



HAL
open science

Modélisation pour le suivi et l'amélioration des performances énergétiques des quartiers

Victor Marty-Jourjon

► **To cite this version:**

Victor Marty-Jourjon. Modélisation pour le suivi et l'amélioration des performances énergétiques des quartiers. Energie électrique. Université Paris sciences et lettres, 2021. Français. NNT : 2021UP-SLM061 . tel-03666882

HAL Id: tel-03666882

<https://pastel.hal.science/tel-03666882>

Submitted on 12 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'école nationale supérieure des mines de Paris

Modélisation pour le suivi et l'amélioration des performances énergétiques des quartiers

Modeling for monitoring and improving the energy performances of districts

Soutenue par

Victor Marty-Jourjon

Le : 09/09/2021

Ecole doctorale n° 621

**Ingénierie des Systèmes,
Matériaux, Mécanique,
Energétique**

Spécialité

**Energétique et génie des
procédés**



Composition du jury :

Stéphane, GINESTET Professeur, INSA Toulouse	<i>Rapporteur-Président</i>
Mindjid, MAIZIA Professeur, Université de Tours	<i>Rapporteur</i>
Sihem, GUERNOUTI Docteur, CEREMA	<i>Examineur</i>
Simon, ROUCHIER Docteur, Université Savoie Mont Blanc	<i>Examineur</i>
Thomas, BERTHOU Docteur, Mines ParisTech	<i>Examineur</i>
Pascal, STABAT Professeur, Mines ParisTech	<i>Directeur de thèse</i>

Remerciements

Je remercie d'abord l'ensemble des membres du jury pour avoir accepté d'évaluer ma thèse de doctorat. Les différents points de vue et nombreuses remarques soulevées m'ont permis d'enrichir ma réflexion sur ces travaux.

Je voudrais ensuite particulièrement remercier mon directeur de thèse Pascal Stabat : par ton excellence scientifique, tu as su répondre avec précision et pertinence à mes nombreuses questions. Tu as également su rester à l'écoute dans les moments difficiles, compris mes inquiétudes, supporter et accompagner mes choix scientifiques. Bravo et merci pour tout ça.

Je voudrais ensuite m'adresser à Thomas Berthou qui est à l'origine de cette thèse : merci de m'avoir fait confiance d'abord en tant qu'ingénieur de recherche puis en tant que doctorant. Merci aussi de m'avoir fortement soutenu sur les différents livrables pour lesquels nous avons travaillé ensemble pendant ces années au sein de l'équipe MDE « Maîtrise de la Demande en Énergie », et merci d'avoir apporté un regard différent sur cette thèse.

Plus généralement merci aussi à toute l'équipe MDE, avec qui j'ai passé tant de bons moments et appris tellement de choses. Je pense notamment à Bruno Duplessis avec qui j'ai été heureux de collaborer sur des projets de recherche et d'enseignement, et à toute l'équipe de doctorants et ex-doctorants, qui en plus de m'avoir appris beaucoup de choses, ont rendu cette aventure plus douce et plus joyeuse : Damien, Romain, Valentin D, Baptiste, Enora, Lucie, Derek, Thomas C, Tam, Valentin M, Rémi, avec une pensée particulière à Antoine Fabre avec qui j'ai partagé mon bureau toutes ces années.

J'ai aussi une pensée pour Baptiste Portelli, mon professeur de physique en classe préparatoire au lycée Etienne Mimard, qui par ses excellents cours et son engagement envers ses élèves, a pris une place importante dans mon parcours jusqu'ici.

Merci à mes amis et ma famille qui m'ont encouragé et félicité le jour de la soutenance. Merci à Samuel pour ses relectures. Merci à mes sœurs, Louli-Betel et Négélina qui ont été pour moi une grande source motivation. Merci à mes parents pour m'avoir permis de grandir dans un environnement heureux qui m'a laissé le temps et la sérénité nécessaire pour me passionner pour les sciences. Et pour finir je m'adresse à Anchal : tu étais en première ligne dans les moments les plus difficiles, tu m'as apporté ton soutien, non seulement moral, mais aussi scientifique, et je ne serais sans doute pas arrivé au bout de cette thèse sans toi, mille mercis.

Table des Matières

Table des Matières.....	5
Nomenclature.....	10
Chapitre 1 : Contexte, enjeux et objectifs.....	13
1.1 Besoin de suivi et d'amélioration des performances énergétiques à l'échelle du quartier.....	15
1.1.1 Les enjeux liés à l'énergie et au bâtiment	15
1.1.2 Les actions possibles dans le secteur du bâtiment.....	15
1.1.2.1 Rénovation	15
1.1.2.2 Flexibilité	16
1.1.2.3 Sobriété	17
1.1.3 Le quartier : une nouvelle échelle d'étude des actions de performance énergétique.....	17
1.1.4 Un modèle pour le suivi et l'amélioration des performances énergétiques à l'échelle du quartier	19
1.2 Données et sources d'informations disponibles à l'échelle du quartier	20
1.2.1 Les données d'exploitation.....	21
1.2.2 Les données de description du quartier	21
1.3 Choix de l'approche de modélisation	23
1.3.1 Modélisation bottom-up ou top-down ?	23
1.3.2 Modèles de signature énergétique.....	24
1.3.3 Modèles de simulation thermique dynamique (STD)	25
1.3.4 Modèles dynamiques simplifiés (RC).....	26
1.3.5 Les principaux logiciels de simulation dynamique simplifiée à l'échelle du quartier	27
1.4 Problématique de la thèse.....	29

Chapitre 2 : Calibration par méthode Bayésienne	33
2.1 Introduction.....	35
2.2 L'approche bayésienne.....	36
2.3 Choix des probabilités a priori.....	37
2.4 Réduction du nombre de paramètres calibrés	38
2.5 Inférence bayésienne	39
2.5.1 Choix de la fonction de vraisemblance.....	39
2.5.1.1 Hypothèse des erreurs indépendantes et identiquement distribuées (i.i.d).....	39
2.5.1.2 Prise en compte des corrélations entre les erreurs et les entrées observables.....	40
2.5.1.3 Proposition d'une formulation statistique originale pour la prise en compte des corrélations temporelles.....	41
2.5.1.4 L'hypothèse du filtre de Kalman.....	42
2.5.2 Calcul de la probabilité a posteriori.....	45
2.5.3 Les moteurs d'inférence markoviens	47
2.5.3.1 Métropolis-Hastings (MH).....	48
2.5.3.2 «Slice sampling»	48
2.5.3.3 Méthodes Hamiltoniennes.....	49
2.6 Vérification de la convergence	50
2.7 Calcul de la prédiction a posteriori.....	51
2.8 Utilisation d'un métamodèle.....	52
2.9 Conclusion	53
Chapitre 3 : Sélection d'un modèle simplifié de type « RC »	55
3.1 Introduction.....	58
3.2 Les méthodes d'évaluation d'un modèle	58
3.2.1 Indicateurs d'écart	58
3.2.2 Données de validation	59
3.2.3 Validation croisée.....	60

3.2.4	Critères d'information	60
3.2.5	Analyse des résidus	61
3.2.6	Identifiabilité.....	61
3.3	Etat de l'art des modèles RC.....	62
3.3.1	Les études d'identifiabilité	64
3.3.2	Les études concernant le modèle ISO (R5C1)	66
3.4	Présentation de la méthode de construction d'un modèle RC identifiable	68
3.5	Description du modèle ISO (R5C1)	69
3.5.1	La méthode horaire	70
3.5.2	Le calcul de C_m et A_m	72
3.5.3	Formulation mathématique et informatique du modèle	73
3.5.4	Calcul des flux thermiques	75
3.5.5	Prise en compte de la ventilation.....	76
3.6	Description du cas d'étude	76
3.7	Etude du modèle ISO sur un cas simplifié mono-zone	78
3.7.1	Estimation physique des paramètres du modèle ISO	78
3.7.2	Comparaison du modèle ISO avec le logiciel TRNSYS et ajout d'une seconde capacité C_a	80
3.7.3	Première calibration du modèle et réduction du nombre de paramètres calibrés.....	82
3.7.4	Re-paramétrage du modèle ISO_ C_a	85
3.8	Etude du modèle ISO_ C_a sur le bâtiment complet multi-zones.....	92
3.9	Etude d'une variante du modèle ISO_ C_a , le modèle ISO_ $C_p C_a$	98
3.10	Conclusion	102
Chapitre 4 : Paramétrage a priori d'un modèle de quartier : cas du quartier de Clichy-Batignolles		
4.1	Introduction.....	108
4.2	Construction d'une table SIG	109

4.2.1	Utilisation de la base de données BD_TOPO®	110
4.2.2	Corrections manuelles de la table SIG	111
4.2.3	Calcul des surfaces de déperdition	113
4.2.4	Calculs des apports solaires.....	114
4.3	Identification et positionnement des locaux associés à des surfaces chauffées et des usages	116
4.3.1	Les différentes définitions de surfaces.....	117
4.3.2	Utilisation de la base de données MAJIC (zone ENS)	119
4.3.3	Utilisation des documents de construction (zone Ouest).....	122
4.3.4	Lien entre les locaux et la table SIG.....	123
4.4	Enrichissement statistique : application aux bâtiments de bureaux.....	124
4.4.1	Classification des bâtiments de bureaux par types.....	125
4.4.2	Classe d'inertie	126
4.4.3	Coefficients de déperditions thermiques	126
4.4.4	Infiltration	127
4.4.5	Profils d'occupation	129
4.4.6	Apports thermiques métaboliques liés aux occupants	133
4.4.7	Apports thermiques liés aux consommations d'électricité spécifique.	134
4.4.8	Ventilation mécanique	134
4.4.9	Scénario de stores extérieurs (apports solaires).....	134
4.4.10	Scénario de températures de consigne	135
4.5	Exemples de modèle a priori sur le quartier de Clichy-Batignolles	136
4.6	Conclusion	138
Chapitre 5 : Application de la méthode de calibration bayésienne.....		141
5.1	Introduction.....	143
5.2	Discussion sur le choix de la fonction de vraisemblance.....	143
5.2.1	Fonction de vraisemblance <i>Li. i. d</i>	144
5.2.2	Fonction de vraisemblance <i>LGP</i>	147

5.2.3	Fonction de vraisemblance <i>LKal</i>	149
5.2.4	Conclusion sur le choix de la fonction de vraisemblance	153
5.3	Discussion sur le nombre de capteurs de température	153
5.4	Identification des capacités et des limites de la méthode de calibration de modèles RC selon les données disponibles.....	156
5.4.1	Estimation des incertitudes a priori	158
5.4.1.1	Scénario d'utilisation des stores extérieurs (apports solaires).....	158
5.4.1.2	Scénario de ventilation mécanique.....	159
5.4.1.3	Scénario d'occupation (apports métaboliques)	160
5.4.1.4	Scénario de températures de consigne	161
5.4.1.5	Scénario de consommation d'électricité spécifique.....	161
5.4.1.6	Résumé des hypothèses de modélisation et des distributions a priori.....	162
5.4.2	Analyse des modèles a posteriori.....	163
5.4.3	Identification de dérives de consommation.....	166
5.4.4	Identification des causes des dérives de consommation	169
5.5	Conclusion	172
	Conclusion générale et perspectives	173
	Table des figures	178
	Liste des tableaux	183
	Bibliographie.....	185
	Annexe 1 : Inventaire des bases de données territoriales	192
	Annexe 2 : Méthodes de calcul de Cm selon (AFNOR, 2010).....	196
1.	Méthode détaillée du calcul de Cm.....	196
2.	Méthode simplifiée du calcul de Cm	199

Nomenclature

Sigles

SED	Simulation Energétique Dynamique
STD	Simulation Thermique Dynamique
SCR	Système de Coordonnées de Référence
SIG	Système d'Information Géographique

Éléments mathématiques

k, i, j	Index
$ A $	Déterminant de A
0_k	Vecteur de zéros $\in \mathbb{R}^k$
I_k	Matrice identité $\in \mathbb{R}^{k \times k}$
D	Vecteur de données mesurées
D_k	Kième composante du vecteur D
η	Vecteur de données simulées, parfois noté $\eta(\theta)$
n	Nombre de données mesurées ou simulées
x	Vecteur des paramètres observables
p	Nombre de paramètres observables
θ	Vecteur des paramètres non observables calibrés
q	Nombre de paramètres non observables calibrés
θ_{ML}	Vecteur des paramètres non observables correspondant au maximum de vraisemblance
θ^s	Vecteur des paramètres non observables correspondant à un échantillon s
Ω	Ensemble des vecteurs de paramètres θ possibles
δ	Vecteur aléatoire caractérisant l'inadéquation du modèle
ε	Vecteur aléatoire caractérisant l'incertitude des mesures
Σ, U	Matrice de covariance, et vecteur de moyenne, associés à une distribution normale multidimensionnelle $N(\Sigma, U)$
σ	Ecart type
S	Nombre d'échantillons
M	Nombre de chaînes d'échantillonnage
N	Nombre d'échantillons par chaîne d'échantillonnage
L	Fonction de vraisemblance
$P(\theta)$	Probabilité a priori du vecteur de paramètre θ
$P(Data \theta)$	Probabilité de vraisemblance des données sachant un vecteur de paramètre θ
$P(Data)$	Fonction de répartition
$P(\theta Data)$	Probabilité a posteriori du vecteur de paramètre θ (sachant les données)

Paramètres physiques

ε	Emissivité totale hémisphérique des parois opaques	-
σ	Constante de Stefan-Boltzmann	$W/(m^2 K^4)$
ρ_{air}	Masse volumique de l'air	kg/m^3
c_{air}	Capacité thermique massique de l'air	$J/(K \cdot kg)$
θ_0^3	Température moyenne entre les parois lourdes et les parois légères, utilisée pour la linéarisation de l'équation de transfert radiatif, supposée égale à 293 K	K
S_{ch}	Surface chauffée	m^2
V_{ch}	Volume chauffé	m^3
Cm, Cp	Capacité thermique caractérisant les effets d'inertie liés aux parois lourdes (selon la norme ISO 13790)	kJ/K
R_{ab}	Résistance entre les nœuds a et b	
S_{roof}	Surface de toit	m^2
S_{wall}	Surface des parois verticales extérieures opaques	m^2
S_{win}	Surface des parois extérieures vitrées	m^2
A_l	Surface des parois légères	m^2
A_m, A_p	Surface massique effective (selon la norme ISO 13790)	m^2
\dot{V}_{inf}	Débit d'infiltration	V_{ch}/h
T_e	Température extérieure	K
T_{sky}	Température de ciel	K
φ_{sky_h}	Flux radiatifs entre les parois extérieures opaques et la voûte céleste	W
φ_{sky_f}	Flux radiatifs entre les parois vitrées et la voûte céleste	W
$\varphi_{sol_{opaque}}$	Flux radiatifs solaires arrivant sur les parois extérieures opaques	W
$\varphi_{sol_{in}}$	Flux radiatifs solaires pénétrant dans le bâtiment	W
φ_{heat}	Puissance du système de chauffage	W
φ_{elec}	Apports thermiques dus à l'éclairage et les appareils électriques	W
φ_{meta}	Apports thermiques dus aux métabolismes des occupants	W
φ_{int}	Apports internes ($\varphi_{elec} + \varphi_{meta}$)	W
$r_{cv_{heat}}$	Proportion de la part convective de la puissance de chauffage φ_{heat}	-
$r_{cv_{int}}$	Proportion de la part convective des apports internes φ_{int}	-
h_i	Coefficient de convection convective interne	$W/(m^2 K)$
h_e	Coefficient de convection convective externe	$W/(m^2 K)$
O_{cc}	Profil occupation moyen	-
r_{elec}	Pic des apports thermiques liés aux consommations d'électricité spécifique, par m^2 de surface chauffée	W/m^2
r_{light}	Pic des apports thermiques liés à l'éclairage, par m^2 de surface chauffée	W/m^2
r_{app}	Pic des apports thermiques liés aux appareils électriques, par m^2 de surface chauffée	W/m^2
r_{pers}	Ratio de m^2 de surface chauffée par occupant	$m^2/pers$

Chapitre 1

Contexte, enjeux et objectifs

Table des matières

Chapitre 1.....	13
1.1 Besoin de suivi et d'amélioration des performances énergétiques à l'échelle du quartier.....	15
1.1.1 Les enjeux liés à l'énergie et au bâtiment	15
1.1.2 Les actions possibles dans le secteur du bâtiment.....	15
1.1.3 Le quartier : une nouvelle échelle d'étude des actions de performance énergétique.....	17
1.1.4 Un modèle pour le suivi et l'amélioration des performances énergétiques à l'échelle du quartier	19
1.2 Données et sources d'informations disponibles à l'échelle du quartier	20
1.2.1 Les données d'exploitation.....	21
1.2.2 Les données de description du quartier	21
1.3 Choix de l'approche de modélisation	23
1.3.1 Modélisation bottom-up ou top-down ?	23
1.3.2 Modèles de signature énergétique.....	24
1.3.3 Modèles de simulation thermique dynamique (STD)	25
1.3.4 Modèles dynamiques simplifiés (RC).....	26
1.3.5 Les principaux logiciels de simulation dynamique simplifiée à l'échelle du quartier	27
1.4 Problématique de la thèse.....	29

1.1 Besoin de suivi et d'amélioration des performances énergétiques à l'échelle du quartier

1.1.1 Les enjeux liés à l'énergie et au bâtiment

La réduction des consommations d'énergie est aujourd'hui un enjeu majeur de notre société. Depuis le début de la période industrielle les consommations d'énergie mondiales ont augmenté de manière exponentielle ((IEA) 2018). Ces augmentations sont à l'origine de nombreuses externalités, dont notamment les émissions de gaz à effet de serre, responsables d'un réchauffement global de la terre (IPCC AR5 2014). Les conséquences de ce changement climatique sont multiples et affectent les systèmes physiques, biologiques et humains (IPCC AR5 2014). Il apparaît ainsi indispensable et urgent de réduire les émissions anthropiques de CO₂. Pour limiter le réchauffement climatique à 2 °C par rapport à l'ère préindustrielle, il est nécessaire de diviser les émissions de CO₂ mondiales par 2 d'ici 2030 et par 3 d'ici 2040, pour finalement atteindre un flux d'émissions net, nul en 2050 (IPCC, 2018). Le secteur du bâtiment, qui inclut le résidentiel et le tertiaire, est responsable d'environ 30 % de la consommation mondiale d'énergie finale et 40 % de la consommation Française (International Energy Agency 2020). Enfin, ce secteur du résidentiel et tertiaire représente en Europe et en France, respectivement 12 % et 16 % des émissions de gaz à effet de serre (Baude et al. 2018).

