeling to the operational context. For example, it can be used in defining congestion zones in order to prevent very low level of service on certain roads in a city. Therefore, instead of defining link categories with respect to their speed limits or capacities, we can also define the categories of roads with respect to different congestion zones. We can then add toll price on these roads. A metamodel can be built by taking into account the toll prices for different categories. Then a GSA can be carried out to study whether or not the traffic flow in these zones would be sensitive to the congestion charging prices, under consideration of

the uncertainty in other inputs such as traffic demand in the O-D matrix. Similarly, one can define restricted zones in a city where users can only drive below a certain speed limit. When building the metamodel of the DTA model, We can add a scalar input with respect to the new speed limits on roads in the restricted zones. Then a GSA can be carried out to see how this measure can influence the spatial distribution and total volume of the road traffic volume and their resulting air pollutant emissions on the whole network.

Uncertainty quantification of dynamic traffic assignment simulation and emission estimation

Uncertainty quantification of the DTA simulation

We can deal with the underestimation of spatial uncertainty of traffic flow by adding spatial uncertainty in inputs, or by taking into account the representativeness error. Besides, the error in traffic flow measurements can also be taken into account when we make use of the observations to evaluate the reliability of the generated ensemble. The traffic flow measurements used in this PhD work have been considered to be perfect. For each observation point, the measured traffic flow was considered as a scalar o. If we consider that there are uncertainties in the observation, the latter can then be considered as a random variable O whose mean value is o. This would require an analysis of error in the loop detector measurements so that the distribution of O can be defined.

Concerning the uncertainty quantification of atmospheric pollutant emissions, only the uncertainty due to the uncertain inputs of traffic assignment was taken into account in our work. Two ensembles of emission simulations were generated, with or without taking into account the heavy duty vehicles (HDVs) on the network. However, the ratio of HDVs was given as constant on certain links and the uncertainty on this ratio has not been taken into account. The uncertainty regarding other inputs of the COPERT IV has not yet been taken into account in the uncertainty analysis. However, in the GSA study for the modeling chain coupling the DTA model with COPERT IV, we have found that the estimated emissions of NOx and CO were sensitive to the share of gasoline passenger cars (PCs). In addition, the share of emission standard of PCs and HDVs can also be taken into consideration when analyzing the uncertainty of emissions. With the built metamodel, the implementation of COPERT IV (*Pollemission* [Chen and Mallet, 2016]) and the built modeling chain, it will be straightforward to generate emission ensembles with consideration of the additional uncertainties in vehicle fleet inputs. It will be possible to generate emission ensembles of other pollutants such as for CO and particulate matters (PM) as well.

Ensemble improvement and calibration

We tried to improve the performance of the generated ensemble of DTA simulations. It is found that optimizing the ensemble according to reliability diagram criterion might worsen other ensemble performance evaluation criteria. In future research, we can build an ensemble with more members, and then select a sub-ensemble which is representative of uncertainty or shows good reliability for probabilistic forecasting of traffic flows. In addition, other criteria can also be used and optimized, in order to evaluate and calibrate the ensemble. For example, the minimization of the *continuous ranked probability score (CRPS)* is commonly used in the calibration of meteorological forecast ensembles [Gneiting et al., 2005; Junk et al., 2015]. It has also been successfully used in calibrating the ensemble for photovoltaic power forecasts [Thorey, 2017].

Application to transportation management and air quality simulation

It is shown in this PhD work that the generated ensemble of dynamic traffic assignments can be used to forecast the level of service (LOS) of roads in the network, under consideration of uncertainties in traffic demand and road network in a full metropolitan area. Although the uncertainty quantification approach presented in this work is based on the DTA simulation in a given agglomeration, the method can be used on another network and with other kinds of traffic models. The sensitivity analysis and uncertainty quantification methods can bring about new insights and establish new methods for assessing transportation systems through LOS forecasts, travel time/speed forecasts, road-traffic related emissions estimations, etc.

In addition, although we cannot directly measure the concentration of the atmospheric pollutants emitted by road traffic in a metropolitan area, the generated ensembles of different pollutants can provide *uncertain* emission inputs for air quality models. An ensemble of air quality simulations can then be generated, using the different traffic emissions and other perturbed input data, in order to quantify the uncertainty in air quality simulations. It is then possible to study the propagation of input uncertainty through the whole modeling chain from dynamic traffic assignment to the final estimation of air pollutant concentrations in a metropolitan area down to street resolution. This ensemble of air quality simulations can then be evaluated and calibrated by using the concentrations measured by air quality monitoring stations. These are in fact the main objectives of the ANR research project ESTIMAIR, of which this PhD work is an important part.

References

Aguiléra, V. (2014). *Rapport Habilitation à Diriger des Recherches: Technologies de l'information et systèmes de transport*. PhD thesis, Université de Paris-Est, 6-8 Avenue Blaise-Pascal, Cité Descartes, Champs-sur-Marne, France. (Cited on pages 7 and 37.)

Aguiléra, V. and Leurent, F. (2009). Large problems of dynamic network assignment and traffic equilibrium: Computational principles and application to paris road network. *Transportation Research Record, Journal of the Transportation Research Board*, 2132:122–132. (Cited on pages 6, 7, 15, 37, 43, 44, 52, 53, 91, 96, 97 and 133.)

Aguiléra, V. and Tordeux, A. (2014). A new kind of fundamental diagram with an application to road traffic emission modeling. *Journal of Advanced Transportation*, 48(2):165–184. (Cited on page 234.)

Aguiléra, V. and Tordeux, A. (2011). Integration of a speed-dependent emission model in dynamic traffic assignment: a large scale application to the paris metropolitan area. *Procedia - Social and Behavioral Sciences*, 20:475 – 484. The State of the Art in the European Quantitative Oriented Transportation and Logistics Research – 14th Euro Working Group on Transportation and 26th Mini Euro Conference and 1st European Scientific Conference on Air Transport. (Cited on page 15.)

Airparif (2014). Inventaire regional des émissions en Île-de-france. Airparif report. Technical report. (Cited on pages 1 and 150.)

Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9(7):1518–1530. (Cited on pages 27 and 187.)

André, M., Roche, A., and Bourcier, L. (2013). Statistiques de parcs et trafic pour le calcul des émissions transports routiers en france, Report IFSTTAR-LTE. Technical report. (Cited on pages 156 and 157.)

Ashok, K. and Ben-Akiva, M. E. (2000). Alternative approaches for real-time estimation and prediction of time-dependent origin–destination flows. *Transportation Science*, 34(1):21–36. (Cited on page 28.)

Azevedo, C. L., Ciuffo, B., Cardoso, J. L., and Ben-Akiva, M. E. (2015). Dealing with uncertainty in detailed calibration of traffic simulation models for safety assessment. *Transportation research part C: emerging technologies*, 58:395–412. (Cited on pages 20 and 89.)

Beckmann, M., McGuire, C., and Winsten, C. B. (1956). Studies in the economics of transportation. Technical report. (Cited on pages 12 and 86.)

Borgonovo, E. and Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3):869 – 887. (Cited on page 24.)

Boulter, P. and McCrae, I. (2007). Artemis: Assessment and reliability of transport emission models and inventory systems-final report. *TRL Published Project Report*. (Cited on pages 2 and 21.)

Boulter, P., McCrae, I. S., and Barlow, T. J. (2007). *A review of instantaneous emission models for road vehicles*. TRL Limited Wokingham, United Kingdom. (Cited on page 22.)

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthey Weather Review*, 78(1):1–3. (Cited on page 27.)

Bröcker, J. and Smith, L. A. (2007). Increasing the reliability of reliability diagrams. *Weather and forecasting*, 22(3):651–661. (Cited on page 188.)

Broomhead, D. S. and Lowe, D. (1988). Radial basis functions, multi-variable functional interpolation and adaptive networks. Technical report, Royal Signals and Radar Establishment Malvern (United Kingdom). (Cited on pages 19, 87, 88 and 93.)

Buhmann, M. D. (2003). *Radial basis functions: theory and implementations*, volume 12. Cambridge university press. (Cited on pages 19 and 88.)

Bureau, o. p. r. (1964). Traffic assignment manual. *US Department of Commerce*. (Cited on pages 13, 52 and 86.)

Calvert, S., Taale, H., Snelder, M., and Hoogendoorn, S. (2018). Improving traffic management through consideration of uncertainty and stochastics in traffic flow. *Case Studies on Transport Policy*. (Cited on page 28.)

Calvert, S. C., Taale, H., Snelder, M., and Hoogendoorn, S. (2012). Probability in traffic: a challenge for modelling. In *DTA2012: 4th International Symposium on Dynamic*

Traffic Assignment, Martha's Vineyard, USA, 4-6 June 2012; Authors version. (Cited on page 28.)

Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental modelling & software*, 22(10):1509–1518. (Cited on page 26.)

Campolongo, F. and Saltelli, A. (1997). Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliability Engineering & System Safety*, 57(1):49–69. (Cited on page 26.)

Campolongo, F., Saltelli, A., and Cariboni, J. (2011). From screening to quantitative sensitivity analysis. a unified approach. *Computer Physics Communications*, 182(4):978–988. (Cited on page 26.)

Candille, G. (2003). *Validation des systèmes de prévisions météorologiques probabilistes*. PhD thesis, Paris VI. (Cited on page 28.)

Candille, G. and Talagrand, O. (2005). Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal of the Royal Meteorological Society*, 131(609):2131–2150. (Cited on page 188.)

Carrese, S., Cipriani, E., Mannini, L., and Nigro, M. (2017). Dynamic demand estimation and prediction for traffic urban networks adopting new data sources. *Transportation Research Part C: Emerging Technologies*, 81:83–98. (Cited on page 28.)

Carteret, M., André, M., and Pasquier, A. (2014). Vehicle fleet composition assessment in the Île-de-France region to calculate the road traffic pollutant emissions. *Pollution Atmospherique*, (221). (Cited on page 3.)

Chen, A., Yang, H., Lo, H. K., and Tang, W. H. (1999). A capacity related reliability for transportation networks. *Journal of advanced transportation*, 33(2):183–200. (Cited on page 28.)

Chen, H.-K. and Hsueh, C.-F. (1998). A model and an algorithm for the dynamic user-optimal route choice problem. *Transportation Research Part B: Methodological*, 32(3):219–234. (Cited on pages 13 and 86.)

Chen, R., Aguiléra, V., Mallet, V., Cohn, F., Poulet, D., and Brocheton, F. (2017). A sensitivity study of road transportation emissions at metropolitan scale. *Journal of Earth Sciences & Geotechnical Engineering*, 7(1):151–173. (Cited on pages 43, 96, 97, 100, 125 and 151.)

Chen, R. and Mallet, V. (2016). Pollemission software computing traffic emissions of atmospheric pollutants with COPERT-IV formulations. https://github.com/pollemission. (Cited on pages 153, 231, 233 and 236.)

Chen, X. M., Zhang, L., He, X., Xiong, C., and Li, Z. (2014). Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network. *Computer-Aided Civil and Infrastructure Engineering*, 29(5):359–381. (Cited on pages 20 and 89.)

Chiu, Y.-C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., and Hicks, J. (2011). Dynamic traffic assignment: A primer. *Transportation Research E-Circular*, (E-C153). (Cited on pages 12, 37 and 86.)

Cho, H.-J., Smith, T. E., and Friesz, T. L. (2000). A reduction method for local sensitivity analysis of network equilibrium arc flows. *Transportation Research Part B: Methodological*, 34(1):31–51. (Cited on page 24.)

Chong, L. and Osorio, C. (2017). A simulation-based optimization algorithm for dynamic large-scale urban transportation problems. *Transportation Science*. (Cited on pages 20 and 89.)

Ciuffo, B., Casas, J., Montanino, M., Perarnau, J., and Punzo, V. (2013). Gaussian process metamodels for sensitivity analysis of traffic simulation models: Case study of aimsun mesoscopic model. *Transportation Research Record: Journal of the Transportation Research Board*, (2390):87–98. (Cited on pages 20, 26, 27 and 88.)

Cukier, R., Fortuin, C., Shuler, K. E., Petschek, A., and Schaibly, J. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i theory. *The Journal of chemical physics*, 59(8):3873–3878. (Cited on page 26.)

Cukier, R., Levine, H., and Shuler, K. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *Journal of computational physics*, 26(1):1–42. (Cited on page 25.)

da Rocha, T. V., Leclercq, L., Montanino, M., Parzani, C., Punzo, V., Ciuffo, B., and Villegas, D. (2015). Does traffic-related calibration of car-following models provide accurate estimations of vehicle emissions? *Transportation Research Part D: Transport and Environment*, 34:267 – 280. (Cited on pages 15, 22, 27 and 29.)

de Dios Ortúzar, J. and Willumsen, L. G. (2011). *Modelling transport*. John Wiley & Sons. (Cited on pages 12, 13, 15, 30, 37 and 52.)

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive theory of functions of several variables*, pages 85–100. Springer. (Cited on pages 19 and 88.)

Dyn, N., Levin, D., and Rippa, S. (1986). Numerical procedures for surface fitting of scattered data by radial functions. *SIAM Journal on Scientific and Statistical Computing*, 7(2):639–659. (Cited on pages 19, 87, 88 and 93.)

EEA (2016). *EMEP/EEA air pollutant emission inventory guidebook - Part B.1.A.3.b.i-iv Road transport*. European Environment Agency. (Cited on pages 2, 21, 22, 23, 152, 153 and 233.)

EEA (2017). *Air quality in Europe — 2017 report*. European Environment Agency. (Cited on page 1.)

Epstein, E. S. (1969a). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8(6):985–987. (Cited on page 27.)

Epstein, E. S. (1969b). Stochastic dynamic prediction. *Tellus*, 21(6):739–759. (Cited on page 27.)

Fellendorf, M., Nokel, K., and Handke, N. (2000). Visum-online-traffic management for the expo 2000 based on a traffic model. In *Proceedings of the 7th World Congress on Intelligent Systems*. (Cited on page 15.)

Fomunung, I., Washington, S., Guensler, R., and Bachman, W. (2001). Validation of the measure automobile emissions model: a statistical analysis. *Journal of Transportation and Statistics (September 2000)*, pages 65–84. (Cited on page 22.)