1.1.2 Les actions possibles dans le secteur du bâtiment

Dans le but de baisser les consommations d'énergie et les impacts environnementaux du secteur résidentiel et tertiaire, de nombreuses actions sont possibles à l'échelle du bâtiment : rénovation, flexibilité, sobriété.

1.1.2.1 Rénovation

Une première action est la rénovation thermique des bâtiments. Près de 80 % des bâtiments résidentiels existants en Europe ont été construits avant 1990 construction (iNSPiRe Project 2014) et consomment environ 90 % de l'énergie finale de ce secteur (The Shift Project 2017a). De même, le parc de bâtiments tertiaires en Europe est majoritairement ancien ; environ 60 % des bâtiments tertiaires ont été construits avant 1980 (The Shift Project 2017b).

Dans le domaine résidentiel, plus de 60 % de l'énergie consommée en Europe est utilisée pour le chauffage (ODYSSEE-MURE 2015). D'après (The Shift Project 2017a), il serait possible par la rénovation, de réduire les émissions de CO₂ issues du stock de bâtiments construits avant 1990, de 50 % à 80 % selon les pays de l'Union Européenne. Ces estimations sont basées sur les résultats des projets de recherche européens : (IEE Project 2016). Malgré ces potentiels, il existe de nombreux obstacles qui ralentissent fortement le processus de rénovation. Entre autres, les rénovations nécessitent des investissements initiaux importants avec des temps de retour sur investissement longs. Ainsi, des aides financières adaptées et un accompagnement des ménages sont nécessaires (The Shift Project 2017a).

1.1.2.2 Flexibilité

Un autre type d'action possible consiste à réaliser de la flexibilité électrique. Les besoins électriques des pays varient selon les moments de la journée et les périodes de l'année. Le système électrique fonctionne en adaptant à chaque instant la production à la demande. En France, RTE (Réseau de transport d'électricité), gestionnaire du réseau de transport de l'électricité, est responsable d'assurer cet équilibre. Selon les besoins, différents moyens de production sont appelés; les premiers appelés sont ceux dont le coût marginal est le plus faible. Ainsi, les coûts de production de l'électricité et les externalités associées à cette production varient à chaque instant. Par exemple, en hiver les périodes de grand froid engendrent des pics de demande qui nécessitent la mise en route de capacités de pointe, telles que les centrales à gaz, fortement émettrices de CO₂. Aussi le développement des énergies renouvelables (éolien et solaire) augmente les besoins de flexibilité. Ces sources d'énergie ont un coût marginal nul et sont donc appelées en premier sur le réseau. La variabilité de ces nouvelles sources de production a tendance à faire baisser les taux de charge (taux d'utilisation) des sources modulables existantes, et en conséquence augmente les coûts de production globaux. Ainsi, dans chacun des scénarios prévisionnels de RTE sur l'approvisionnement en électricité de la France, une augmentation des besoins en flexibilité est envisagée pour répondre aux nouvelles sources de production et aux nouveaux usages électriques (Figure 1).

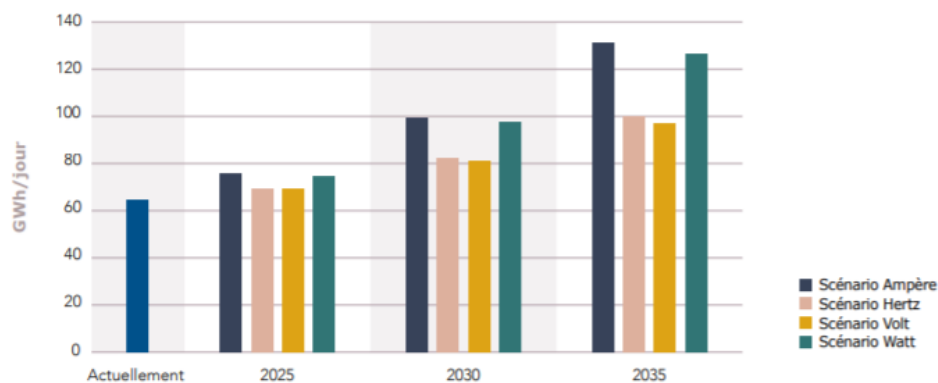


Figure 1 : Evolutions des besoins de flexibilité en volume d'énergie déplacée par jour pour répondre aux besoins du système électrique suivant 4 scénarios (RTE 2017).

1.1.2.3 Sobriété

Enfin un dernier levier d'action dans le bâtiment est la sobriété. La sobriété consiste à changer ses habitudes de vie pour moins consommer d'énergie et éviter les externalités associées. Il apparaît de plus en plus certain que l'efficacité énergétique et l'utilisation d'énergie propre ne suffiront pas à répondre aux enjeux liés aux consommations d'énergie et au changement climatique. L'énergie économisée dans de nombreux secteurs par l'amélioration de l'efficacité énergétique est en partie réutilisée pour produire davantage de biens et services, c'est l'effet rebond. La sobriété implique donc une prise de conscience globale des problèmes énergétiques. Dans le scénario énergétique Négawatt, la sobriété pourrait permettre de réduire les consommations d'énergie finale en France d'environ 30 % (Négawatt 2017a). Dans le domaine du bâtiment, la sobriété passe d'abord par une limitation des températures de consigne de chauffage, ainsi qu'une maîtrise des surfaces bâties et une modération du nombre d'équipements électriques et de l'éclairage (Négawatt 2017b).

1.1.3 Le quartier : une nouvelle échelle d'étude des actions de performance énergétique

Le quartier est aujourd'hui une échelle de plus en plus privilégiée des collectivités pour leur politique de transition énergétique. De nombreux projets sont menés à cette échelle, tels que le projet Cordees (CoResponsability in District Energy Efficiency and Sustainability) utilisée comme cas d'étude dans cette thèse. Ce projet avait pour but d'améliorer les performances énergétiques du nouveau quartier de Clichy-Batignolles à Paris, en impliquant l'ensemble des acteurs du quartier : maître d'ouvrage,

locataires, bailleurs, exploitants. Des campagnes de sensibilisation à destination des habitants, des écoles, et des travailleurs ont été mises en place, un audit des réseaux de chaleur a été réalisé, un important travail a été mené pour mesurer et centraliser des données d'exploitation du quartier : données de sous-stations, données Linky, données de mesure des températures intérieures.

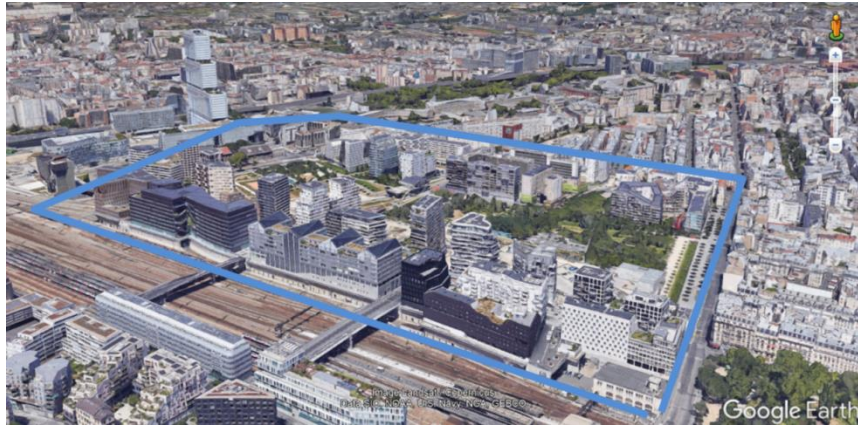


Figure 2 : Quartier de Clichy-Batignolles

Plus généralement l'échelle du quartier ouvre de nouveaux leviers d'action pour la réduction des consommations d'énergie et des émissions de CO₂ :

- Sur un groupe de bâtiments, les bâtiments prioritaires et les actions de rénovation les plus pertinentes, peuvent être identifiées.
- Des systèmes de chauffage et climatisation mutualisés peuvent être utilisés. Des réseaux de chaleur peuvent être installés, étendus et optimisés.
- Des capacités d'effacement électrique conséquentes, pourraient être mises en place, et à terme être valorisées sur le marché de l'électricité.
- Dans le cas où les bâtiments sont reliés à un réseau de chaleur alimenté par plusieurs systèmes de production, des actions d'effacement thermique peuvent favoriser l'utilisation d'une source d'énergie moins polluante.
- Des actions de rénovation ou sensibilisation, peuvent être menées de manière commune sur des bâtiments proches géographiquement et architecturalement, dans le but de réaliser des économies d'échelle (mutualisation des moyens).
- Enfin, de plus en plus de projets de construction sont réalisés à l'échelle du quartier, on parle souvent d'éco-quartiers. Il apparaît pertinent de pouvoir suivre les performances de ces quartiers qui très souvent diffèrent fortement des objectifs.

Ainsi un quartier d'étude peut être délimité de diverses manières :

- Les bâtiments partagent un système énergétique en commun relié par un réseau de chaleur. Dans ce cas, les performances du réseau sont influencées par chacune des sous-stations alimentant les différents bâtiments.
- Les bâtiments sont regroupés géographiquement en un îlot, qui présente des contraintes en alimentation électrique.
- Les bâtiments appartiennent à un quartier neuf, construit par un aménageur commun.
- Les bâtiments appartiennent à un même acteur. Cela peut être le cas pour un groupe de bâtiments universitaires ou des logements sociaux.
- Les bâtiments appartiennent à un quartier délimité par le tissu urbain. Il est possible d'imaginer que n'importe quel quartier présentant une cohérence urbaine (architecture, usages, années de construction), puisse être considéré pour la mise en place d'actions de performance énergétique mutualisée.

1.1.4 Un modèle pour le suivi et l'amélioration des performances énergétiques à l'échelle du quartier

Selon le contexte présenté ci-dessus, il semble pertinent de disposer de modèles de quartier capables d'étudier des scénarios de performance énergétique visant la baisse des consommations d'énergie et des émissions de CO₂, et de rendre possible la mise en place d'un suivi des performances énergétiques du quartier. Le suivi de performances est aujourd'hui principalement cadré par l'IPMVP (International Performance Measurement and Verification Protocol)(EVO 2010) : un protocole qui intègre différentes méthodes de mesure et vérification d'actions d'économie d'énergie, et met en avant les questions à se poser pour obtenir des mesures significatives. La démarche IPMVP est principalement composée de deux phases: la période de référence et la période de suivi. Sur la période de suivi, une courbe de référence ajustée doit être estimée ; cette courbe représente ce qu'on s'attendrait à observer si le système étudié était resté identique à ce qu'il était lors de la période de référence. Si aucune action spécifique n'a été réalisée entre la période de référence et la période de suivi, la comparaison entre la courbe de référence ajustée et la courbe de suivi peut permettre d'identifier des dérives de consommation. Si des actions de performance énergétique ont été entreprises entre la période de référence et la période de suivi, la comparaison permet d'évaluer les gains obtenus grâce à ces actions.

Ainsi cette thèse vise à développer une méthode de modélisation pour le suivi et l'amélioration des performances énergétiques d'un quartier (Figure 3). Dans une

première phase, un modèle de quartier est paramétré et calibré sur une période de référence. Ce modèle calibré peut être utilisé pour étudier différents scénarios de performance énergétique (rénovation, flexibilité, sobriété). Ensuite le modèle doit permettre de simuler une consommation de référence ajustée, pour évaluer des actions de performance énergétique mises en place, ou identifier des dérives de consommation (modifications des comportements, des systèmes de régulation ou de la structure des bâtiments). Des retours d'expérience pourront aiguiller la mise en place de nouvelles actions sur le même site ou sur d'autres cas d'étude, une plus grande confiance pourra être accordée à l'efficacité des actions, et à terme, des contrats de performance pourraient être mis en place.

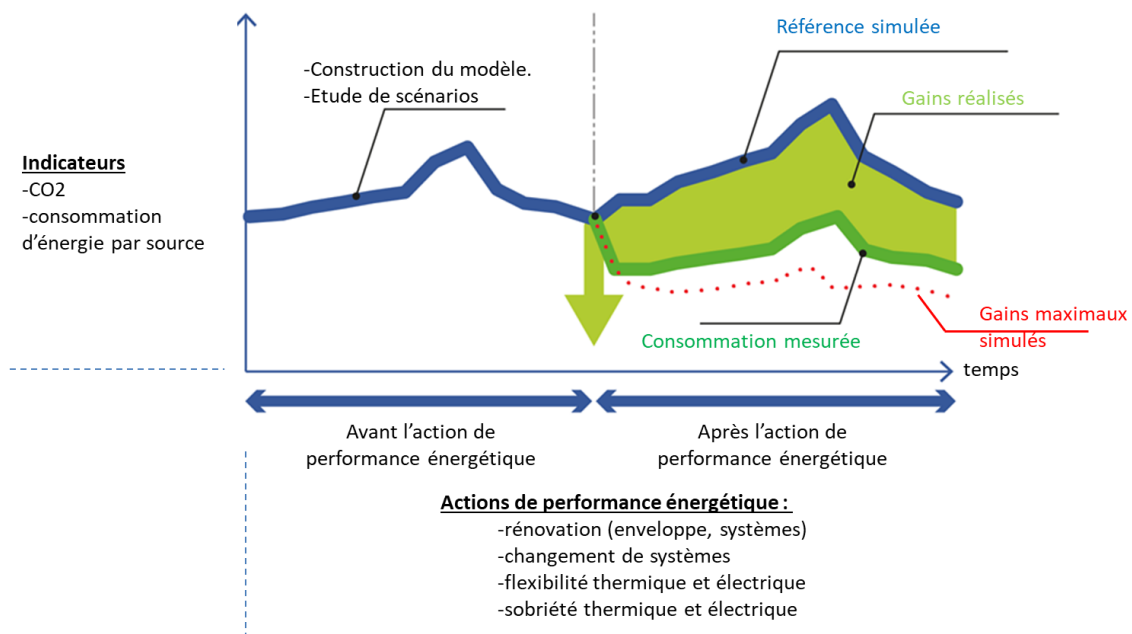


Figure 3 : Processus d'évaluation d'actions de performance énergétique sur un quartier

1.2 Données et sources d'informations disponibles à l'échelle du quartier

La construction de modèles de bâtiments à l'échelle du quartier nécessite des données à cette même échelle. Ces données, peuvent être classées en deux catégories : les données de description du quartier et les données d'exploitation.

1.2.1 Les données d'exploitation

Les données d'exploitation sont des séries temporelles issues du fonctionnement du quartier : consommation de chauffage, consommation d'électricité spécifique, température intérieure des bâtiments.

Les compteurs communicants (Linky pour l'électricité, Gazpar pour le gaz) se développent fortement et permettent de remonter automatiquement des données de consommation sur des pas de temps fins. L'utilisation des données issues des compteurs communicants est soumise à des règles d'anonymisation (Légifrance 2017). Le déploiement à l'échelle nationale ne concerne que le secteur résidentiel.

Les réseaux de chaleur sont de plus en plus répandus à l'échelle du quartier. Les consommations d'énergie peuvent être mesurées au niveau des sous-stations qui alimentent en chaleur les différents bâtiments d'un quartier pour le chauffage et l'eau chaude sanitaire.