Franco, V., Kousoulidou, M., Muntean, M., Ntziachristos, L., Hausberger, S., and Dilara, P. (2013). Road vehicle emission factors development: A review. *Atmospheric Environment*, 70:84–97. (Cited on page 150.)

Friesz, T. L., Bernstein, D., Smith, T. E., Tobin, R. L., and Wie, B.-W. (1993). A variational inequality formulation of the dynamic network user equilibrium problem. *Operations Research*, 41(1):179–191. (Cited on pages 13 and 87.)

Friesz, T. L., Cho, H.-J., Mehta, N. J., Tobin, R. L., and Anandalingam, G. (1992). A simulated annealing approach to the network design problem with variational inequality constraints. *Transportation Science*, 26(1):18–26. (Cited on pages 13 and 87.)

Friesz, T. L., Luque, J., Tobin, R. L., and Wie, B.-W. (1989). Dynamic network traffic assignment considered as a continuous time optimal control problem. *Operations Research*, 37(6):893–901. (Cited on pages 13 and 87.)

Garaud, D. and Mallet, V. (2011). Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: Application to air quality. *Journal of Geophysical Research: Atmospheres*, 116(D19). (Cited on page 209.)

Garaud, D. and Mallet, V. (2012). Uncertainty estimation and decomposition based on monte carlo and multimodel photochemical simulations. (Cited on page 29.)

Ge, Q., Ciuffo, B., and Menendez, M. (2015). Combining screening and metamodel-based methods: An efficient sequential approach for the sensitivity analysis of model outputs. *Reliability Engineering & System Safety*, 134:334–344. (Cited on pages 26 and 27.)

Gendreau, M., Laporte, G., and Séguin, R. (1996). Stochastic vehicle routing. *European Journal of Operational Research*, 88(1):3–12. (Cited on page 28.)

Girard, S., Mallet, V., Korsakissok, I., and Mathieu, A. (2016). Emulation and sobol' sensitivity analysis of an atmospheric dispersion model applied to the fukushima nuclear accident. *Journal of Geophysical Research: Atmospheres*, 121(7):3484–3496. (Cited on pages 21, 26, 27 and 90.)

Gkatzoflias, D., Kouridis, C., Ntziachristos, L., and Samaras, Z. (2007). Copert 4: Computer programme to calculate emissions from road transport. *European Environment Agency*. (Cited on page 153.)

Gneiting, T., Raftery, A. E., Westveld III, A. H., and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review*, 133(5):1098–1118. (Cited on page 237.)

Grassot, L., Nicolas, J.-P., and Pluvinet, P. (2012). De l'intérêt de contrôler l'impact des hypothèses de composition du parc automobile sur l'estimation des émissions liées au trafic routierthe benefit of verifying the impact of the car fleet arrangement hypotheses on the estimation of road traffic emissions. *Recherche Transports sécurité*, 28(3-4):215–227. (Cited on page 3.)

Greenshields, B., Channing, W., Miller, H., et al. (1935). A study of traffic capacity. In *Highway research board proceedings*, volume 1935. National Research Council (USA), Highway Research Board. (Cited on page 257.)

Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129(3):550–560. (Cited on page 27.)

Hamill, T. M. and Colucci, S. J. (1997). Verification of Eta–RSM short-range ensemble forecasts. *Monthly Weather Review*, 125(6):1312–1327. (Cited on page 187.)

Hardy, R. L. (1990). Theory and applications of the multiquadric-biharmonic method 20 years of discovery 1968–1988. *Computers & Mathematics with Applications*, 19(8-9):163–208. (Cited on pages 20 and 88.)

He, X., Chen, X., Xiong, C., Zhu, Z., and Zhang, L. (2016). Optimal time-varying pricing for toll roads under multiple objectives: A simulation-based optimization approach. *Transportation Science*, 51(2):412–426. (Cited on pages 20 and 89.)

Helton, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, 42(2-3):327–367. (Cited on page 24.)

Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17. (Cited on page 132.)

Hoogendoorn, S., Taale, H., Wilmink, I., Van Katwijk, R., Immers, B., and Schuurman, H. (2012). The future of traffic management: State of the art, current trends and perspectives for the future. (Cited on page 28.)

Hugosson, M. B. (2005). Quantifying uncertainties in a national forecasting model. *Transportation Research Part A: Policy and Practice*, 39(6):531–547. (Cited on page 28.)

INFRAS (2017). The handbook of emission factors for road transport. `http://www.hbefa.net/e/index.html`. (Cited on page 22.)

Iooss, B. (2011). Revue sur l'analyse de sensibilité globale de modèles numériques. *Journal de la Société Française de Statistique*, 152(1):3–25. (Cited on page 25.)

Jacobson, L. N., Nihan, N. L., and Bender, J. D. (1990). *Detecting erroneous loop detector data in a freeway traffic management system*. Number 1287. (Cited on page 30.)

Jin, R., Chen, W., and Simpson, T. W. (2001). Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and multidisciplinary optimization*, 23(1):1–13. (Cited on pages 19, 87 and 88.)

Jones, L. K., Gartner, N. H., Shubov, M., Stamatiadis, C., and Einstein, D. (2017). Modeling origin-destination uncertainty using network sensor and survey data and new approaches to robust control. *Transportation Research Part C: Emerging Technologies*. (Cited on page 28.)

Junk, C., Delle Monache, L., and Alessandrini, S. (2015). Analog-based ensemble model output statistics. *Monthly Weather Review*, 143(7):2909–2917. (Cited on page 237.)

Kioutsioukis, I., Tarantola, S., Saltelli, A., and Gatelli, D. (2004). Uncertainty and global sensitivity analysis of road transport emission estimates. *Atmospheric Environment*, 38(38):6609–6620. (Cited on pages 26 and 150.)

Kleijnen, J. P. (2009). Kriging metamodeling in simulation: A review. *European journal of operational research*, 192(3):707–716. (Cited on pages 18, 19, 87, 88 and 92.)

Kouridis, C., Gkatzoflias, D., Kioutsioukis, I., Ntziachristos, L., Pastorello, C., and Dilara, P. (2010). Uncertainty estimates and guidance for road transport emission calculations. *Publications Office of the European Union, EUR*, 24296. (Cited on page 150.)

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139. (Cited on pages 19 and 87.)

Kuwahara, M. and Akamatsu, T. (1993). Dynamic equilibrium assignment with queues for a one-to-many od pattern. *Transportation and Traffic Theory*, 12:185–204. (Cited on pages 14, 43 and 86.)

Lee, L., Carslaw, K., Pringle, K., Mann, G., and Spracklen, D. (2011). Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmospheric Chemistry and Physics*, 11(23):12253–12273. (Cited on pages 21 and 90.)

Leurent, F. (2003). On network assignment and supply-demand equilibrium: An analysis framework and a simple dynamic model. In *European Transport Conference Proceedings*. (Cited on pages 7, 15, 37, 40, 44, 52, 91, 96, 97, 133 and 151.)

Leurent, F. and Aguiléra, V. (2009). An atomic Dijkstra algorithm for dynamic shortest paths in traffic assignment. (Cited on page 44.)

Lophaven, S. N., Nielsen, H. B., and Søndergaard, J. (2002). *DACE: a Matlab kriging toolbox*, volume 2. Citeseer. (Cited on pages 19 and 87.)

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141. (Cited on page 27.)

Lu, S. (2008). Sensitivity of static traffic user equilibria with perturbations in arc cost function and travel demand. *Transportation science*, 42(1):105–123. (Cited on page 24.)

Lu, S. and Nie, Y. M. (2010). Stability of user-equilibrium route flow solutions for the traffic assignment problem. *Transportation Research Part B: Methodological*, 44(4):609–617. (Cited on page 24.)

Ma, W. and Qian, Z. S. (2018). Statistical inference of probabilistic origin-destination demand using day-to-day traffic data. *Transportation Research Part C: Emerging Technologies*, 88:227 – 256. (Cited on page 28.)

Madu, C. N. and Kuei, C.-H. (1994). Regression metamodeling in computer simulation — the state of the art. *Simulation Practice and Theory*, 2(1):27 – 41. (Cited on pages 18 and 87.)

Mahut, M. and Florian, M. (2010). Traffic simulation with dynameq. In *Fundamentals of Traffic Simulation*, pages 323–361. Springer. (Cited on pages 14 and 86.)

Mallet, V., Tilloy, A., Poulet, D., and Brocheton, F. (2013). Reduction and emulation of ADMS Urban. In *15th international conference on harmonisation within atmospheric dispersion modelling for regulatory purposes*, Madrid, Spain. (Cited on pages 19, 21, 88 and 90.)

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266. (Cited on pages 19 and 87.)

Meunier, F. and Wagner, N. (2010). Equilibrium results for dynamic congestion games. *Transportation Science*, 44(4):524–536. (Cited on pages 12, 37, 38, 43 and 52.)

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174. (Cited on page 25.)

Mounce, R. (2006). Convergence in a continuous dynamic queueing model for traffic networks. *Transportation Research Part B: Methodological*, 40(9):779–791. (Cited on pages 12, 14 and 43.)

Mounce, R. (2007). Existence of equilibrium in a continuous dynamic queueing model for traffic networks. In *4th IMA International Conference on Mathematics in Transport*. (Cited on pages 12 and 43.)

Murphy, A. H. (1971). A note on the ranked probability score. *Journal of Applied Meteorology*, 10(1):155–156. (Cited on page 28.)

Ntziachristos, L., Gkatzoflias, D., Kouridis, C., and Samaras, Z. (2009). COPERT: a european road transport emission inventory model. In *Information technologies in environmental engineering*, pages 491–504. Springer. (Cited on page 22.)

Oeuvray, R. and Bierlaire, M. (2009). BOOSTERS: A derivative-free algorithm based on radial basis functions. *International Journal of Modelling and Simulation*, 29(1):26–36. (Cited on pages 19, 20 and 88.)

Orda, A. and Rom, R. (1990). Shortest-path and minimum-delay algorithms in networks with time-dependent edge-length. *Journal of the ACM (JACM)*, 37(3):607–625. (Cited on page 13.)

Osorio, C. and Bierlaire, M. (2013). A simulation-based optimization framework for urban transportation problems. *Operations Research*, 61(6):1333–1345. (Cited on pages 20 and 89.)

Osorio, C. and Chong, L. (2015). A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems. *Transportation Science*, 49(3):623–636. (Cited on pages 20 and 89.)

Osorio, C. and Nanduri, K. (2015). Energy-efficient urban traffic management: a microscopic simulation-based approach. *Transportation Science*, 49(3):637–651. (Cited on pages 20 and 89.)

Patriksson, M. (2004). Sensitivity analysis of traffic equilibria. *Transportation Science*, 38(3):258–281. (Cited on page 24.)

Peeta, S. and Ziliaskopoulos, A. K. (2001). Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1(3):233–265. (Cited on pages 12, 13, 37 and 86.)

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79:214–232. (Cited on page 26.)

Pianosi, F., Sarrazin, F., and Wagener, T. (2015). A Matlab toolbox for global sensitivity analysis. *Environmental Modelling & Software*, 70:80–85. (Cited on page 26.)

Pu, Y., Yang, C., Liu, H., Chen, Z., and Chen, A. (2015). Impact of license plate restriction policy on emission reduction in Hangzhou using a bottom-up approach. *Transportation Research Part D: Transport and Environment*, 34:281–292. (Cited on page 23.)

Punzo, V. and Ciuffo, B. F. (2011). Sensitivity analysis of car-following models: methodology and application. Technical report. (Cited on pages 26 and 27.)

Ran, B. and Boyce, D. (1996). *Modeling dynamic transportation networks: an intelligent transportation system oriented approach*. Springer. (Cited on pages 13 and 86.)

Ran, B., Hall, R. W., and Boyce, D. E. (1996). A link-based variational inequality model for dynamic departure time/route choice. *Transportation Research Part B: Methodological*, 30(1):31–46. (Cited on page 87.)

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer. (Cited on pages 19 and 87.)

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press. (Cited on pages 19, 87, 88 and 92.)

Richards, P. I. (1956). Shock waves on the highway. *Operations Research*, 4(1):42–51. (Cited on pages 14 and 257.)

Roquigny, Q. (2013). L'évaluation socio-économique en période de transition - valeurs du temps. Technical report, Commissaria général à la stratégie et à la prospective - département développement durable. (Cited on page 70.)

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by Kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1):1–55. (Cited on pages 18, 19, 87, 88 and 92.)

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, pages 409–423. (Cited on pages 19, 87 and 92.)

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297. (Cited on pages 27, 131, 132 and 133.)

Saltelli, A., Chan, K., Scott, E. M., et al. (2000). *Sensitivity analysis*, volume 1. Wiley New York. (Cited on pages 26 and 133.)

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons. (Cited on pages 21, 24, 26, 90, 130, 131, 133 and 235.)

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. John Wiley & Sons. (Cited on pages 24 and 130.)

Sarrazin, F., Pianosi, F., and Wagener, T. (2016). Global sensitivity analysis of environmental models: convergence and validation. *Environmental Modelling & Software*, 79:135–152. (Cited on page 26.)

Sarri, A., Guillas, S., and Dias, F. (2012). Statistical emulation of a tsunami model for sensitivity analysis and uncertainty quantification. *arXiv preprint arXiv:1203.6297*. (Cited on pages 21 and 90.)

Shao, H., Lam, W. H., Sumalee, A., Chen, A., and Hazelton, M. L. (2014). Estimation of mean and covariance of peak hour origin–destination demands from day-to-day traffic counts. *Transportation Research Part B: Methodological*, 68:52–75. (Cited on page 28.)

Sheffi, Y. (1985). *Urban Transportation Networks:Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall. (Cited on pages 12 and 86.)

Shorshani, M. F. (2014). *Modélisation de l'impact du trafic routier sur la pollution de l'air et des eaux de ruissellement*. PhD thesis, Université de Paris-Est, École Doctorale Sciences, Ingénierie et Environnement, 6-8 Avenue Blaise-Pascal, Cité Descartes, Champs-sur-Marne, France. (Cited on page 22.)

Shorshani, M. F., André, M., Bonhomme, C., and Seigneur, C. (2015a). Modelling chain for the effect of road traffic on air and water quality: Techniques, current status and future prospects. *Environmental Modelling & Software*, 64:102 – 123. (Cited on pages 15, 22 and 23.)