Dans des quartiers démonstrateurs, tels que le quartier de Clichy-Batignolles, des capteurs supplémentaires peuvent être installés, pour disposer notamment des températures intérieures des bâtiments. Cependant, l'installation de capteurs à l'échelle du quartier est généralement très coûteuse et nécessite l'accord des consommateurs et de l'autorité de protection des données personnelles (CNIL).

1.2.2 Les données de description du quartier

Le second type de données regroupe les données descriptives des bâtiments du quartier. Ces dernières peuvent par exemple prendre la forme de grandes bases de données couvrant l'ensemble du territoire français. Un inventaire de ces bases de données est présenté Annexe 1. La base de données BD TOPO® publiée par l'Institut National de Géographie (IGN), apparaît particulièrement intéressante ; elle contient des formes géométriques géolocalisées selon un système de coordonnées de référence (SCR) et caractérisées par des attributs. Plusieurs thématiques sont disponibles : bâtiments, réseaux routiers, zones de végétation (Figure 4). Cette base de données peut être utilisée pour le calcul des surfaces de déperdition et de masques solaires. La base de données MAJIC (Mise À Jour des Informations Cadastrales), semble également intéressante. Cette base de données, disponible dans un cadre de recherche, regroupe les données foncières des locaux résidentiels et tertiaires non-public. Les locaux peuvent être localisés par adresse postale et de nombreuses informations sont

disponibles, telles que la surface et le code d'activité NAF des locaux tertiaires. Les données issues du recensement de population diffusées par l'INSEE présentent aussi un intérêt. Les logements résidentiels, localisés par IRIS, sont caractérisés (type de logement, profil des habitants, combustible pour le chauffage).

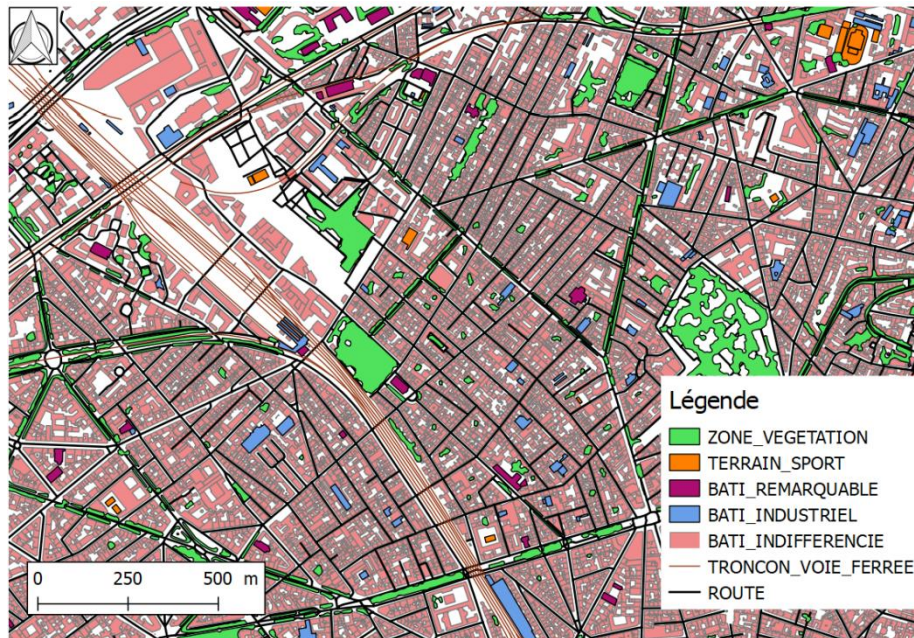


Figure 4 : Extrait de la base de données BD TOPO® autour du quartier de Clichy-Batignolles (75017 Paris)

Une autre source d'information réside dans les bases de données Google®, en consultation libre via diverses API : des images satellites (Google Satellite®), une visualisation 3D modélisée (Google Earth®), et une collection de photos haute définition, prises sur 360° depuis des voitures (Google Earth®/ Google Steet View®). Toutes ces images peuvent s'avérer utiles pour identifier des taux de vitrage, des nombres d'étages et des protections solaires.

Enfin, il est possible d'utiliser des documents de construction, issus du dossier des ouvrages exécutés (DOE) : rapports de simulations thermiques dynamiques (STD), calcul de la Réglementation thermique, plans de bâtiments, rapports de délibération communale sur un projet de construction. Ces documents peuvent être pertinents pour obtenir des informations, comme les surfaces chauffées, les types de logements et éventuellement les systèmes énergétiques utilisés. Cependant la récupération et le traitement de ces documents peuvent s'avérer particulièrement difficiles et chronophages.

1.3 Choix de l'approche de modélisation

L'objectif d'amélioration continue des performances d'un quartier, poursuivi dans cette thèse, peut être décomposé en deux sous-objectifs : l'étude de scénarios et le suivi des performances. Ci-dessous, différentes approches de modélisation sont passées en revue au regard des capacités prédictives liées à ces deux objectifs et des contraintes liées à la disponibilité des données et la rapidité des calculs.

1.3.1 Modélisation bottom-up ou top-down ?

La modélisation pour la gestion de l'énergie à l'échelle urbaine est un sujet déjà largement abordé dans la littérature. Les différentes approches ont été décrites et catégorisées (Li et al. 2017; Swan et Ugursal 2009). Selon (Swan et Ugursal 2009), deux grandes voies coexistent : les méthodes descendantes (top-down), et les méthodes ascendantes (bottom-up) (Figure 5).

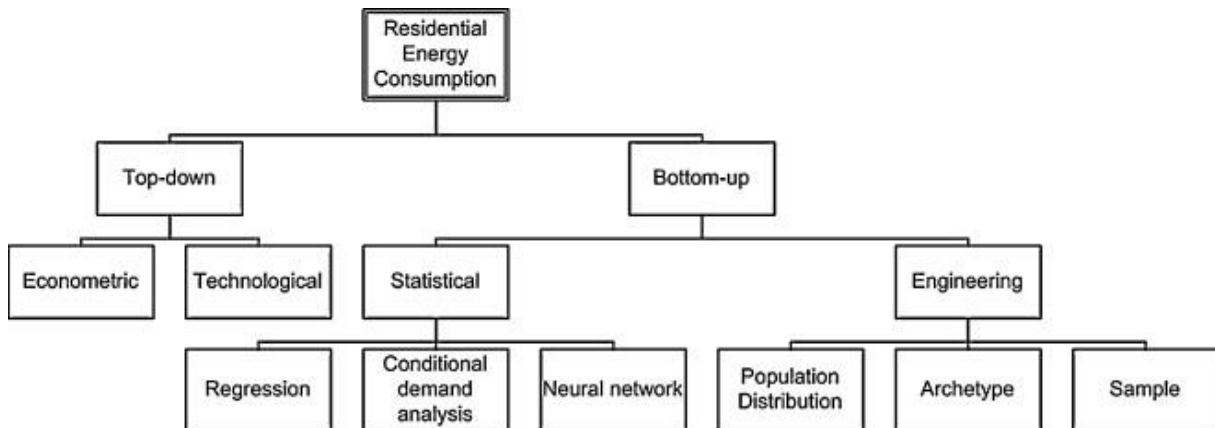


Figure 5 : Classification des approches de modélisation des consommations d'énergie territoriales (Swan et Ugursal 2009)

L'approche top-down est basée sur des données caractérisant un territoire large et pouvant être désagrégées à une échelle plus fine. On parle souvent d'approches top-down pour les modèles de simulation de consommations d'énergie à échelle macroscopique (ville, région, pays) selon des hypothèses économiques (Farzana et al. 2014) ou techniques (Sartori et al. 2009). Cependant, dans ces cas, l'approche reste macroscopique et ne considère pas l'échelle locale. Un exemple typique de modélisation top-down est l'étude de (Mikkola et Lund 2014). Dans cette étude des profils types de consommation par usage sont calibrés selon des données de consommation macro (ville). Les courbes de charge obtenues sont ensuite désagrégées

à l'échelle du quartier selon des données statistiques disponibles à cette échelle. Ces résultats ont pour but la définition de stratégies de gestion locale de l'énergie (Smart-Grid). L'avantage des modèles top-down est qu'ils nécessitent peu de données et sont robustes à l'échelle macro. Cependant, cette approche, basée sur l'attribution de profils de consommation types, ne permet pas d'étudier l'évolution de ces profils selon des paramètres tels que l'efficacité des systèmes, l'enveloppe des bâtiments, le comportement des occupants ou même la température extérieure. Cette approche est écartée par la suite. En revanche, l'approche bottom-up reposant sur l'utilisation de modèles plus détaillés par bâtiment d'un quartier s'avère nécessaire pour répondre à la problématique de cette thèse (Figure 5).

Dans l'approche bottom-up, des données locales de description du territoire sont utilisées pour simuler les consommations de chaque usage (bâtiments, systèmes, appareils électriques...). Ces consommations unitaires peuvent ensuite être agrégées et étudiées à une échelle plus large comme le quartier. Au sein de cette approche, (Swan et Ugursal 2009) distinguent deux branches : les modèles statistiques et les modèles physiques (Figure 5). L'approche statistique relie par des modèles de régression les consommations des utilisateurs finaux à leurs données descriptives (âge des occupants, âge des bâtiments...), on parle de modèle « boîte noire » ; des données d'apprentissage/calibration sont nécessaires. La complexité des modèles statistiques est variable, allant de modèles simples de signature énergétique à des modèles complexes de machine learning tels que les réseaux de neurones. La seconde approche modélise les processus physiques liés aux consommations d'énergie. A nouveau, les modèles peuvent être détaillés tels que les logiciels de Simulation Energétique Dynamique (SED) : TRNSYS, EnergyPlus, Pleiades Comfie ou simplifiés en un nombre réduit de résistances et capacités thermiques (R6C2, R5C1...)(Berthou 2013)(ISO 2008); on parlera alors de modèles dynamiques simplifiés. Dans les paragraphes suivants, les modèles de signature énergétique, machine learning, STD, et dynamique simplifiés sont discutés.

1.3.2 Modèles de signature énergétique

Les modèles statistiques pour la prévision des consommations de chauffage comprennent notamment les modèles de signature énergétique qui consistent en des régressions du premier ou du deuxième degré entre des données de consommation et de température extérieure. Les données de température peuvent être des moyennes ou

des degrés-jours. Des coefficients supplémentaires liés à l'occupation ou à l'ensoleillement peuvent également être ajoutés. Les modèles sont rapides, nécessitent peu de données d'entrée et sont relativement justes sur des pas de temps mensuels. Cette méthode est aujourd'hui largement utilisée pour le suivi de performances de bâtiments. Cela s'explique par la disponibilité des données et sa facilité d'utilisation et d'interprétation des résultats. Des modèles plus complexes, sont pour le moment, jugés non pertinents car le travail requis pour augmenter la justesse en augmentant la quantité de données utilisées n'est pas estimé rentable. Cependant, sur des quartiers neufs, les consommations sont difficiles à prévoir avec des modèles de signature énergétique, car elles sont particulièrement sensibles à l'activité des occupants et à l'ensoleillement. Aussi, cette méthode n'est pas robuste à des changements d'usages ou de systèmes. Il est intéressant d'étudier de nouvelles possibilités de suivi de performance en s'appuyant sur une modélisation plus complexe et une quantité croissante de données disponibles. De plus, ces modèles de signature ne permettent pas d'étudier des scénarios de performance énergétique, on s'intéressera donc à des modèles plus complexes, capables de représenter une plus grande diversité de phénomènes et d'intégrer plus d'informations.

1.3.3 Modèles de simulation thermique dynamique (STD)

La simulation thermique dynamique (STD) des bâtiments permet de modéliser en détail les différents phénomènes physiques. Ainsi, cette approche peut permettre une très bonne précision pour l'étude de scénarios et le suivi de performance, mais son utilisation est contrainte par des temps de paramétrage et de calibration élevés. Le paramétrage impose une connaissance fine des bâtiments, dont une description précise des géométries, des systèmes CVC, des matériaux de construction et des usages. Ce niveau de connaissance n'est aujourd'hui pas envisageable à l'échelle du quartier. En effet, il est difficile d'obtenir l'ensemble des données nécessaires à une modélisation détaillée des bâtiments d'un quartier (plans détaillés de chaque bâtiment, description précise des systèmes et de l'occupation). Quand bien même ces documents seraient disponibles, il serait long de construire un modèle STD de chaque bâtiment. Le BIM (Building Information Modeling) vise à rendre possible de telles modélisations, cependant ces sources d'information ne seront probablement disponibles que dans plusieurs années, et uniquement pour un nombre réduit de quartiers neufs. Il est cherché ici à pouvoir modéliser n'importe quel quartier, ancien, ou récent, en un

temps raisonnable. De plus, la généralisation du BIM sur un large territoire pourrait s'avérer coûteuse des points de vue financiers et environnementaux, car très intensif en calculs. La calibration de ces modèles sera nécessaire pour réduire les incertitudes de paramétrage et permettre le suivi de performance des bâtiments du quartier. Mais cette calibration risque aussi de s'avérer difficile et chronophage, étant donné le nombre de paramètres incertains et l'intensité des calculs.

1.3.4 Modèles dynamiques simplifiés (RC)

Les modèles dynamiques simplifiés, dits RC, se positionnent, en termes de justesse, de temps de calcul, de facilité de paramétrage, et de facilité de calibration, comme un niveau intermédiaire entre les modèles de signature énergétique et les modèles STD. Ce compromis est synthétisé dans la Figure 6. En effet, cette approche consiste à utiliser des modèles de bâtiments simplifiés inspirés des phénomènes physiques. Un paramétrage à partir d'une description physique réduite des bâtiments sera alors possible ; et la calibration, moins coûteuse en calcul, pourra être envisagée à l'échelle du quartier. C'est donc cette approche qui a été retenue dans cette thèse. On cherchera à explorer quelles sont les possibilités offertes par cette approche en matière de justesse et temps de calcul, dans le but de simuler les consommations en énergie d'un quartier pour évaluer des actions de performance énergétique et faire du suivi de performance. Plusieurs plateformes de simulation des consommations d'énergie à large échelle utilisant cette approche ont déjà été développées, et sont présentées dans le paragraphe suivant.

Approche de modélisation Critères de choix	Modèle de signature ($kWh = a \cdot DJU + b$)	Modèles dynamiques simplifiés (RC)	Simulation thermique dynamique (ex. Energy+, PleiadeComfie)
Justesse pour le suivi de performances	Performant	Très performant	Très performant
Justesse pour la simulation de scénarios de performance énergétique	Impossible	Performant	Très performant
Temps de calcul	Très performant	Très performant	Peu performant
Facilité de calibration	Très performant	Performant	Peu performant
Facilité de paramétrage physique	Impossible	Très performant	Peu performant





	Très performant		Performant
	Peu performant		Impossible

Figure 6 : Comparaison de trois approches de modélisation

1.3.5 Les principaux logiciels de simulation dynamique simplifiée à l'échelle du quartier

Anciennement appelé Sun-Tool; CitySim (Robinson 2009) est un des premiers modèles adaptés à la simulation dynamique du bâtiment à l'échelle urbaine. Il est basé sur un modèle d'enveloppe R5C2 (2 capacités et 5 résistances) multizones. Cet outil propose notamment un modèle détaillé d'ensoleillement nécessitant une description 3D des bâtiments. Aussi les systèmes de chauffage, climatisation et ventilation sont modélisés. CitySim était, au départ, paramétré manuellement, il a plus tard été réutilisé dans la plateforme MEU (Management énergétique urbain)(Darmayan et al. 2016).

MEU est une plateforme de simulation des flux d'énergie à échelle urbaine à destination des différents acteurs de l'énergie (collectivités, distributeurs, gestionnaires). Le modèle se base sur des données de description des bâtiments à large échelle, renseignant notamment les : périodes de construction, emprises au sol, hauteurs et usages des bâtiments. Ce type de données est particulièrement bien renseigné en Suisse, pays d'origine du logiciel. À partir de ces données, deux méthodes de modélisation sont possibles. La première se base sur des données statistiques de consommations annuelles selon les usages et les périodes de

construction. La seconde fait appel à l’outil CitySim décrit précédemment. Des modèles de réseaux de chaleur, de froid, d’électricité et de gaz ont en plus été développés. Aussi, la plateforme dispose d’une interface graphique pour l’étude de scénarios de gestion de l’énergie.

Teaser (Remmen et al. 2018) est un outil conçu dans le but de limiter le travail de collecte de données et de paramétrage en proposant une structure pouvant inclure un large panel de données : données statistiques de description des bâtiments, archétypes, standards internationaux. L’outil permet de paramétrer un modèle RC de simulation des demandes thermiques (chaud et froid) d’un parc de bâtiments (quartier ou ville). Les comportements sont modélisés par des profils horaires issus de (Swiss Society of Engineers and Architects 2006). Les propriétés thermiques des murs sont estimées selon l’âge des bâtiments à partir d’une étude allemande (BMVBS 2010).