Shorshani, M. F., André, M., Bonhomme, C., and Seigneur, C. (2015b). Modelling chain for the effect of road traffic on air and water quality: techniques, current status and future prospects. *Environmental Modelling & Software*, 64:102–123. (Cited on pages 22 and 23.)

Shorshani, M. F., Seigneur, C., Rehn, L. P., Chanut, H., Pellan, Y., luc Jaffrezo, J., Charron, A., and André, M. (2015c). Atmospheric dispersion modeling near a roadway under calm meteorological conditions. *Transportation Research Part D: Transport and Environment*, 34:137 – 154. (Cited on page 15.)

Smit, R., Ntziachristos, L., and Boulter, P. (2010). Validation of road vehicle and traffic emission models–A review and meta-analysis. *Atmospheric environment*, 44(25):2943–2953. (Cited on pages 22 and 150.)

Smit, R., Smokers, R., and Rabé, E. (2007). A new modelling approach for road traffic emissions: VERSIT+. *Transportation Research Part D: Transport and Environment*, 12(6):414–422. (Cited on page 22.)

Smith, M. J. (1979). The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B: Methodological*, 13(4):295–304. (Cited on pages 12 and 86.)

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1(4):407–414. (Cited on pages 25, 26 and 130.)

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, 55(1):271–280. (Cited on page 130.)

Song, W., Han, K., Wang, Y., Friesz, T., and del Castillo, E. (2017). Statistical metamodeling of dynamic network loading. *Transportation Research Procedia*, 23:263–282. (Cited on pages 20 and 89.)

Soulhac, L., Salizzoni, P., Mejean, P., Didier, D., and Rios, I. (2012). The model sirane for atmospheric urban pollutant dispersion: Part II, validation of the model on a real case study. *Atmospheric Environment*, 49:320 – 337. (Cited on pages 1 and 23.)

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media. (Cited on pages 19, 87 and 88.)

Talagrand, O. (1999). Evaluation of probabilistic prediction systems. In *Workshop proceedings" Workshop on predictability", 20-22 October 1997, ECMWF, Reading, UK*. (Cited on page 187.)

Tang, J., McNabola, A., Misstear, B., and Caulfield, B. (2017). An evaluation of the impact of the Dublin Port Tunnel and HGV management strategy on air pollution emissions. *Transportation Research Part D: Transport and Environment*, 52:1–14. (Cited on page 23.)

Tarantola, S. and Becker, W. (2016). SIMLAB software for uncertainty and sensitivity analysis. *Handbook of Uncertainty Quantification*, pages 1–21. (Cited on page 26.)

Theory, N. R. C. U. T. R. B. T. F. and Committee, C. (2011). *75 Years of the Fundamental Diagram for Traffic Flow Theory: Greenshields Symposium : July 8-10, 2008, Woods Hole, Massachusetts*. Transportation research circular. Transportation Research Board. (Cited on page 257.)

Thorey, J. (2017). *Prévision d'ensemble par agrégation séquentielle appliquée à la prévision de production d'énergie photovoltïque*. PhD thesis, Université Pierre et Marie Curie, École Doctorale Sciences, Mathématiques de Paris Centre, Place Jussieu, Paris, France. (Cited on page 237.)

Thouron, L., Seigneur, C., Kim, Y., Mahé, F., André, M., Bruge, B., Chanut, H., and Pellan, Y. (2018). Intercomparison of three modeling approaches for traffic-related road dust resuspension using two experimental data. *Transportation Research Part D: Transport and Environment*, 58:108–121. (Cited on pages 15 and 22.)

Tobin, R. L. and Friesz, T. L. (1988). Sensitivity analysis for equilibrium network flow. *Transportation Science*, 22(4):242–250. (Cited on page 24.)

TRB, T. R. B. (2000). *Highway Capacity Manual*. US Transportation Research Board, Washington DC. (Cited on page 195.)

Treiber, M., Kesting, A., and Thiemann, C. (2012). *Traffic Flow Dynamics: Data, Models and Simulation*. Springer Berlin Heidelberg. (Cited on page 16.)

Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, 59(2):251–260. (Cited on pages 43, 52 and 74.)

Wagner, N. (2012). *Dynamic user equilibrium on a transport network: mathematical properties and economical applications*. PhD thesis, Université de Paris-Est, École Doctorale Ville Environment Territoire, 6-8 Avenue Blaise-Pascal, Cité Descartes, Champs-sur-Marne, France. (Cited on pages 12, 37, 40, 52 and 96.)

Wardrop, J. G. (1952). Road paper. Some theoretical aspects of road traffic research. In *ICE Proceedings: engineering divisions*, volume 1, pages 325–362. Thomas Telford. (Cited on pages 3, 11, 52, 86 and 133.)

Whitham, G. (1955). On kinematic waves II. A theory of traffic flow on long crowded roads. In *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, volume 229, pages 317–345. The Royal Society. (Cited on pages 14 and 257.)

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press. (Cited on page 27.)

Yang, H. (1997). Sensitivity analysis for the elastic-demand network equilibrium problem with applications. *Transportation Research Part B: Methodological*, 31(1):55–70. (Cited on page 24.)

Yang, H., Sasaki, T., Iida, Y., and Asakura, Y. (1992). Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research Part B: Methodological*, 26(6):417–434. (Cited on page 28.)

Zallinger, M., Tate, J., and Hausberger, S. (2008). An instantaneous emission model for the passenger car fleet. In *16th International Transport and Air Pollution CongressTechnical University Graz*. (Cited on page 22.)

Zhang, C., Osorio, C., and Flötteröd, G. (2017). Efficient calibration techniques for large-scale traffic simulators. *Transportation Research Part B: Methodological*, 97:214–239. (Cited on pages 20 and 89.)

Zhou, X. and Mahmassani, H. S. (2006). Dynamic origin-destination demand estimation using automatic vehicle identification data. *IEEE Transactions on intelligent transportation systems*, 7(1):105–114. (Cited on page 28.)

Zhu, D. and Marcotte, P. (2000). On the existence of solutions to the dynamic user equilibrium problem. *Transportation Science*, 34(4):402–414. (Cited on page 12.)

Zhu, S. and Ferreira, L. (2013). Quantifying errors in micro-scale emissions models using a case-study approach. *Transportation Research Part D: Transport and Environment*, 21:19–25. (Cited on page 29.)