City Energy Analyst (Fonseca et al. 2016) est un logiciel open-source particulièrement abouti. En effet, cet outil traite tous les secteurs : résidentiel, tertiaire et industriel. Un grand nombre de modèles de systèmes sont disponibles : production locale d’énergie, CVC, réseaux de chaleur. Le modèle thermique de bâtiments R5C1, issus de la norme ISO 13790, est utilisé. La plateforme considère la géométrie des bâtiments et les phénomènes de rayonnement solaire et d’ombres portées sont modélisés par le logiciel Daysim (Reinhart s. d.). La plateforme est en accès libre sur Github¹, seule l’interface de visualisation des résultats n’est pas directement accessible, car elle fait appel au logiciel sous une licence privée ArcGis®.

PowerDis est un logiciel développé par l’institut semi-privé Efficacity, proche de City Energy Analyst, ayant pour but de réaliser des études sur la performance énergétique prévisionnelle de projets d’aménagement à l’échelle du quartier et de la ville. Comme pour CEA, de nombreux systèmes sont modélisés (production locale d’énergie, CVC, réseaux de chaleur). Cependant, la plateforme utilise comme cœur de calcul thermique le logiciel DIMOSIM, développé par le CSTB, basé sur des modèles RC multizones, présentant un haut niveau de détail. Les bases de données et les modèles ne sont pas publics.

Smart-E est une plateforme de simulation des consommations énergétiques des territoires développée au laboratoire CES de l’école Mines ParisTech (Berthou et al.

¹ <https://github.com/architecture-building-systems/CityEnergyAnalyst>

2015, 2019). Cette plateforme a pour but de simuler automatiquement les consommations d'énergie à échelle territoriale. Pour cela, elle fait appel aux bases de données nationales de description du territoire (recensement de la population INSEE et BD_TOPO IGN® (voir annexe 2)). Ces bases de données possèdent des informations relatives aux logements et à leurs occupants. Les informations sur les occupants (âges et activités) permettent de reconstituer des courbes d'utilisation d'appareils domestiques. Aussi, les informations sur les types de logements (individuels ou collectifs), les surfaces et les sources d'énergie utilisées pour le chauffage (gaz, électricité, fioul) permettent de simuler les consommations d'énergie.

Malgré ces travaux, aucune de ces plateformes ne propose une méthode permettant d'utiliser l'ensemble des informations disponibles à l'échelle du quartier pour l'étude de scénarios et le suivi de performance. Dans le paragraphe suivant, une formulation bayésienne du problème est présentée, les verrous scientifiques sont identifiés, et les travaux réalisés dans cette thèse sont introduits.

1.4 Problématique de la thèse

Dans cette thèse, il sera proposé une méthode de modélisation de quartiers pour étudier des scénarios de performance énergétique (rénovation, flexibilité, sobriété), identifier des dérives de consommation et les causes de ces dérives, et évaluer des actions de performance énergétique mises en place (IPMVP). Une idée centrale est que la méthode puisse utiliser l'ensemble des données disponibles d'un quartier : données de description et données d'exploitation. Cette idée peut être présentée sous l'angle de la théorie Bayésienne.

Selon l'approche Bayésienne, la probabilité d'une théorie, qui traduit la robustesse de la théorie ou la confiance que l'on a en la théorie, évolue au fil du temps. À chaque acquisition de nouvelles données, une inférence bayésienne est effectuée ; c'est la mise à jour de la probabilité de la théorie. Cette inférence qui calcule la probabilité de la théorie sachant les nouvelles données est régie par la formule de Bayes (équation (1)). Cette formule regroupe deux termes principaux : la probabilité de vraisemblance $P(Data|théorie)$ qui considère l'information apportée par les nouvelles données, et la probabilité a priori de la théorie $P(théorie)$: la probabilité que l'on attribue à la théorie avant de connaître les nouvelles données. Cette dernière probabilité doit refléter l'ensemble des informations recueillies dans le

passé, telles que : des jeux de données similaires récemment acquis, des connaissances issues de la littérature scientifique, ou encore une expertise pratique professionnelle. On reviendra plus en détail sur l’approche bayésienne dans le chapitre 2.

$$P(\text{théorie}|\text{Data}) = \frac{P(\text{Data}|\text{théorie})P(\text{théorie})}{P(\text{Data})} \quad (1)$$

Dans cette thèse, une théorie correspondra à un modèle de quartier simulant les consommations d’énergie des bâtiments. Les nouvelles données acquises au fil du temps seront des données d’exploitation (consommations d’énergie, températures intérieures). Les probabilités a priori sur les modèles devront considérer la littérature scientifique ainsi que l’ensemble des données disponibles sur chaque quartier décrivant les usages et les caractéristiques de construction : bases de données, documents d’architecte, données expertes. Pour réduire le domaine d’étude, une catégorie de modèle a déjà été sélectionnée pour la simulation des consommations de chauffage : les modèles RC. Une structure spécifique de modèle RC sera aussi sélectionnée dans le chapitre 3 et les paramètres du modèle seront calibrés par inférence bayésienne.

La méthode globale est résumée en Figure 7. Deux périodes sont considérées : une période de référence et une période de suivi sur laquelle l’évolution des performances du quartier est étudiée. Ces deux périodes peuvent correspondre à deux saisons hivernales consécutives. Les données de description du quartier sont utilisées pour définir des probabilités a priori associées aux paramètres calibrés du modèle. Les données d’exploitation sont utilisées pour calibrer le modèle par inférence bayésienne. Une fois calibré, le modèle peut être utilisé pour simuler une consommation de référence utilisée pour évaluer une action de rénovation ou identifier d’éventuelles dérives de consommation. Le modèle peut aussi être utilisé pour simuler des scénarios de rénovation ou de flexibilité et identifier les causes de dérives de consommation.

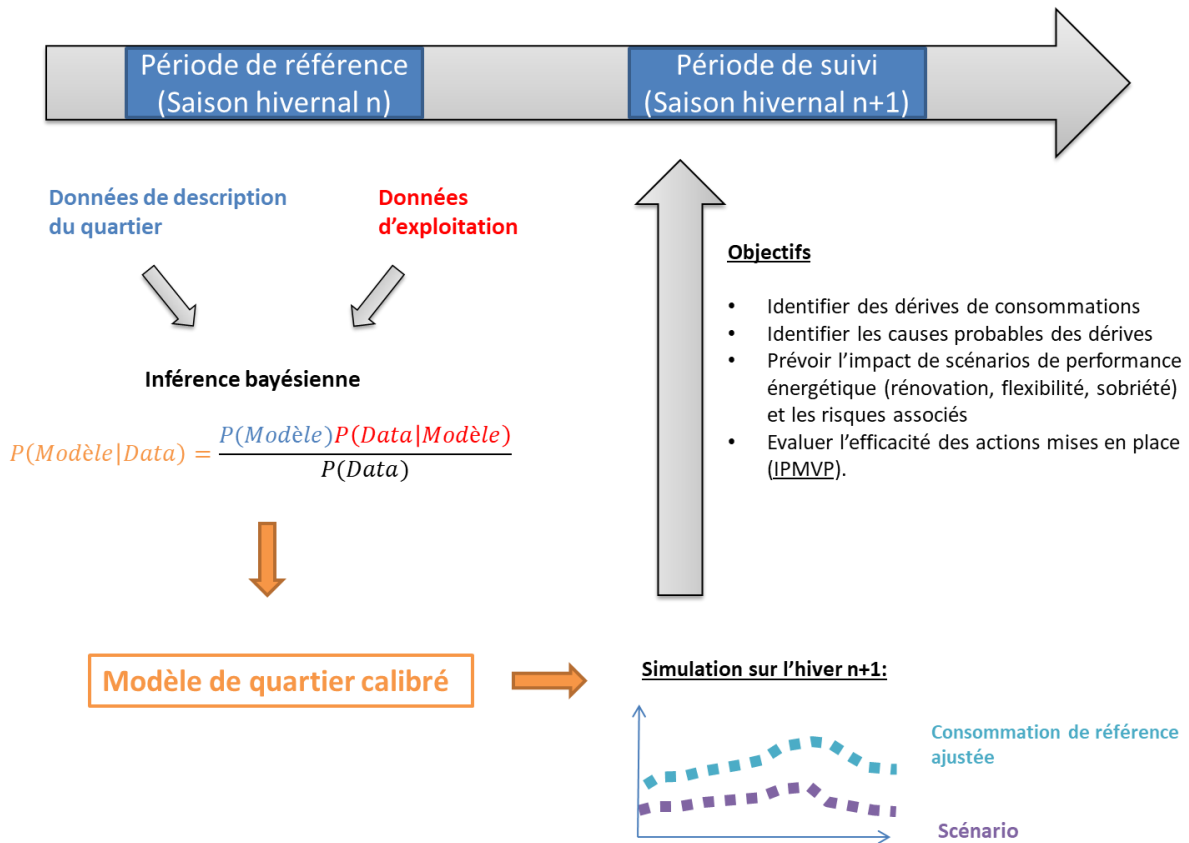


Figure 7 : Méthode globale d'estimation de scénarios et de courbes de référence ajustées sur un quartier

Cette méthode globale pose de nombreuses questions auxquelles il sera cherché à répondre dans les chapitres de la thèse :

- Quelle méthode d'inférence bayésienne utiliser pour la calibration des modèles ? Dans le chapitre 2, les formulations mathématiques, les hypothèses, et les outils informatiques utilisés pour les inférences bayésiennes sont présentés en détail.
- Quel modèle RC utiliser ? Le choix des modèles RC repose notamment sur le fait que les paramètres de ces modèles ont un lien avec les propriétés physiques des bâtiments. Cela est nécessaire pour l'estimation des paramètres a priori, pour l'étude de scénarios, et pour l'identification des causes de dérives, on parlera notamment d'« identifiabilité » physique du modèle. Le chapitre 3 cherche à sélectionner un modèle RC identifiable et présentant de bonnes capacités prédictives, adapté à la calibration bayésienne à l'échelle du quartier.
- Comment, et à partir de quelles données, estimer des modèles a priori pour les bâtiments d'un quartier ? Le chapitre 4 présente une méthode de construction d'une base de données permettant la création d'un modèle de quartier a priori. La

méthode est illustrée sur le cas d'étude du quartier de Clichy-Batignolles. Des bases de données territoriales sont utilisées.

- En fonction des données disponibles, que peut-on attendre de cette méthode de calibration de modèles RC au regard des objectifs de la thèse : étude de scénarios performance énergétique, évaluation d'action de performance énergétique mises en place (IPMVP), identification de dérives de consommation et des causes des dérives ? Dans le chapitre 5, la méthode de calibration est appliquée sur plusieurs cas d'étude, considérant différents niveaux de connaissance des bâtiments; les hypothèses sont discutées, et les limites de la méthode sont identifiées.

Chapitre 2

Calibration par méthode Bayésienne

Table des matières

Chapitre 2.....	33
2.1 Introduction.....	35
2.2 L'approche bayésienne.....	36
2.3 Choix des probabilités a priori.....	37
2.4 Réduction du nombre de paramètres calibrés	38
2.5 Inférence bayésienne	39
2.5.1 Choix de la fonction de vraisemblance	39
2.5.2 Calcul de la probabilité a posteriori.....	45
2.5.3 Les moteurs d'inférence markoviens	47
2.6 Vérification de la convergence	50
2.7 Calcul de la prédiction a posteriori.....	51
2.8 Utilisation d'un métamodèle.....	52
2.9 Conclusion	53

2.1 Introduction

Dans le domaine de la simulation thermique du bâtiment, il existe toujours des paramètres incertains qui influent de manière significative sur les consommations d'énergie. Ainsi d'importantes dérives peuvent être observées entre les résultats de simulations et les consommations réelles. Pour réduire ces dérives, il est possible de faire appel à des méthodes de calibration des paramètres incertains. A l'échelle du bâtiment, de nombreuses méthodes existent (Coakley et al. 2014). Une première possibilité est la calibration manuelle, cette approche nécessite des connaissances expertes, et peut difficilement être optimale. En conséquence, des méthodes automatiques ont été développées (Chaudhary et al. 2016). L'utilisation de ces méthodes est aujourd'hui fortement facilitée par le développement de bibliothèques informatiques disponibles en open-source. On privilégiera alors une approche automatique, et plus particulièrement l'approche bayésienne, déjà évoquée en introduction et présentée en détail dans ce chapitre. Les principales caractéristiques de l'approche bayésienne sont résumées ci-dessous :

- Toutes les sources d'informations et d'incertitudes sont considérées : connaissances a priori (données locales et avis d'experts) et données de mesures.
- Les estimateurs sont robustes aux problèmes liés aux multi-dimensionnalités, aux minimums locaux et aux distributions multimodales.
- Les résultats de calibration sont des distributions de probabilités et l'incertitude associée aux paramètres calibrés et aux prédictions peut être évaluée.
- En contrepartie, cette approche nécessite de très nombreuses simulations, et peut ainsi s'avérer très coûteuse en temps de calcul.

2.2 L'approche bayésienne

Dans le but d'étudier la pertinence des théories scientifiques, deux approches se distinguent : le fréquentisme et le bayésianisme. L'approche fréquentiste se focalise sur ce qu'on appelle la probabilité de vraisemblance : $P(Data|Théorie)$. Ce terme représente la probabilité d'observer des données en supposant une théorie vraie. L'approche fréquentiste fonctionne notamment par rejet d'hypothèses : si la probabilité d'observer des données en supposant une théorie vraie, est inférieure à un seuil fixé (la p-value), alors l'hypothèse est rejetée (Hoang 2018). Dans l'approche bayésienne on s'intéresse au terme $P(Théorie|Data)$ qui est la probabilité de la théorie sachant les données observées, on parlera de crédence ou de probabilité a posteriori (après prise en compte des données). D'après la formule de Bayes présentée dans les équations (2) et (3), cette probabilité dépend de la vraisemblance $P(Data|Théorie)$ utilisée dans le fréquentisme, mais aussi de la probabilité, dite a priori, de la théorie $P(Théorie)$. Cette dernière probabilité est fondamentale dans l'approche bayésienne, elle traduit le fait qu'il n'est pas raisonnable de croire à 100 % aux données recueillies, il faut également considérer l'a priori que l'on a sur la théorie. Cela peut être illustré par une des maximes de l'approche bayésienne énoncée par Carle Sagan « Toute affirmation extraordinaire requiert des preuves extraordinaires » (Hoang 2018). Le dernier terme $P(Data)$, parfois appelé fonction de répartition, représente la somme des probabilités d'observer les données selon toutes les théories possibles ; il garantit que la somme des probabilités de toutes les théories soit égale à 1. Dans l'approche bayésienne on ne rejette pas les théories, mais plusieurs théories coexistent avec des probabilités différentes, ces probabilités peuvent être régulièrement mises à jour avec les nouvelles données.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

$$P(Théorie A|Data) = \frac{P(Data|Théorie A)P(Théorie A)}{P(Data) = \sum_{Théorie} P(Data|Théorie)P(Théorie)} \quad (3)$$

L'approche bayésienne peut être utilisée pour calibrer les paramètres d'un modèle avec des données issues de mesures. Les théories possibles deviennent les différentes valeurs possibles du vecteur de paramètres θ :

$$P(\theta|Data) = \frac{P(Data|\theta)P(\theta)}{\int_{\Omega} P(Data|\tau)P(\tau)d\tau} \quad (4)$$

L'approche bayésienne a déjà été utilisée plusieurs fois pour la calibration de modèles thermiques de bâtiments (Chong et Menberg 2018; Heo et al. 2012; Kang et Krarti 2016; Kristensen et al. 2017; Raillon et Ghiaus 2018; Yuan et Su 2017). Une méthode générale de calibration semble aujourd'hui établie (voir Figure 8). Dans un premier temps, les paramètres incertains sont associés à des lois de probabilité a priori. Les paramètres peu influents sont identifiés et écartés pour limiter le temps de calcul nécessaire à la calibration. Ensuite, à partir des distributions a priori et des données de calibration, une inférence bayésienne (calcul de la distribution a posteriori) est effectuée. Enfin, la consistance des résultats est analysée et des prédictions a posteriori sont calculées.