Appendices

Appendix A

Hydrodynamic model and fundamental diagram

Inspired from fluid mechanics, hydrodynamic models are used to model the traffic flow and congestion phenomenon on a road. The most widely used hydrodynamic model is the LWR model, developed by Whitham [1955] and Richards [1956]. The LWR model gives a relation between the three most important variables of an homogeneous flow of vehicles: the flow rate (in $\mathrm{veh\,h^{-1}}$), traffic density (in $\mathrm{veh\,km^{-1}}$) and average speed (in $\mathrm{km\,h^{-1}}$).

x is the spatial dimension of traffic flow while t is the temporal dimension. Let $q(x,t)$, $k(x,t)$ and $v(x,t)$ denote the traffic flow, density and velocity of a flux of traffic. Based on the homogeneity assumption and hydrodynamic in fluid mechanics, we have the conservation law and flow-density-velocity relation:

$$\begin{cases} \frac{\partial q(x,t)}{\partial x} + \frac{\partial k(x,t)}{\partial t} = 0 \\ q(x,t) = k(x,t)v(x,t). \end{cases} \tag{A.1}$$

The assumption in LWR model is that the velocity of traffic flow only depends on the traffic density on the road ($v(x,t) = f_{vk}(k(x,t))$). The related *fundamental diagram* describes thus the relation between traffic flow and density:

$$q(x,t) = k(x,t)v(x,t) = k(x,t)f_{vk}(k(x,t)) = f_{kq}(k(x,t)). \tag{A.2}$$

Therefore, the final LWR model can be represented by the partial derivation equation

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial f_{kq}(k(x,t))}{\partial x} = 0. \tag{A.3}$$

There have been various formulations to model the *fundamental diagram* presented in Equation A.2 since its first formulation from experimental observations by Greenshields et al. [1935]. Detailed reviews can be found in Theory and Committee [2011]. The modeling of the *fundamental diagram* depends on properties of roads (width of the lanes), composition of traffic flow (passenger cars, trucks, buses) and external conditions (weather...). It can be calibrated using measured traffic flow data. Here we use an

example to illustrate a fundamental diagram and the fundamental values we can get from it.

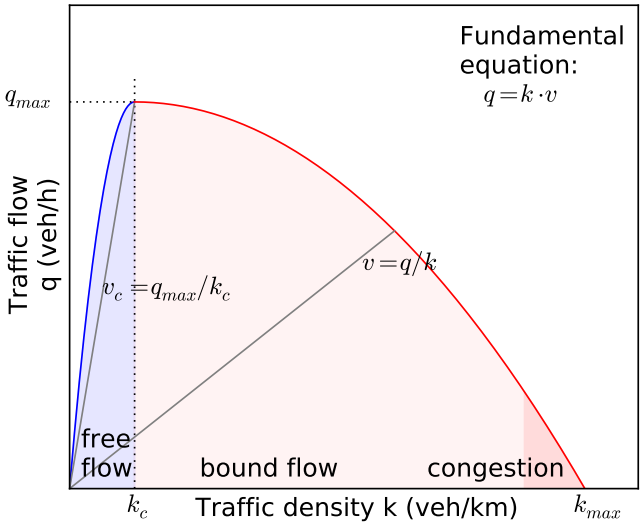


Figure A.1: Illustration of fundamental diagram. Density (k) defines the number of vehicles passing through a given road with fixed length. It is expressed in veh km^{-1}. Traffic flow q is the number of vehicles passing a point during an unit of time. It is in veh h^{-1}. With the assumption that vehicles are homogeneous, the speed V is the average speed of a flux of traffic. k_c is the critical density of traffic. When $k < k_c$, vehicles travel with free-flow speed. When $k = k_c$, the traffic flows reaches its maximum value q_{max}. It defines the maximum traffic flow that can pass through a road: the capacity of a road.

The traffic density is derived from the number of vehicles (n) on a given distance (L): $k = \frac{n}{L}$. For security, there is a minimum distance between each vehicle in a traffic flow and let d_c denote this *critical* distance. If the inter-distance between two vehicles is bigger than d_c, vehicles can travel with free-flow speed. In this case, the traffic flow increases with the traffic density. However, when the distance between two vehicles is less than d_c, vehicles should decrease their speed. Finally, if the density continues to increase and reaches the maximum value, the congestion happens and the speed of the traffic flow is zero. Therefore, the *fundamental diagram* gives information for maximum traffic flow (q_{max}) and this is often the definition of *capacity* of a road. Before the critical density k_c, the flowing situation on the road is *free-flow* condition. When $k > k_c$, the traffic flow is bounded.

Nomenclature

Greek Symbols

α_{day}	ratio between (i) the total traffic demand during the whole day and (ii) the traffic demand during evening peak hour on the same *day*	–
α_{VDF}	coefficients to be adjusted in VDF	–
β	coefficients to be adjusted in VDF	–
$\beta_k^{temporal}$	temporally-averaged error of the simulated flow at detector k	$\mathrm{veh\,h^{-1}}$
$\beta_t^{spatial}$	spatially-averaged error of the simulated flow at time step t	$\mathrm{veh\,h^{-1}}$
$\mathbf{\Psi}$	reduced basis obtained from PCA	–
δ_{0-5}	evening peak coefficient multiplied with the traffic demand between O-D pair between which the distance is from 0 to 5 km	–
δ_0	evening peak coefficient multiplied with the traffic demand between O-D pair in the same zone	–
δ_{10-15}	evening peak coefficient multiplied with the traffic demand between O-D pair between which the distance is from 10 to 15 km	–
δ_{5-10}	evening peak coefficient multiplied with the traffic demand between O-D pair between which the distance is from 5 to 10 km	–
$\delta_{>15}$	evening peak coefficient multiplied with the traffic demand between O-D pair between which the distance is larger than 15 km	–
$\varepsilon_{\mathrm{diesel\_2.0}}$	proportion of diesel PCs with engine capacity less than 2.0 L among all diesel PCs	–
η	transposition coefficient for input O-D matrix	–
$\gamma_{\mathrm{gaso\_1.4}}$	share of gasoline PCs with engine capacity less than 1.4 L among all gasoline PCs	–
$\lambda_{big,small}$	coefficient multiplied with default link capacities for links in big capacity category and small capacity category, respectively	–

$\lambda_{high,low}$	coefficient multiplied with default link speed limits for links in high speed category and low speed category, respectively	–
ϕ	chosen RBF function	–
Ψ_j	the j^{th} component of the reduced basis	–
$\rho(h)$	temporal variation ratio between (i) the total cumulated traffic from h_0 till h and (ii) the total cumulated traffic during the whole simulation period of one day	–
σ_{HDV}	ratio between the number of HDVs and PCs	–
$\tau_{\text{diesel},1.4-2.0l}$	proportion of diesel passenger cars with engine capacity from 1.4 L to 2.0 L	–
$\tau_{\text{diesel},<1.4l}$	proportion of diesel passenger cars with engine capacity less than 1.4 L	–
$\tau_{\text{diesel},>2.0l}$	proportion of diesel passenger cars with engine capacity bigger than 2.0 L	–
τ_{euro1}	proportion of PCs with emission standard *EURO 1*	–
τ_{euro2}	proportion of PCs with emission standard *EURO 2*	–
τ_{euro3}	proportion of PCs with emission standard *EURO 3*	–
τ_{euro4}	proportion of PCs with emission standard *EURO 4*	–
τ_{euro5}	proportion of PCs with emission standard *EURO 5*	–
$\tau_{\text{gaso},1.4-2.0l}$	proportion of gasoline passenger cars with engine capacity from 1.4 L to 2.0 L	–
$\tau_{\text{gaso},<1.4l}$	proportion of gasoline passenger cars with engine capacity less than 1.4 L	–
$\tau_{\text{gaso},>2.0l}$	proportion of gasoline passenger cars with engine capacity bigger than 2.0 L	–
$\tau_{\text{pre-euro}}$	proportion of PCs with emission standard *pre-EURO*	–
θ_{deisel}	proportion of diesel PCs	–
θ_{gaso}	proportion of gasoline PCs	–
$\varphi_{\text{diesel\_euro4}}$	proportion of diesel PCs with emission standard of Euro 4 and higher	–
ξ_{day}	total demand coefficient for the simulation *day* divided by the total demand on the reference day	–

ξ_d	total demand coefficient for adjusting the total traffic demand for the day d with respect to the reference day	–
$\zeta_{\text{gaso\_euro4}}$	proportion of gasoline PCs with emission standard of Euro 4 and higher	–

Roman Symbols

(o, d)	an *O-D pair*	–
$(q_{o,d}^{\text{peak}})$	average traffic demand during the evening peak from o to d	$\text{veh}\,\text{h}^{-1}$
$[a_i, b_i]$	uniform distribution for sampling p_i in the Monte Carlo simulation	–
$\bar{\mathcal{H}}$	interval outside $\mathcal{H}_{\text{atomic}}$	–
$\bar{o}_h$	spatially-averaged traffic flow over all detectors during $[h, h + \Delta h]$	$\text{veh}\,\text{h}^{-1}$
$\boldsymbol{C}^{big}$	vector of big link capacities	–
$\boldsymbol{C}^{small}$	vector of small link capacities	–
$\boldsymbol{K}$	vector of link capacity	–
$\boldsymbol{p}$	input vector of $\mathcal{M}$	–
$\boldsymbol{p}_{\text{ref}}$	input vector of $\mathcal{M}$ on the reference day	–
$\boldsymbol{Q}^{\text{peak}}$	O-D matrix of the agglomeration of Clermont-Ferrand during the evening peak hour 17:00 – 18:00	–
$\boldsymbol{R}(h)$	vector of minimum cost routes	–
$\boldsymbol{T}(h)$	vector of link travel times	–
$\boldsymbol{T}_0$	vector of free-flow travel time	–
$\boldsymbol{U}(h)$	vector of generalized travel cost	–
$\boldsymbol{V}(h_{\text{simu}})$	average speed during $[h_{\text{simu}}, h_{\text{simu}} + \Delta h]$ computed by DTA model	–
$\boldsymbol{V}_0^{high}$	vector of high speed limits	–
$\boldsymbol{V}_0^{low}$	the vector of low speed limits ($V_0 \leq 50\,\text{km}\,\text{h}^{-1}$)	–
$\boldsymbol{V}_0$	vector of speed limit	–
$\boldsymbol{X}^r$	cumulated traffic flow on route r	–
$\boldsymbol{Y}$	vector of link cumulated flow	–

$\boldsymbol{y}$	output vector of the original model	–
Δh	time step in DTA simulation	h
$\mathbb{R}$	set of all real numbers	–
$\mathbb{R}^D$	real coordinate space of D dimensions, the space of outputs in this dissertation	–
$\mathbb{R}^K$	real coordinate space of K dimensions, the space of inputs in this dissertation	–
$\mathbb{R}_+$	set of all positive real numbers	–
$\widehat{\mathcal{M}}(\boldsymbol{p})$	metamodel of $\mathcal{M}(\boldsymbol{p})$	–
$\mathcal{A}$	set of links	–
$\mathcal{A}_{big}$	set of links with $C_a > 900\,\mathrm{veh\,h^{-1}}$	–
$\mathcal{A}_{high}$	set of links with $V_0 > 50\,\mathrm{km\,h^{-1}}$	–
$\mathcal{A}_{low}$	set of links with $V_0 \leq 50\,\mathrm{km\,h^{-1}}$	–
$\mathcal{A}_{small}$	set of links with $C_a \leq 900\,\mathrm{veh\,h^{-1}}$	–
$\mathcal{C}(\mathbb{R}_+)$	set of all continuous maps from $\mathbb{R}_+$ to $\mathbb{R}$	–
$\mathcal{G}(\mathcal{N}, \mathcal{A})$	oriented graph to model the road network	–
$\mathcal{H}$	bounded time interval for carrying out DTA simulation	–
$\mathcal{H}_{\text{atomic}}$	simulation period for the LCF model	–
$\mathcal{I}_d(h)$	set of indices of links on the minimum cost route to the destination d for users departing from origin at h	–
$\mathcal{K}^{last}$	set of locations of detector where the simulation ensemble underestimates the traffic flow	–
$\mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$	set of positive measurable functions on $\mathcal{H}$	–
$\mathcal{M}(\boldsymbol{p})$	original model	–
$\mathcal{M}(\mathbb{R})$	the set of measures on the set $\mathbb{R}$ with $\mathcal{M}(]-\infty, h]) = \int_{-\infty}^{h} m(\tilde{h})d\tilde{h}$	–
$\mathcal{N}$	set of nodes	–
$\mathcal{Q}(h)$	dynamic O-D matrix	–
$\mathcal{S}$	user strategy set during $\mathcal{H}$	–

$\mathcal{T}^{last}$	set of time intervals when the simulation ensemble underestimates the traffic flow	–
$\mathcal{W}^{peak}$	spatially-averaged traffic flow in Clermont-Ferrand during $17:00-18:00$ on day d	$\mathrm{veh\,h^{-1}}$
$\mathcal{W}_d(h)$	spatially-averaged traffic flow in Clermont-Ferrand during $[h, h+\Delta h]$ on day d	$\mathrm{veh\,h^{-1}}$
$\mathcal{W}_{ref}(h)$	spatially-averaged traffic flow in Clermont-Ferrand during $[h, h+\Delta h]$ on the reference day	$\mathrm{veh\,h^{-1}}$
$\mathcal{W}_{ref}^{peak}$	spatially-averaged traffic flow in Clermont-Ferrand during $17:00-18:00$ on the reference day	$\mathrm{veh\,h^{-1}}$
$\mathcal{W}_{type}(h)$	spatially-averaged traffic flow in Clermont-Ferrand during $[h, h+\Delta h]$ on the typical weekday *type*	$\mathrm{veh\,h^{-1}}$
$\mathcal{Z}_D$	set of nodes for Destination zone	–
$\mathcal{Z}_O$	set of nodes for Origin zone	–
$\widehat{\boldsymbol{y}}$	output vector of the metamodel of $\mathcal{M}(\boldsymbol{p})$	–
$\widehat{f}(\boldsymbol{p})$	emulator of f	–
C	(in VDF) link capacity	$\mathrm{veh\,h^{-1}}$
C_i	lane capacity at observation point i	$\mathrm{veh\,h^{-1}\,lane^{-1}}$
$C_{a,general}$	generalized cost on link a	h
D	total number of links, $D = card(\mathcal{A})$	–
$d(\cdot,\cdot)$	Euclidean distance	–
$E_{hot,i,j,a}$	hot emission on the road a for pollutant i, for vehicle of technology j	g
$e_{hot,i,j,a}$	HEF on the road a for pollutant i, for vehicle of technology j	$\mathrm{g\,km^{-1}}$
e_{hot}	hot emission factor	$\mathrm{g\,km^{-1}}$
$e_{k,t}$	error of the simulated flow at detector k and at time t with respect to observed traffic flow	$\mathrm{veh\,h^{-1}}$
$f(\boldsymbol{p})$	original function f	–
h	time when users enter the road network	h
h_0	beginning time of the simulation	h