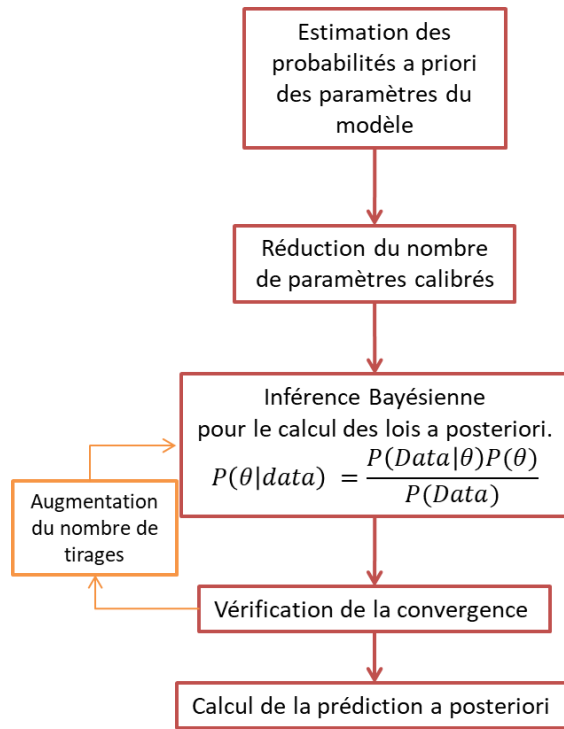


Figure 8 : Méthode globale de calibration bayésienne

2.3 Choix des probabilités a priori

Une étape importante consiste à estimer les lois de probabilité a priori des paramètres calibrés θ . Ces probabilités doivent rendre compte des connaissances et incertitudes de l'expérimentateur, et peuvent être plus ou moins informatives. Plus la distribution de probabilité est resserrée autour d'une valeur, plus elle est dite

informative. Utiliser une probabilité non-informative, revient à faire complètement confiance aux données. Une loi a priori informative permet d'éviter le sur-apprentissage. Cette dernière loi ne fait pas complètement confiance aux données, on parle parfois de lois de régularisation (Martin 2018).

2.4 Réduction du nombre de paramètres calibrés

Réduire le nombre de paramètres calibrés permet de limiter les coûts de calcul. Aussi, calibrer un grand nombre de paramètres peut s'avérer peu productif lorsque les données d'entrée ne permettent pas d'identifier ces paramètres (Lim et Zhai 2017). Ainsi, il est pertinent d'identifier et écarter les paramètres peu influents. Cela peut être fait par étude bibliographique et par analyse de sensibilité. Il existe plusieurs méthodes d'analyse de sensibilité, mais seule l'analyse de Sobol (Sobol 1993) permet de considérer en entrée des lois de probabilité a priori (Kristensen & Petersen, 2016) (Figure 9).

Property	Sensitivity analysis method		
	Local	Morris	Sobol'
Scale and shape of input parameter PDF	No	No	Yes
Multi-dimensional averaging	No	No/Yes	Yes
Model independent	No	Yes	Yes
Relative computational time	Low	Medium	High

Figure 9 : Comparaison de diverses méthodes d'analyse de sensibilité (Kristensen et Petersen 2016)

Aussi (Yamaguchi et al. 2013), propose une méthode de calibration hiérarchique, ou plusieurs inférences bayésiennes sont réalisées de manière hiérarchique sur des paramètres et résolutions temporelles différentes. Certains paramètres sont supposés avoir une influence sur les consommations annuelles et d'autres sont supposés influencer plus fortement sur les données de consommation mensuelles. Cette approche pourrait apporter une expertise physique supplémentaire et rendre la calibration plus rapide ; elle n'a cependant pas été abordée dans cette thèse et pourra faire l'objet de futurs travaux.

2.5 Inférence bayésienne

2.5.1 Choix de la fonction de vraisemblance

L'étape d'inférence bayésienne nécessite une hypothèse sur la fonction de vraisemblance $P(Data|\theta)$. Le lien entre les données de calibration observées et les résultats de simulations issues d'un modèle η peut être formulé statistiquement selon l'équation (5). Chaque donnée D_k du vecteur de données D , de longueur n , est égale au résultat de simulation η additionné à deux variables aléatoires δ et ε caractérisant respectivement l'inadéquation du modèle et l'incertitude des mesures. Le modèle considère en entrée deux types de paramètres : les paramètres dits observables x , tels que la température extérieure, ou le pas de temps, dans le cas d'un modèle dynamique, et les paramètres non-observables θ , calibrés, tels qu'une résistance thermique.

$$D_k = \eta_k(\theta, x) + \varepsilon_k + \delta_k \quad (5)$$

Les vecteurs aléatoires ε et δ sont généralement associés à des lois normales multidimensionnelles :

$$\varepsilon \sim N(U_\varepsilon, \Sigma_\varepsilon) \quad (6)$$

$$\delta \sim N(U_\delta, \Sigma_\delta) \quad (7)$$

La fonction de vraisemblance L représente la probabilité d'observer un jeu de données, sachant un modèle, et sachant l'incertitude sur les données observées :

$$L = P(D|\eta(\theta), \varepsilon, \delta) = \frac{1}{(2\pi)^{n/2} |\Sigma_L|^{1/2}} e^{-\frac{1}{2}(D-\eta(\theta)+U_L)^T \Sigma_L^{-1} (D-\eta(\theta)+U_L)} \quad (8)$$

$$\Sigma_L = \Sigma_\varepsilon + \Sigma_\delta \quad (9)$$

$$U_L = U_\varepsilon + U_\delta \quad (10)$$

2.5.1.1 Hypothèse des erreurs indépendantes et identiquement distribuées (i.i.d)

Une première approche est de supposer que les erreurs sont non biaisées ($U_L = \mathbf{0}_n$), indépendantes (Σ_L est une matrice diagonale), et identiquement distribuées (les valeurs des éléments diagonaux de Σ_L sont identiques) (Hastie et al. 2008; Manfren et

al. 2013). La probabilité d'observer chaque donnée D_k est alors issue d'une loi normale centrée sur la valeur simulée par le modèle $\eta(\theta, \mathbf{x}_k)$, et d'écart-type $\sigma_\varepsilon + \sigma_\delta$:

$$P(D_k | \eta_k(\theta, \mathbf{x}), \varepsilon, \delta) = \frac{1}{(\sqrt{\sigma_\varepsilon^2 + \sigma_\delta^2})\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{D_k - \eta_K(\theta, \mathbf{x})}{\sqrt{\sigma_\varepsilon^2 + \sigma_\delta^2}} \right)^2} \quad (11)$$

$$\varepsilon \sim N(0_p, I_n \sigma_\varepsilon) \quad (12)$$

$$\delta \sim N(0_p, I_n \sigma_\delta) \quad (13)$$

Cette approche revient à modéliser les erreurs entre les données et les simulations par un seul vecteur aléatoire g regroupant les erreurs de mesure et de modélisation :

$$D_k(\mathbf{x}_k) = \eta_k(\theta, \mathbf{x}) + g \quad (14)$$

$$g \sim N(0_n, I_n \sigma_g) \quad (15)$$

$$\sigma_g = \sqrt{\sigma_\varepsilon^2 + \sigma_\delta^2} \quad (16)$$

La fonction de vraisemblance associée à cette hypothèse sera appelée $L_{i.i.d}$, elle correspond au produit des probabilités d'observer indépendamment chaque donnée D_k :

$$L_{i.i.d} = \prod_{k=1}^n \frac{1}{\sigma_g \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{D_k - \eta_K(\theta, \mathbf{x})}{\sigma_g} \right)^2} \quad (17)$$

Le paramètre d'écart-type σ_g est souvent appelé hyper-paramètre, et est ajouté au vecteur des paramètres non observables calibrés θ .

Une limite importante de cette approche est qu'elle ne prend pas en compte les corrélations temporelles entre les données. L'existence de ces corrélations temporelles peut être illustrée par le fait que deux données très proches temporellement l'une de l'autre ne seront pas deux fois plus informatives qu'une seule des deux données.

2.5.1.2 Prise en compte des corrélations entre les erreurs et les entrées observables

L'inadéquation du modèle δ est parfois modélisée par un processus gaussien défini par une moyenne $\mathbf{0}_n$, et une fonction de covariance notée K , permettant de calculer la covariance associée à n'importe quel couple de données (D_i, D_j) . La fonction de covariance K , aussi appelée noyau, doit respecter certaines propriétés pour être valide et assurer que la matrice de covariance soit toujours symétrique et définie positive

(Larochelle 2015). Un noyau souvent utilisé est le noyau gaussien, ou noyau RBF (Radial Basis Fonction), qui fait intervenir l'exponentielle de la distance euclidienne entre les données. Cette distance peut être calculée à partir des données observables :

$$\delta \sim GP\left(\mathbf{0}_n, K(D_i, D_j)\right) \quad (18)$$

$$K(D_i, D_j) = \Sigma_{\delta, ij} = \frac{1}{\lambda_\delta} \exp\left(-\sum_{k=1}^p \beta_k^\delta |x_{ik} - x_{jk}|^\alpha\right) \quad (19)$$

Dans la littérature, ce modèle d'inadéquation est utilisé pour calibrer des modèles thermiques avec des données mensuelles ou annuelles (Chong et al. 2017; Ling et al. 2014). A ces échelles de temps, les erreurs ne sont pas corrélées temporellement et le temps ne fait pas partie des données observables \mathbf{x} considérées. Les hyper-paramètres λ_δ et β_k^δ , sont ajoutés au vecteur des paramètres calibrés θ . La prise en compte des corrélations des erreurs avec l'ensemble des paramètres observables augmente fortement le nombre de paramètres à calibrer, cette approche ne sera donc pas retenue dans la thèse.

2.5.1.3 Proposition d'une formulation statistique originale pour la prise en compte des corrélations temporelles

Dans le cas de la calibration d'un modèle thermique dynamique avec des données horaires, ou infra horaires, il semble pertinent de prendre en compte les corrélations temporelles entre les erreurs. Pour cela, il est d'abord proposé de modéliser le vecteur aléatoire \mathbf{g} par un processus gaussien défini par une moyenne $\mathbf{0}_n$, et un noyau K de type RBF, ne considérant pour le calcul de la distance euclidienne entre les données, que le temps t . Ainsi plus les données sont proches temporellement, plus elles sont corrélées :

$$\mathbf{g} \sim GP\left(\mathbf{0}_n, K(D_i, D_j)\right) \quad (20)$$

$$K(D_i, D_j) = \sigma_g^2 e^{-\frac{(t_i - t_j)^2}{2l_k^2}} \quad (21)$$

D'autres noyaux valides peuvent être obtenus par addition et multiplication de noyaux valides (Larochelle 2015). Il est ainsi proposé d'ajouter deux composantes au noyau défini ci-dessus :

$$K(D_i, D_j) = \Sigma_{LGP,ij} = \sigma_g^2 e^{-\frac{(t_i-t_j)^2}{2l_1^2}} e^{-\frac{2\sin^2\left(\frac{\pi(t_i-t_j)}{24}\right)}{l_2^2}} e^{-\frac{(P_i-P_j)^2}{2l_3^2}} \quad (22)$$

La première composante correspond au noyau présenté en équation (20). La seconde composante est un noyau périodique qui permet prendre en compte le fait que les données sont corrélées sur une période de 24h. Enfin, un troisième noyau permet de prendre le paramètre observable correspondant à la puissance thermique fournie par le système de chauffage. Pour des puissances éloignées, les données seront alors peu corrélées. Ces dernières seront notamment peu corrélées entre elles aux moments des changements de température de consigne. Les hyper-paramètres l_1 , l_2 , et l_3 sont ajoutés au vecteur des paramètres calibrés θ . La fonction de vraisemblance résultante sera appelée L_{GP} :

$$L_{GP} = \frac{1}{(2\pi)^{n/2} |\Sigma_{LGP}|^{1/2}} e^{-\frac{1}{2}(D-\eta(\theta))^T \Sigma_{LGP}^{-1} (D-\eta(\theta))} \quad (23)$$

2.5.1.4 L'hypothèse du filtre de Kalman

Les corrélations temporelles entre les données peuvent aussi être prises en compte en faisant appel à un filtre de Kalman. Le filtre de Kalman est un algorithme qui permet d'estimer l'état d'un système dynamique à partir de données incertaines (bruitées). Cet algorithme s'applique à un modèle d'état linéaire, s'écrivant sous la forme suivante :

$$\dot{X} = AX + Bu + w \quad (24)$$

$$y = CX + v \quad (25)$$

- $X \in \mathbb{R}^n$: Vecteur des variables d'état du système
- $\dot{X} = \frac{dX}{dt}$
- $u \in \mathbb{R}^m$: Vecteur des commandes
- $y \in \mathbb{R}^p$: Vecteur des sorties (estimations des données observées)
- $A \in \mathbb{R}^{n \times n}$: Matrice d'état
- $B \in \mathbb{R}^{n \times m}$: Matrice de commande
- $C \in \mathbb{R}^{p \times n}$: Matrice d'observation

- $w \in \mathbb{R}^n$: Bruit du processus, associé à une distribution normale multivariée, de matrice de covariance $Q \in \mathbb{R}^{n \times n}$
- $v \in \mathbb{R}^p$: Bruit d'observation, associé à une distribution normale multivariée, de matrice de covariance $R \in \mathbb{R}^{p \times p}$

L'algorithme se décompose en une phase de prédiction et une phase de mise à jour :

Prédiction :

$$\hat{X}_{k|k-1} = A_k \hat{X}_{k-1|k-1} + B_k u_k \quad (26)$$

$$P_{k|k-1} = A_k P_{k-1|k-1} A_k^T + Q \quad (27)$$

Mise à jour :

$$\tilde{e}_k = D_k - C_k \hat{X}_{k|k-1} \quad (28)$$

$$S_k = C_k P_{k|k-1} C_k^T + R_k \quad (29)$$

$$K_k = P_{k|k-1} C_k^T S_k^{-1} \quad (30)$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \tilde{e}_k \quad (31)$$

$$P_{k|k} = (I - K_k C_k) P_{k|k-1} \quad (32)$$

- $\hat{X}_{k|k-1} \in \mathbb{R}^n$: Vecteur d'état prédit (a priori)
- $\hat{X}_{k|k} \in \mathbb{R}^n$: Vecteur d'état mis à jour (a posteriori)
- $P_{k|k-1} \in \mathbb{R}^{n \times n}$: Matrice d'estimation a priori de la covariance de l'erreur
- $P_{k|k} \in \mathbb{R}^{n \times n}$: Matrice d'estimation a posteriori de la covariance de l'erreur
- $D_k \in \mathbb{R}^p$: Observation ou mesure du processus à l'instant k
- $\tilde{e}_k \in \mathbb{R}^p$: Innovation
- $S_k \in \mathbb{R}^{p \times p}$: Covariance de l'innovation
- $K_k \in \mathbb{R}^{n \times p}$: Gain de Kalman optimal

Lors de la phase de prédiction, sont calculés le vecteur d'état et sa matrice de covariance au pas de temps k connaissant le vecteur d'état et la matrice de covariance au pas de temps $k - 1$. Lors de la phase de mise à jour (ou d'actualisation), sont estimés le vecteur d'état et la matrice de covariance au pas de temps k, sachant la prédiction, et sachant les observations à ce même pas de temps. Ainsi, à chaque pas de temps, il est cherché à estimer au mieux le vecteur d'état d'un système et son incertitude associée, sachant : un modèle, toutes les données recueillies jusqu'à ce pas de temps, et les incertitudes liées au modèle et aux données observées.

La probabilité d'observer une nouvelle donnée D_k sachant le modèle $\eta(\theta)$, et les données précédentes, peut s'écrire selon l'équation suivante :

$$P(D_k|D_{0:k-1}, \eta(\theta)) = \frac{1}{\sqrt{2\pi S_k}} e^{-\frac{1(\hat{e}_k(\theta))^2}{2 S_k}} \quad (33)$$

Selon cette formulation statistique, la fonction de vraisemblance d'un jeu de données prend la forme suivante :

$$L_{Kal} = P(D_0|\eta(\theta)) \prod_{k=1}^n P(D_k|D_{0:k-1}, \eta(\theta)) = P(D_0|\eta(\theta)) \prod_{k=1}^n \frac{1}{\sqrt{2\pi S_k}} e^{-\frac{1(\hat{e}_k(\theta))^2}{2 S_k}} \quad (34)$$

Cette fonction de vraisemblance a été plusieurs fois utilisée dans la littérature pour identifier les paramètres de modèles RC de bâtiments (Madsen et Holst 1995; Raillon et Ghiaus 2018; Rouchier 2018).

Contrairement à la fonction de vraisemblance $L_{i,i,d}$ qui suppose les erreurs indépendantes, cette approche considère les corrélations temporelles entre les erreurs. La probabilité d'observer chaque donnée dépend non seulement du modèle considéré, mais aussi des données précédemment observées : $P(D_k|D_{0:k-1}, \eta(\theta))$. De plus, l'actualisation permet d'identifier les paramètres d'un modèle dynamique en évitant d'éventuels biais engendrés par des phénomènes non modélisés. Ce point est illustré sur la Figure 10 qui représente les erreurs entre des données, et des résultats de simulation, actualisé et non actualisé. Dans le cas de la simulation actualisée, le biais non modélisé par le modèle risque d'avoir un fort impact sur les paramètres identifiés par minimisation des erreurs. L'actualisation calculée par le filtre de Kalman permet de limiter cet impact.