$H_a(Y_a)(h)$	link exit time function	–
$H_a(Y_a)^{-1}(\mathcal{H})$	all the instants at which link a can be entered in order to leave it at some instant in $\mathcal{H}$	–
h_{simu}	the instant at which we want to compute traffic flow, travel speed and time	h
I_{atomic}	set of discrete time instants during 3.25 h for the LCF model	–
I_{CI90}	number of observations whose value is in $CI90$ predicted by the ensemble	–
I_{CI90}^{flow}	$5^{th} - 95^{th}$ percentile range of traffic flows over the whole network of Clermont-Ferrand	–
I_{CI90}^{NOx}	$5^{th} - 95^{th}$ percentile range of NOx emissions from passenger cars over the whole network of Clermont-Ferrand	–
I_{day}	set of discrete time instants on the *day*	–
I_{IQR}	number of observations whose value is in IQR predicted by the ensemble	–
I_{IQR}^{flow}	interquartile range of traffic flows over the whole network of Clermont-Ferrand	–
I_{IQR}^{NOx}	interquartile range of NOx emissions from passenger cars over the whole network of Clermont-Ferrand	–
k	(in LWR model) traffic density: number of vehicles per unit of distance	$\text{veh}\,\text{km}^{-1}$
k_c	(in LWR model) critical traffic density separating *free-flow* and *congested* condition of the fundamental diagram	$\text{veh}\,\text{km}^{-1}$
$K_a(h)$	capacity on link a	$\text{veh}\,\text{h}^{-1}$
$L(h)$	number of users queuing at h in the bottleneck	veh
L_a	length of link a	km
M	sample size for conducting metamodeling	–
M_{ob}	total number of available observation points	–
N_{det}	total number of loop detectors	–
$N_{j,a}$	number of vehicles on the road a	veh
N_{link}	total number of links in the modeled network	–
N_{type}	total number of days of the same *type* from September 2013 to August 2014	–

$O^{total}_{(17,18)}$	total observed traffic of all detectors during 17:00 - 18:00	veh
$O^{total}_{(h_1,h)}$	total observed traffic of all detectors during $[h_1, h]$	veh
$O_{k,(h_0,h)}$	number of cumulated vehicles passing through detector k from h_0 to h	veh
$O_{k,(h_1,h)}$	number of cumulated vehicles passing through detector k from h_1 to h	veh
$o_{k,h}$	average traffic flow measured at detector k during $[h, h+\Delta h]$	veh h^{-1}
$o_{k,t}$	traffic flow observation at detector k and at time t	veh h^{-1}
$P(h)$	temporal variation coefficient	–
$P^*(h_i)$	temporal variation coefficient for the deterministic simulation at time h_i	–
P_a	toll price of link a	€
p_i	the i^{th} component of input $\boldsymbol{p}$	–
p_i^*	value of input p_i in the deterministic simulation	–
Q_1	value of the 1st quartile of a sequence of numbers	–
Q_3	value of the 3rd quartile of a sequence of numbers	–
$Q_{(o,d)}(h)$	cumulated traffic demand from o to d from h_0 until h	veh
$q_{(o,d)}(h)$	traffic demand from o to d at instant h	veh h^{-1}
$q^{17-18}_{(o,d)}$	hourly average traffic flow during 17:00 - 18:00 for each O-D pair	veh h^{-1}
$q^{\mathcal{H}}_{(o,d)}(h)$	density of traffic demand from o to d during $\mathcal{H}$	veh h^{-1}
q_{max}	(in LWR model) the maximum traffic flow that can pass through a road: the capacity of a road	veh h^{-1}
$R_{(o,d)}$	route from the original node o to the destination node d	–
$R_{(o,d)}(h)$	minimum cost route from o to d when a user departing from o at the instant h	–
S_i	first-order Sobol' sensitivity index with respect to the i^{th} component of input $\boldsymbol{p}$	–
$s_{j,i}$	computed traffic flow on link i from the j^{th} simulation from the Monte Carlo approach	veh h^{-1}

$s_{j,k,t}$	traffic flow computed at detector k and at time t by the j^{th} simulation from the Monte Carlo approach	veh h^{-1}
S_{Ti}	total-effect Sobol' sensitivity index with respect to the i^{th} component of input $\boldsymbol{p}$	–
T	total number of time steps during the simulation period	–
t	link travel time	s
t_0	free-flow link travel time	s
T_{0a}	free-flow travel time on link a	h
$T_a(h)$	computed travel time on link a	h
$t_a(Y_a)(h)$	link travel time functions for the network $\mathcal{G}(\mathcal{N}, \mathcal{A})$	–
$U_a(h)$	generalized travel cost on link a	h
V	(in VDF) traffic volume on a link	veh h^{-1}
v	vehicle average speed	km h^{-1}
V/C	volume-to-capacity ratio on a link or on a road	–
v_t	value of time	€/h
V_{0a}	speed limit on link a	km h^{-1}
$v_{a,h}$	average speed at link a during $[h, h + \Delta h]$	km h^{-1}
X	distribution of users strategy in DUE problem, $X \in \mathcal{M}(\mathcal{S})$	veh
$x^r(h)$	density of X^r	veh h^{-1}
$y_a(h)$	traffic flow at h, $y(h) \in \mathcal{L}^1(\mathcal{H}, \mathbb{R}_+)$	veh h^{-1}
$y_{a,h}$	computed traffic flow during $[h, h + \Delta h]$	veh h^{-1}
Y_a	cumulated flow on link a	veh
$z \times z$	dimension of the O-D matrix	–
z	number of zones in the simulation area	–
X^r	cumulated traffic flow on route r	veh

Acronyms

BC	black carbon

CI	confidence interval
CI90	90 % confidence interval: the interval between the 5th and 95th percentiles of a sequence of value predicted by the ensemble of simulations
DNL	dynamic network loading
DNLP	dynamic network loading problem
DOP	dynamic optimization problem
DTA	dynamic traffic assignment
DUE	dynamic user-equilibrium
EEA	European Environment Agency
EF	emission factor
FIFO	first-in-first-out
GPS	global positioning system
GSA	global sensitivity analysis
HDV	heavy duty vehicle
HEF	hot emission factor
IQR	interquartile range
LADTA	Lumped Analytical DTA
LCF	atomic simulation of LADTA applied to Clermont-Ferrand
LHS	Latin Hypercube Sampling
LOS	level of service of a road network
LSA	local sensitivity analysis
LTK	LADTA Tool Kit
Meta-LCF	metamodel of the original LCF model
MNBE	mean normalized bias error
NRMSE	normalized root mean square error
O-D	Origin-Destination
OAT	one at a time

PC	passenger car
PCA	principal component analysis
pdf	probability density function
PM	particulate matter
RBF	radial basis function
RC	route choice
RMSE	root mean square error
RPS	ranked probability score
RSD	relative standard deviation
SA	sensitivity analysis
TA	traffic assignment
TF	traffic flowing
UE	user equilibrium
VBSA	variance-based sensitivity analysis
VDF	volume-delay function
VI	variational inequality
VL	volume loading