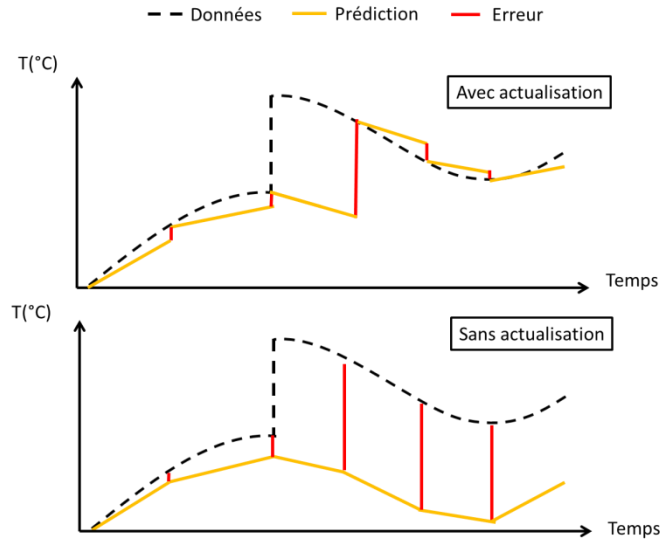


Figure 10 : Illustration de l'effet de l'actualisation sur des données simulées

Les trois fonctions de vraisemblance : $L_{i.i.d}$, L_{GP} , et L_{Kal} sont envisagées pour la calibration des modèles de bâtiments à l'échelle du quartier. Ces fonctions seront testées et comparées sur un cas d'application au chapitre 5.

2.5.2 Calcul de la probabilité a posteriori

Une fois les distributions de probabilité priori définies, les paramètres calibrés et la fonction de vraisemblance sélectionnés, on peut alors chercher à estimer le maximum de la probabilité a posteriori $P(\theta|D)$, on parle de maximum a posteriori (MAP).

Cependant, la calibration par le calcul du maximum a posteriori (MAP) présente plusieurs limites. Le calcul du MAP, consiste à chercher par un algorithme d'optimisation, le point le plus haut (le mode) de la probabilité a posteriori. L'algorithme d'optimisation risque alors de trouver des maximums locaux, non-représentatifs de la distribution. Aussi, le mode ne sera pas représentatif de la distribution dans le cas d'une distribution bimodale, telle que celle représentée en Figure 11. Sur cet exemple, le mode est entouré par un faible volume de probabilité ; bien que ce dernier correspond à la densité de probabilité la plus élevée, la probabilité de tirer aléatoirement une valeur proche de ce maximum sera faible. Enfin, le maximum a posteriori se heurte à « la malédiction de la dimensionnalité ». En haute dimension, la majorité de la masse de probabilité ne se trouve pas proche du mode, ou de la moyenne, mais dans une « coquille » appelée « typical set », qui s'éloigne du mode quand le nombre de dimensions augmente (Martin 2018). Bien que la densité de

probabilité soit plus élevée au niveau du mode, le volume sur lequel la densité est intégrée pour obtenir la masse de probabilité, se répartit principalement loin du mode (Betancourt 2018). Le mode devient alors un mauvais estimateur de la distribution. Ce comportement est illustré en Figure 12 avec une loi normale : en dimension 1, la majorité des valeurs tirées aléatoirement seront proches du mode (au sens de la distance euclidienne); cependant, lorsque le nombre de dimensions augmente, la masse de probabilité s'éloigne de plus en plus du mode.

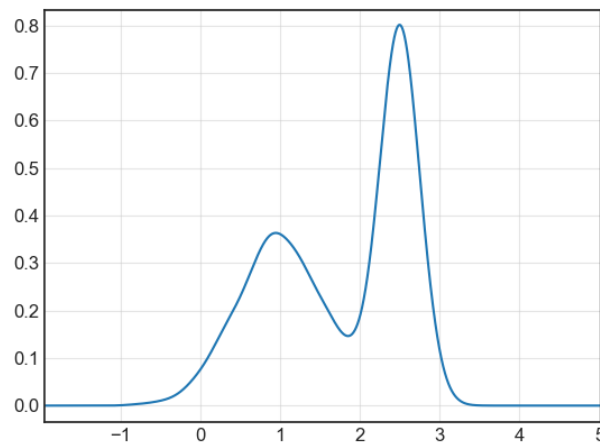


Figure 11 : Exemple d'une distribution bimodale incompatible avec l'utilisation du maximum a posteriori

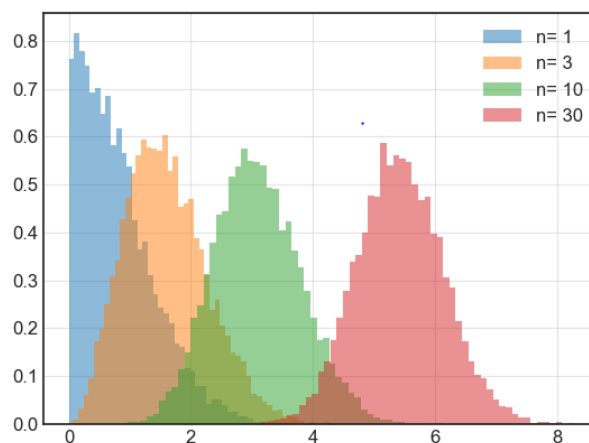


Figure 12 : Distribution de la masse de probabilité de lois normales $N(O_n, I_n)$, en fonction de la distance par rapport au mode

Dans l'approche Bayésienne, il n'est pas retenu une seule théorie, ou un seul modèle calibré (tel que le modèle correspondant au MAP), mais un ensemble de théories ou un ensemble de modèles plus ou moins probables. Dans le cas de la calibration des paramètres d'un modèle thermique de bâtiment, il sera cherché à estimer les distributions de probabilité a posteriori de chacun de ces paramètres. Une

distribution a posteriori est souvent résumée par la moyenne et le plus petit intervalle contenant 95% de la masse de probabilité, appelée intervalle de crédence a posteriori ou HPD (Highest Posterior Density) (Martin 2018) (Figure 13). Cet intervalle peut aussi être utilisé pour décrire une distribution a priori, on parlera alors d'un intervalle de crédence a priori.

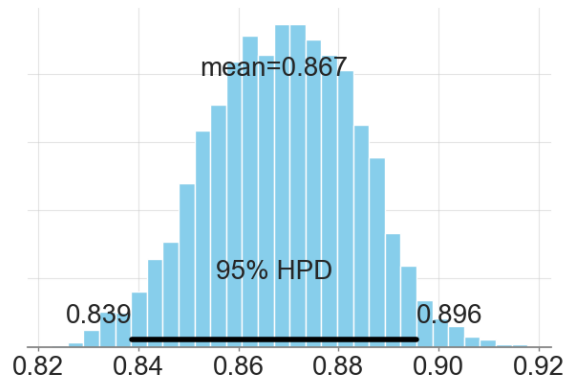


Figure 13 : Illustration d'une distribution de probabilité a posteriori résumée par une moyenne et un intervalle de crédence à posteriori HPD95

2.5.3 Les moteurs d'inférence markoviens

Pour mieux identifier la solution optimale et estimer les incertitudes liées aux résultats, il est nécessaire de parcourir l'ensemble de la distribution de probabilité a posteriori. Cela peut être fait par des techniques d'échantillonnage régulier de l'espace des variables ; mais ces techniques sont peu efficaces, principalement à cause de la « malédiction de la dimensionnalité » (Martin 2018). En effet plus le nombre de dimensions augmente plus la taille de l'espace où se trouve la majorité de la masse de probabilité est faible par rapport à la taille de l'espace total exploré. Beaucoup de ressources informatiques seront alors inutilement mobilisées pour parcourir de nombreux points très peu probables. Pour limiter les temps de calcul, des techniques d'échantillonnage appelées MCMC (Monte-Carlo par chaînes de Markov) ont été développées ; on parle aussi de moteurs d'inférence markoviens (Martin 2018). Le terme Monte-Carlo réfère aux méthodes permettant d'approximer une valeur numérique par échantillonnage aléatoire. Une chaîne de Markov est une séquence d'états, ou la probabilité de se déplacer d'un état à l'autre dépend uniquement de l'état qui précède. Ainsi, les méthodes MCMC sont des algorithmes d'échantillonnage pseudo-aléatoires utilisant des chaînes de Markov. Si une chaîne de Markov vérifie certaines propriétés mathématiques, il peut être montré que la chaîne garantit un échantillonnage dans la bonne distribution (Martin 2018). L'algorithme

d'échantillonnage parcourt chaque région de l'espace des paramètres, proportionnellement aux densités de probabilité. Il existe de nombreuses techniques d'échantillonnage MCMC, les principales sont présentées ci-dessous.

2.5.3.1 Métropolis-Hastings (MH)

L'algorithme d'échantillonnage Métropolis-Hastings (MH) est probablement la méthode MCMC la plus répandue :

- A partir d'un échantillon initial θ_0 :
- Un nouvel échantillon θ_{k+1} est tiré aléatoirement dans une distribution dite instrumentale $q(\theta_{k+1}|\theta_k)$, qui dépend de l'échantillon précédant θ_k .
- La probabilité d'acceptation du nouvel échantillon est calculée selon la formule de Métropolis-Hastings:

$$p_a(\theta_{k+1}|\theta_k) = \min \left(1, \frac{p(\theta_{k+1})q(\theta_k|\theta_{k+1})}{p(\theta_k)q(\theta_{k+1}|\theta_k)} \right) \quad (35)$$

- Si cette probabilité d'acceptation est supérieure à une valeur tirée aléatoirement dans l'intervalle $[0,1]$, alors l'échantillon est accepté, sinon il est rejeté.

L'efficacité de la méthode dépend fortement de la distribution instrumentale $q(\theta_{k+1}|\theta_k)$; si cette distribution est dispersée, la probabilité de rejet sera élevée. À l'inverse, si la distribution est resserrée, la probabilité de rejet sera plus faible mais l'espace sera exploré plus lentement. En pratique, cette loi instrumentale est généralement une loi gaussienne multidimensionnelle dont la variance peut être optimisée de manière à avoir un taux de rejet correspondant à des valeurs connues pour être efficaces (autour de 25 % selon (Martin 2018)).

2.5.3.2 «Slice sampling»

Une alternative à la méthode de Métropolis-Hastings est l'algorithme de «Slice sampling» (Neal 2003). Cette méthode a l'avantage d'être beaucoup moins dépendante de son hyperparamètre que la méthode de Métropolis-Hastings (MacKay 2003; Neal 2003) (la variance de la fonction instrumentale dans le cas de Métropolis-Hastings, et le pas d'échantillonnage w pour l'algorithme de «Slice sampling») :

- A partir d'un échantillon initial θ_0 :
- Figure 14 (a) : Une valeur est tirée aléatoirement, de manière uniforme, dans l'intervalle $[\theta_0, f(\theta_0)]$

- Figure 14 (b) : A partir de cette valeur une tranche « slice » est définie : un intervalle de taille w est positionné aléatoirement autour de θ_0 , puis la tranche est étendue de chaque côté par ajouts successifs de l'intervalle w , jusqu'à ce que l'intervalle soit en dehors de chaque extrémité (droite : qr et gauche : ql) : $f(qr) < f(\theta_0)$ et $f(ql) < f(\theta_0)$.
- Figure 14 (c) : Un nouvel échantillon θ_1 est tiré aléatoirement, de manière uniforme, dans la tranche (entre qr et ql). Si le point tiré est à l'extérieur de la tranche ($f(ql) < f(x_0)$), la tranche est tronquée au niveau de ce point. θ_1 est alors tiré aléatoirement dans cette tranche réduite.

Dans le cas d'une distribution à multiples variables, chaque variable peut être échantillonnée à tour de rôle (Neal 2003).

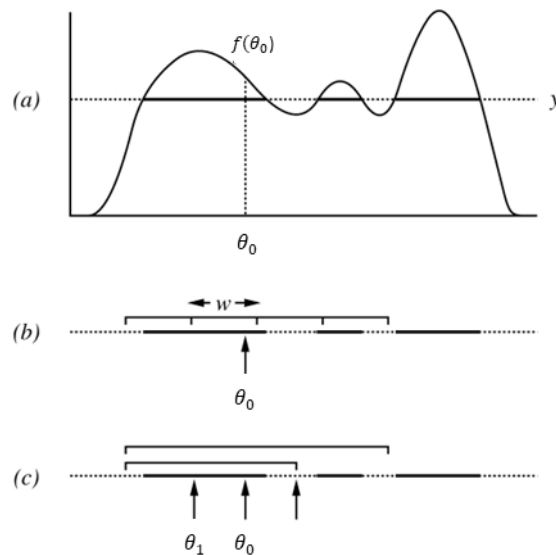


Figure 14 : Illustration de la méthode de « Slice sampling » par (Neal 2003)

2.5.3.3 Méthodes Hamiltoniennes

En haute dimension, dans le but de réduire les temps d'échantillonnage des distributions de probabilité, des techniques dites Hamiltoniennes ont été développées (Betancourt 2018). En effet, les méthodes de Metropolis-Hasting et Slice, sont aussi soumises à la « malédiction de la dimensionnalité », plus le nombre de dimensions augmente plus le nombre de rejets augmente (Betancourt 2018). Les méthodes dites Hamiltoniennes utilisent entre autres des gradients pour mieux suivre le « typical set » ou la masse de probabilité est élevée (Betancourt 2018). Aujourd'hui l'algorithme Hamiltonien « No-U-Turn Sampler » NUTS (Hoffman & Gelman, 2011)

apparaît particulièrement intéressant, notamment car il ne dépend pas du paramétrage réalisé par le modélisateur. Ce dernier a récemment été utilisé par (Chong & Menberg, 2018) pour calibrer des modèles EnergyPlus selon la méthode (Kennedy et O’Hagan 2001).

Cependant, dans cette thèse les échantillonnages de distributions a posteriori seront réalisés par l’algorithme de ‘Slice-sampling’; car, étant plus simple et ne nécessitant pas de calcul de gradient, cet algorithme est plus facile à appliquer aux différents modèles étudiés. L’utilisation d’un algorithme à gradient tel que le NUTS pourra faire l’objet de travaux futurs visant à améliorer la méthode.

2.6 Vérification de la convergence

Lors de la calibration, M chaînes de N échantillons chacune sont exécutées en parallèle sur les différents cœurs du processeur. Les valeurs initiales des différentes chaînes sont tirées au hasard dans les distributions a priori. Les premiers tirages de chaque chaîne sont généralement écartés pour éviter un potentiel biais dû à l’initialisation, c’est l’étape de «burn-in ». Une fois l’inférence bayésienne effectuée, il est pertinent d’évaluer la convergence de l’algorithme d’échantillonnage. En première approche une observation visuelle des chaînes d’échantillonnage peut être faite ; la similarité des chaînes suggère la convergence de l’algorithme. Afin d’évaluer plus précisément la convergence du processus, des indicateurs ont été proposés dans la littérature. Il existe notamment l’indicateur \hat{R} (Gelman et Rubin 1992), calculé à partir des variances inter-chaînes W et intra-chaînes B :

$$\hat{R} = \frac{\frac{N-1}{N}W + \frac{1}{N}B}{W} \quad (36)$$

Avec :

$$B = \frac{N}{M-1} \sum_{m=1}^M (\bar{\theta}^m - \bar{\theta})^2 \quad (37)$$

$$W = \frac{1}{M} \sum_{m=1}^M \frac{1}{N-1} \sum_{n=1}^N (\theta_n^m - \bar{\theta}^m)^2 \quad (38)$$

$$\bar{\theta}^m = \frac{1}{N} \sum_{n=1}^N \theta_n^m \quad (39)$$

$$\bar{\theta} = \frac{1}{M} \sum_{m=1}^M \bar{\theta}^m \quad (40)$$

Plus récemment l'indicateur split- \hat{R} (Gelman et al. 2014) a remplacé le \hat{R} . Le split- \hat{R} est calculé de la même manière que \hat{R} , mais les chaînes d'échantillonnage sont séparées en deux, le nombre de chaînes est alors doublé. La notation \hat{R} est souvent utilisée pour parler du split- \hat{R} , il en sera de même dans la suite de la thèse.

Une bonne convergence de l'algorithme est généralement admise quand $\hat{R} < 1.1$ (Vehtari et al. 2019); ce qui correspond à une variance inter-chaînes B faible comparée à la variance intra-chaînes W . Si la convergence n'est pas satisfaisante le nombre de tirages doit être augmenté.

Il est à noter que (Vehtari et al. 2019) ont mis en avant certaines limites à l'utilisation du \hat{R} pour évaluer la convergence de l'algorithme d'échantillonnage. Il reste cependant performant pour évaluer la convergence des échantillonnages effectués dans cette thèse.

2.7 Calcul de la prédiction a posteriori

Une fois l'inférence Bayésienne effectuée ; il est possible de calculer la probabilité d'observer de nouvelles données \tilde{D} , sachant les données de calibration D , c'est la probabilité de prédiction a posteriori $P(\tilde{D}|D)$:

$$P(\tilde{D}|D) = \int P(\tilde{D}|\theta)P(\theta|D)d\theta \quad (41)$$

Il est ainsi possible d'échantillonner des observations probables \tilde{D} , en procédant ainsi :

- Un échantillon de paramètres θ est tiré aléatoirement dans la loi a posteriori $P(\theta|D)$ déterminée par inférence bayésienne.
- Un échantillon de données \tilde{D} est tiré dans la fonction de vraisemblance $P(\tilde{D}|\theta)$.

Comme pour les distributions a posteriori des paramètres calibrés, les distributions de prédictions a posteriori peuvent être représentées par les plus faibles intervalles comprenant, par exemple, 90 % et 50 % des données échantillonnées, appelés HPD90 et HPD50 (Highest Posterior Density (HPD)). Ces intervalles sont illustrés sur un exemple en Figure 15.

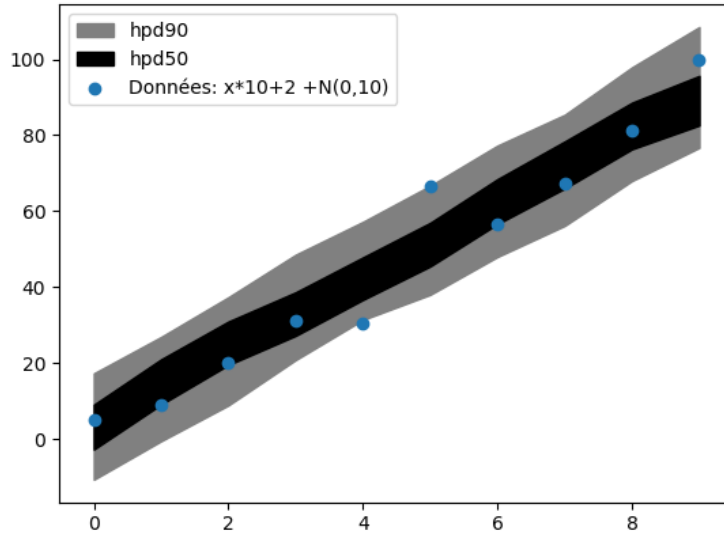


Figure 15 : Exemple d'un échantillonnage de prédictions a posteriori issues de la calibration d'un modèle linéaire sur des données bruitées

2.8 Utilisation d'un métamodèle

L'étude (Kennedy et O'Hagan 2001) propose une méthode de calibration d'un modèle informatique par une approche bayésienne incluant l'utilisation d'un métamodèle (ou émulateur). Ce métamodèle est un modèle de modèle ayant pour but de réduire le temps nécessaire à la calibration lorsque le modèle initial est trop coûteux en calcul. Sur la base de ce travail, (Higdon et al. 2006) présentent un cadre plus général distinguant les cas où il est possible d'exécuter un grand nombre de fois le modèle informatique et les cas où le nombre de simulations possibles est limité par le temps unitaire d'une simulation et nécessite donc l'utilisation d'un métamodèle. Plus récemment, d'autres auteurs ont appliqué cette méthode au cas de la simulation énergétique de bâtiments. (Chong et Menberg 2018) présentent une méthode de calibration bayésienne d'un modèle énergétique de bâtiment basée sur la méthode (Kennedy et O'Hagan 2001).

Dans un premier temps, un nombre réduit de m simulations sont effectuées avec le modèle calibré, pour différents jeux de paramètres connus (x_k^*, θ_k^*) . Un échantillonnage par hypercube latin peut être utilisé pour couvrir correctement l'espace des paramètres. Ces résultats de simulation sont ensuite concaténés avec les données observées au sein d'un même vecteur z :

$$z = \begin{pmatrix} y(x_0, \theta) \\ \vdots \\ y(x_n, \theta) \\ \eta(x_0^*, \theta_0^*) \\ \vdots \\ \vdots \\ \eta(x_m^*, \theta_m^*) \end{pmatrix} \quad (42)$$

Un processus gaussien associé à ce vecteur z est utilisé comme métamodèle. La fonction de covariance est définie à partir des paramètres observables x , et non observables θ :

$$\Sigma_{z,ij} = \frac{1}{\lambda_z} \exp \left(- \sum_{k=1}^p \beta_k^z |x_{ik} - x_{jk}|^\alpha - \sum_{k'=1}^q \beta_{p+k'}^z |\theta_{ik'} - \theta_{jk'}|^\alpha \right) \quad (43)$$

Des matrices de covariance liées aux incertitudes de mesures et à l'inadéquation du modèle, peuvent être ajoutées :

$$\Sigma_l = \Sigma_z + \begin{pmatrix} \Sigma_\varepsilon + \Sigma_\delta & 0 \\ 0 & 0 \end{pmatrix} \quad (44)$$

Les hyperparamètres λ_k et β_k^δ , sont ajoutés au vecteur des paramètres calibrés θ . Le paramètre α est fixé, généralement à 2. La fonction de vraisemblance correspond alors à l'expression suivante :

$$L(z|\theta, \beta_k^z, \lambda_z, \delta, \varepsilon) = |\Sigma_l|^{-1/2} \exp \left(- \frac{1}{2} (z)^T \Sigma_l^{-1} (z) \right) \quad (45)$$

L'utilisation d'un métamodèle augmente fortement le nombre de paramètres à calibrer. Dans le domaine de la thermique du bâtiment, cette méthode est généralement utilisée avec des données mensuelles ou annuelles ; les paramètres dynamiques, tels que la température extérieure, sont donc moyennés sur ces échelles de temps. De plus, les modèles retenus, dits RC, sont peu coûteux en calculs. Cette méthode, incluant l'utilisation d'un métamodèle, est donc écartée.

2.9 Conclusion

Une méthode générale de calibration bayésienne adaptée aux objectifs de cette thèse a été identifiée (Figure 16). Dans un premier temps, le modèle de quartier est utilisé pour estimer les distributions de probabilité a priori des paramètres du modèle. Une analyse de sensibilité de Sobol peut ensuite permettre d'écarter du problème de calibration des paramètres peu influents. Les distributions de probabilité a posteriori

$P(\theta|D)$ sont échantillonnées par l’algorithme «Slice sampling», sélectionné pour sa robustesse et sa praticité. Trois fonctions de vraisemblance ont été envisagées. Notamment un processus gaussien a été proposé afin de tenir compte des corrélations temporelles entre les données. Ces fonctions de vraisemblance seront testées et comparées au chapitre 5. L’indicateur quantitatif \hat{R} , permettant d’évaluer la convergence de l’algorithme a été présenté. Enfin, Le recours à un méta-modèle n’a pas été jugé pertinent étant donné le choix d’une modélisation par modèles simplifiés RC, déjà peu couteux en temps de calcul.

Les algorithmes d’échantillonnage utilisés sont issus de la bibliothèque Python : PYMC3 spécialisée en statistiques bayésiennes et machine Learning². Cette bibliothèque est avec STAN³, l’outil le plus utilisé dans ce domaine. La bibliothèque est open source et dispose d’une importante documentation et communauté d’utilisateurs⁴. Le livre “Bayesian analysis with python” (Martin 2018), lié à cette bibliothèque, a notamment été utilisé.

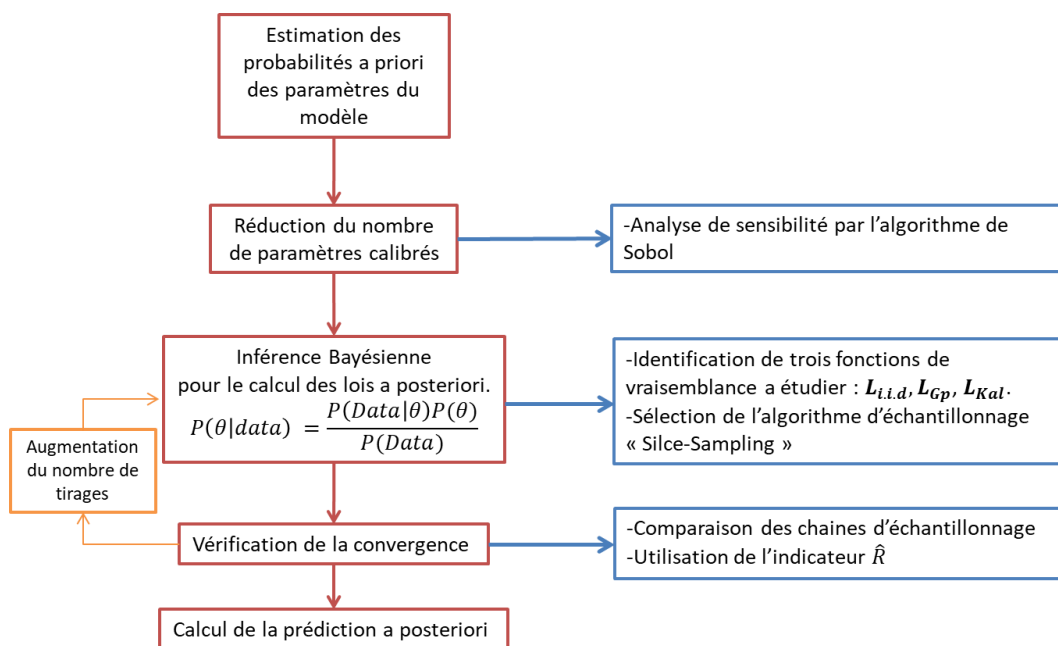


Figure 16 : Méthode globale retenue de calibration bayésienne

² <https://docs.pymc.io/>, <https://en.wikipedia.org/wiki/PyMC3>

³ <https://mc-stan.org/>

⁴ <https://discourse.pymc.io/>

Chapitre 3

Sélection d'un modèle simplifié de
type « RC »

Table des matières

Chapitre 3.....	55
3.1 Introduction.....	58
3.2 Les méthodes d'évaluation d'un modèle	58
3.2.1 Indicateurs d'écart	58
3.2.2 Données de validation	59
3.2.3 Validation croisée.....	60
3.2.4 Critères d'information.....	60
3.2.5 Analyse des résidus	61
3.2.6 Identifiabilité.....	61
3.3 Etat de l'art des modèles RC.....	62
3.3.1 Les études d'identifiabilité	64
3.3.2 Les études concernant le modèle ISO (R5C1)	66
3.4 Présentation de la méthode de construction d'un modèle RC identifiable....	68
3.5 Description du modèle ISO (R5C1)	69
3.5.1 La méthode horaire.....	70
3.5.2 Le calcul de C_m et A_m	72
3.5.3 Formulation mathématique et informatique du modèle	73
3.5.4 Calcul des flux thermiques.....	75
3.5.5 Prise en compte de la ventilation.....	76
3.6 Description du cas d'étude	76
3.7 Etude du modèle ISO sur un cas simplifié mono-zone	78
3.7.1 Estimation physique des paramètres du modèle ISO	78
3.7.2 Comparaison du modèle ISO avec le logiciel TRNSYS et ajout d'une seconde capacité C_a	80

3.7.3	Première calibration du modèle et réduction du nombre de paramètres calibrés	82
3.7.4	Re-paramétrage du modèle ISO_Ca	85
3.8	Etude du modèle ISO_Ca sur le bâtiment complet multi-zones	92
3.9	Etude d'une variante du modèle ISO_Ca, le modèle ISO_CpCa	98
3.10	Conclusion	102

3.1 Introduction

Dans le chapitre 1, une approche de modélisation des besoins thermiques a déjà été présélectionnée : les modèles simplifiés de type « RC ». Cette approche apparaît comme un compromis entre les modèles physiques détaillés et les modèles statistiques. Ils sont facilement paramétrables, les paramètres peuvent être physiquement interprétés, et les coûts de calcul sont faibles. De nombreux modèles RC existent avec des structures et des propriétés différentes. L’objectif de ce chapitre est d’identifier un modèle RC adapté à la calibration bayésienne à l’échelle du quartier et pouvant être intégré à une démarche de suivi et d’amélioration de la performance énergétique d’un quartier.

3.2 Les méthodes d’évaluation d’un modèle

Il existe de nombreuses approches pour valider, diagnostiquer et comparer des modèles, les principales sont présentées ci-dessous.

3.2.1 Indicateurs d’écart

Les indicateurs d’écart sont souvent utilisés pour valider des modèles. Ils permettent de caractériser l’écart entre un jeu de données de référence et une simulation. Il existe plusieurs indicateurs qui apportent des informations différentes. Deux indicateurs fréquemment utilisés sont le biais et le RMSE (Root Mean Square Error). Le biais se définit par la moyenne des écarts à la moyenne, cela permet d’estimer si de manière globale le modèle est centré sur les données, dans le cas contraire, on dit que le modèle présente un biais. Le biais ne permet pas de quantifier l’amplitude des écarts, pour cela, il est possible de faire appel au RMSE qui représente l’écart moyen. Plus le RMSE est élevé plus les écarts de prédictions sont en moyenne élevés.

$$Biais = \frac{\sum_{k=1}^n (\eta_k - D_k)}{n} \quad (46)$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (\eta_k - D_k)^2}{n}} \quad (47)$$

Pour interpréter plus facilement ces indicateurs, il peut être pertinent de normaliser ces derniers en les divisant, par exemple, par la valeur moyenne des observations. Les

indicateurs ainsi obtenus sont notamment utilisés dans (ASHRAE 2002), sous les noms de NMBE (Normalized mean bias error) et CVRMSE (Coefficient of variation of the root mean square error):

$$\text{NMBE} = \frac{100}{\hat{D}} \text{Biais} \quad (48)$$

$$\text{CVRMSE} = \frac{100}{\hat{D}} \text{RMSE} \quad (49)$$

Les indicateurs d'écart ne sont pas suffisants pour évaluer un modèle. Il est toujours possible de construire un modèle qui, correctement calibré, reproduira parfaitement un jeu de données fini, mais malgré cela, le modèle peut ne pas être capable de reproduire un jeu de données supplémentaires différent des données d'apprentissage, c'est le sur-apprentissage. Ainsi, d'autres techniques existent pour évaluer la robustesse d'un modèle.

3.2.2 Données de validation

Pour évaluer un modèle, il est pertinent de séparer les données disponibles en deux : une série d'apprentissage, ou d'entraînement, et une série dite de test, ou de validation. Le modèle est alors calibré sur les données d'apprentissage et l'erreur de prédiction est étudiée à la fois sur les données d'entraînement et les données de validation. Ces résultats sont souvent analysés au regard de la complexité du modèle (Figure 17). Si le modèle est trop simplifié, il ne sera pas capable d'expliquer correctement ni les données d'entraînement ni les données de test, on parle de sous-apprentissage. Cependant pour une trop forte complexité, en essayant de reproduire parfaitement toutes les données d'apprentissage, le modèle dégrade ses performances sur les données de test, c'est le problème du sur-apprentissage. On cherche alors le modèle qui minimise l'erreur sur les données de validation (Hastie et al. 2008).

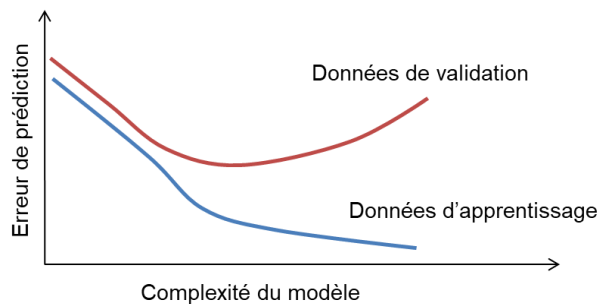


Figure 17 : Illustration des problèmes sur-apprentissage et sous-apprentissage

3.2.3 Validation croisée

L'utilisation de données de validation peut être améliorée en divisant le jeu de données disponibles, non plus en deux, mais en k échantillons. Chacun de ces k échantillons est successivement considéré comme le jeu de données de validation, pendant que le reste des données est utilisé comme jeu d'apprentissage. Les indicateurs d'écart sont calculés pour chacun des cas ainsi définis. Une moyenne et une variance peuvent alors être estimées pour les indicateurs d'écarts sur les données d'entraînement et les données de validation. C'est ce qu'on appelle la validation croisée.

3.2.4 Critères d'information

Les techniques de validation croisée sont très performantes, mais parfois difficiles à mettre en place en pratique. Entraîner un modèle sur de nombreux échantillons peut avoir un coût de calcul élevé. Ainsi des indicateurs, souvent appelés « critères d'information », ayant pour objectif d'évaluer la robustesse des modèles sans passer par des méthodes de validation croisée ont été proposées. Ces critères sont liés au domaine de la théorie de l'information (Martin 2018). Il existe le critère d'information d'Akaike dit AIC (Akaike Information Criterion) présenté en équation (50). Ce critère est composé de deux termes. Le premier terme est appelé la déviance, il caractérise la capacité du modèle à reproduire les données d'apprentissage. θ_{ML} correspond à l'estimation du maximum de vraisemblance, ou au maximum a posteriori (MAP) dans le cas d'une calibration bayésienne. Le second terme correspond au nombre de paramètres calibrés du modèle (q), il permet de pénaliser les modèles complexes. Cette idée correspond au principe du rasoir d'Ockham (ou principe de parcimonie) selon lequel les modèles les plus probables sont les modèles les plus simples. On privilégiera alors le modèle qui présente un AIC le plus faible possible.

$$AIC = -2\log(p(D|\theta_{ML})) + 2q \quad (50)$$

Le critère AIC est pertinent dans le cas d'un apprentissage déterministe, cependant dans le cas d'un apprentissage bayésien, ce critère n'est pas capable de considérer toute l'information contenue dans les courbes de probabilité a posteriori. Ainsi, le critère WAIC (Widely Applicable Information Criterion), présenté en équation (51), a été proposé pour comparer des modèles dans le cas d'un

apprentissage bayésien. À nouveau ce critère se décompose en deux termes. Le premier terme correspond à la moyenne de la log-probabilité a posteriori et caractérise l'écart avec les données d'apprentissage. Les modèles complexes sont à nouveau pénalisés par le second terme défini comme la variance du log probabilité a posteriori.

$$WAIC = -2 \sum_{k=1}^n \log \left(\frac{1}{S} \sum_{s=1}^S p(D_k | \theta^s) \right) + 2 \sum_{k=1}^n (Var_{s=1}^S (\log (p(D_k | \theta^s)))) \quad (51)$$

3.2.5 Analyse des résidus

Pour évaluer un modèle et diagnostiquer des problèmes, il est aussi intéressant d'étudier les résidus de simulation. Les résidus sont les écarts entre les valeurs simulées et les données :

$$e_k = D_k - \eta_k \quad (52)$$

Il est cherché à ce que les résidus du modèle forment une distribution normale centrée en zéro caractéristique d'un bruit blanc, le modèle est alors non biaisé. Les résidus peuvent aussi être tracés en fonction de paramètres endogènes ou exogènes à l'étude, cela pour essayer d'identifier des erreurs de modélisation et des potentielles améliorations du modèle. Enfin, dans le cas de données temporelles, il est possible d'étudier l'autocorrélation des résidus qui mesure la corrélation moyenne entre deux points décalés temporellement d'une valeur fixée τ . Idéalement, la corrélation doit être nulle pour toute valeur de τ . Des tests d'hypothèse peuvent être utilisés pour évaluer statistiquement toutes ces hypothèses de normalité et de corrélations.

3.2.6 Identifiabilité

Dans la littérature, plusieurs définitions de l'identifiabilité existent. Premièrement, l'identifiabilité pratique peut être dissociée de l'identifiabilité structurelle. Un modèle est structurellement identifiable si chacun de ses paramètres a un effet mathématique distinct sur la sortie du modèle (le modèle est inversable). Par exemple, un modèle de bâtiment RC ne sera pas structurellement identifiable s'il possède deux résistances en série sans flux thermique ou capacité intermédiaire. L'identifiabilité structurelle d'un modèle RC peut être vérifiée en s'assurant que l'unicité de la fonction de transfert implique l'unicité des paramètres (Rouchier 2018).

En pratique pour le cas d'un modèle RC simple, l'identifiabilité structurelle peut souvent être évaluée visuellement en observant le schéma RC du modèle.

L'identifiabilité pratique est atteinte lorsqu'un algorithme de calibration, utilisé pour l'identification de paramètres, converge de manière consistante. Un modèle structurellement identifiable n'est pas toujours identifiable en pratique. Notamment, dans certaines conditions d'étude, des paramètres peuvent ne pas être sollicités, les données d'apprentissage ne posséderont alors pas l'information nécessaire à leur identification. Par exemple, un paramètre de capacité thermique de bâtiment ne peut pas être identifié avec des données de consommations journalières. De fortes corrélations entre les paramètres peuvent aussi rendre l'identification pratique difficile. Dans le cas d'une calibration bayésienne, l'identifiabilité pratique est atteinte lorsque l'algorithme converge vers des distributions a posteriori plus resserrées que les distributions a priori.

Enfin, quand les modèles sont construits selon un raisonnement physique, les paramètres identifiés peuvent correspondre à des paramètres physiques connus ; mais ce lien n'est pas toujours évident. C'est notamment le cas pour les modèles dit « boîte grise », tels que les modèles RC, qui simplifient fortement les phénomènes physiques. Un modèle identifiable en pratique, c'est-à-dire pour lequel un algorithme de calibration converge de manière consistante, ne garantit pas la compréhension du lien entre les paramètres identifiés et les propriétés physiques du système étudié.

Vérifier ces différents niveaux d'identifiabilité peut permettre d'améliorer la confiance accordée à un modèle. Sans identifiabilité pratique, un algorithme de calibration aura des difficultés à converger, l'optimisation sera alors difficile et le risque de sur-apprentissage sera élevé. Au-delà de cette identifiabilité pratique, une bonne compréhension du lien entre les paramètres et les propriétés physiques réduit également les risques de sur-apprentissage, et permet de définir des probabilités a priori dans le cas de l'apprentissage bayésien. Dans la suite, c'est donc ce dernier niveau d'identifiabilité qui sera recherché et auquel il sera fait référence.

3.3 Etat de l'art des modèles RC

Dans le Tableau 1, différentes articles évaluant des modèles RC sont inventoriées. Les informations renseignées sont : les modèles étudiés, les cas d'étude, l'utilisation

ou non d'une méthode de calibration « Cali », les objectifs des études, et les méthodes d'évaluation utilisées (en référence au paragraphe précédent).

Publications	Modèles	Cas d'étude	Cali	Objectif	Méthode d'évaluation						
					a	b	c	d	e	f	
(Madsen et Holst 1995)	R3C2	Cas réel : appartement individuel	✓	Commande prédictive Etude de scénarios	✓					✓	✓
(Bacher et Madsen 2011)	R1C1 R2C2 R3C3 R4C4	Cas réel : bâtiment d'expérimentation	✓	Commande prédictive Etude de scénarios Évaluation de performances	✓					✓	
(Reynders et al. 2014)	R1C1 R3C2 R4C3 R6C2 R8C5	Simulation : 2 maisons familiales	✓	Étude de scénarios « Demand Side Management (DSM) »	✓	✓				✓	✓
(Berthou et al. 2014)	R4C2 R6C2 R6C3 R7C3	Simulation : un bâtiment de bureaux	✓	Étude de scénarios « Demand side management (DSM) »	✓	✓				✓	
(Harb et al. 2016)	1R1C 3R2C 4R2C 8R3C	Cas réel : bâtiment de bureaux de 33140 m ²	✓	Étude de scénarios « Demand Side Management (DSM) »	✓					✓	
(Hedegaard et Petersen 2017)	2R2C 4R3C 3C2C 4R3Cw	Simulation : 2 appartements individuels	✓	Évaluation de performances Etude de scénarios	✓	✓				✓	✓
(Raillon et Ghiaus 2018)	R4C3	Cas réel : appartement individuel	✓	Commande prédictive Évaluation de performances	✓	✓			✓	✓	✓
(Brastein et al. 2018)	R3C2	Cas réel : appartement expérimental de 9.4 m ³	✓	Commande prédictive	✓						✓
(Wang et Chen 2019)	R3C2	Cas réel : appartement individuel	✓	Commande prédictive Évaluation de performances Etude de scénarios	✓	✓				✓	
Études concernant le modèle ISO 13790 (AFNOR 2013) :											
(Roujol et al. 2003)	R5C1 (ISO)	Simulation : méthode BESTEST -Cas réel : un bâtiment de bureaux		Logiciel de simulation thermique CONSOCLIM	✓						
(Kokogiannakis et al. 2008)	R5C1 (ISO)	Simulation : un bâtiment de bureaux		Réglementation thermique	✓						
(Michalak 2014)	R5C1 (ISO)	Simulation : un bâtiment résidentiel		Réglementation thermique Étude de scénarios	✓						

(Da Silva et al. 2016)	R5C1 (ISO)	Simulation : méthode BESTEST		Réglementation thermique	✓						
(Vivian et al. 2017)	R5C1 (ISO) R7C2	Simulation : -une chambre -un appartement		Commande prédictive, Étude de scénarios à l'échelle du bâtiment ou du quartier.	✓					✓	
(Michalak 2019)	R5C1 (ISO)	Simulation : méthode BESTEST -un bâtiment résidentiel		Réglementation thermique Étude de scénarios	✓						
(Hedegaard et al. 2019)	R5C2 (ISO) Ajout d'une capacité	Cas réel : un quartier de 159 maisons individuelles	✓	Commande prédictive, Flexibilité à l'échelle quartier	✓	✓					
(Zarella et al. 2020)	R5C1 (ISO)	R5C1 (ISO) campus universitaire		Étude de scénarios à l'échelle du quartier	✓						

Tableau 1 : Inventaire de publications sur les modèles RC ;
« Cali » : utilisation ou non d'une méthode de calibration ;
les méthodes d'évaluation font référence à la numérotation suivante :

a : Indicateurs d'écarts b : Données d'entraînement et données de validation,
c : Validation croisée d : Critères d'information,
e : Analyse de résidus f : Identifiabilité

Il apparaît que les modèles RC sont développés dans plusieurs objectifs : le contrôle prédictif, le calcul d'une consommation de référence (réglementation thermique) et l'évaluation de scénarios d'actions de performance énergétique. Ces modèles sont le plus souvent utilisés avec des données de calibration, mais sont aussi parfois utilisés uniquement avec un paramétrage physique, notamment dans le cas du modèle dit R5C1 issu de la norme ISO 13790 (AFNOR 2013). Les indicateurs d'écart sont systématiquement utilisés pour évaluer ces modèles. À l'inverse, aucune étude de validation croisée n'a été répertoriée. Les autres méthodes d'évaluation de modèles sont utilisées de manière hétérogène.

3.3.1 Les études d'identifiabilité

L'identifiabilité physique des modèles RC apparaît comme un point important et relativement peu étudié dans la littérature. En effet, étant donné que ces modèles simplifient fortement la complexité physique des bâtiments, le lien entre les paramètres des modèles et les caractéristiques physiques des bâtiments n'est pas toujours évident et peut varier significativement avec la structure des modèles. Un des objectifs de cette thèse est la bonne compréhension de ce lien pour permettre de

définir des probabilités a priori qui seront utilisées pour la calibration Bayésienne du modèle. Aussi, cela permettrait de pouvoir construire un modèle de quartier et étudier des scénarios sans disposer de données d'exploitation (température intérieure, consommation d'énergie), ou en disposant de données fortement dégradées. Enfin, si ce lien est maîtrisé, de tels modèles peuvent être utilisés pour caractériser un bâtiment par identification de ses paramètres physiques.

Dans la littérature, lorsque les modèles sont calibrés, les paramètres obtenus sont souvent évalués selon leur plausibilité physique. Dans l'étude (Wang et Chen 2019), le modèle R3C2 est sectionné parmi 4 modèles pour ses capacités prédictives et la plausibilité physique des paramètres calibrés. Dans (Harb et al. 2016), les auteurs définissent pour les paramètres calibrés, des limites hors desquelles l'interprétation physique des paramètres est supposée non-plausible ; 2 modèles sur 4 sont ainsi considérés physiquement plausibles par les auteurs.

Certaines publications étudient en détail l'identifiabilité physique. (Brastein et al. 2018) posent comme condition nécessaire à l'interprétation physique d'un modèle calibré, que l'algorithme de calibration doit converger indépendamment des conditions initiales et des jeux de données d'apprentissage utilisés. La publication étudie la calibration d'un modèle R3C2 et conclut que pour qu'une interprétation physique soit possible les paramètres liés aux résistances thermiques des fenêtres et des portes doivent être fixés.

(Madsen and Holst 1995) étudient le cas d'un appartement de deux zones à température homogène et soumis à une puissance de chauffage dite « pseudo-aléatoire » composée de différentes fréquences de variation. Les paramètres calibrés sont comparés à des estimations physiques de ces derniers. La résistance globale apparaît plutôt bien identifiée (entre 10 % et 20 % d'erreur). La capacité totale affiche cependant environ 30 % d'erreur.

(Raillon et Ghiaus 2018) étudient le cas d'une maison résidentielle instrumentée. La maison est soumise à une puissance de chauffage fortement variable dite ROLBS (Randomly Ordered Logarithmic Binary Sequence). La température considérée est une moyenne des températures de chaque pièce pondérée par les surfaces au sol. Des intervalles de confiance regroupant 95 % des distributions de probabilité identifiées par calibration bayésienne sont comparés aux estimations

physiques des paramètres. Seule la résistance globale est correctement identifiée. Les capacités affichent des ordres de grandeur correcte, mais des écarts de plus de 100 %.

Dans (Reynders et al. 2014), un appartement résidentiel instrumenté est étudié. La température est supposée homogène dans tout le bâtiment. La température de consigne est fixée à 20 °C avec un réduct à 16 °C la nuit. Les paramètres calibrés sont comparés aux valeurs théoriques issues d'un modèle thermique détaillé. Deux modèles, d'ordres 3 et 4, présentent un bon niveau d'identifiabilité du coefficient de transfert thermique global des parois opaques, de la capacité du nœud d'air et d'un coefficient caractérisant les pertes thermiques par ventilation. Pour ces paramètres, les écarts sont inférieurs à 10 %. La capacité globale présente une erreur plus importante (plus de 20 %) et les résultats varient en fonction des données de calibration utilisées.

Enfin, l'étude (Hedegaard and Petersen 2017) compare aussi des paramètres identifiés à des estimations théoriques. Deux appartements modélisés sur EnergyPlus sont utilisés comme cas d'étude. Pour tous les modèles testés, le coefficient de transmission thermique global et la constante de temps liée au nœud d'air sont précisément identifiés (moins de 10 % d'écart). La capacité thermique globale est surestimée de 10 % à 50 % selon les modèles et les données utilisées. Les résultats d'identification des paramètres liés aux pertes thermiques par infiltration et par transmission à travers les parois sont peu cohérents, les écarts varient d'environ -80 % à +50 %.

Il ressort de cette bibliographie que le paramètre caractérisant la résistance thermique globale des bâtiments est généralement assez bien identifié. L'identification des autres paramètres est en revanche moins probante. De plus, les méthodes d'estimation des paramètres théoriques, tels que les capacités thermiques globales, sont souvent mal ou non décrites. En outre, les travaux de recherche menés portent sur un nombre limité de cas d'études spécifiques ne permettant pas de tirer des conclusions plus générales.

3.3.2 Les études concernant le modèle ISO (R5C1)

Aucune étude d'identifiabilité n'a été répertoriée pour le modèle dit R5C1 issu de la norme (AFNOR 2010). Pourtant, ce modèle apparaît particulièrement intéressant, car il dispose d'une méthode complète de paramétrage à partir de caractéristiques

physiques des bâtiments a priori connues. C'est notamment le seul modèle répertorié qui propose une méthode pour estimer le paramètre de capacité thermique globale. Ce modèle est adaptable aux données disponibles et aux spécificités des bâtiments. Différents niveaux de détail sont possibles, des valeurs par défaut de certains paramètres sont proposées, ainsi que des méthodes de prise en compte de cas spécifiques (couplages thermiques entre zones, protections solaires, façades, rideaux, etc...). Dans la suite, il sera souvent fait référence à ce modèle sous le raccourci « modèle ISO ».

Le modèle ISO a été souvent étudié et utilisé dans des travaux de recherche. Il a notamment été validé pour des cas d'études mono-zone selon la méthode dite BESTEST (ASHRAE140) (Michalak 2019; Roujol et al. 2003; Da Silva et al. 2016). Dans ces études, des consommations d'énergie mensuelles ou annuelles, et des températures intérieures maximales, sont comparées à des résultats issus de logiciels SED validés.

A l'échelle d'un bâtiment entier le modèle ISO a aussi été utilisé pour simuler des consommations mensuelles ou annuelles, en supposant une température uniforme dans tout le bâtiment. L'étude (Kokogiannakis et al. 2008) s'intéresse à la classification réglementaire de bâtiments de bureaux selon leurs performances énergétiques. Très peu de différences sont observées entre les résultats issus du modèle ISO et ceux issus de modèles SED validés. (Zarrella et al. 2020) ont simulé avec les modèles ISO et EnergyPlus, les consommations annuelles de chauffage des bâtiments d'un campus. Les écarts obtenus sont compris entre 3 % et 10 %.

Bien que ce modèle ait initialement été conçu pour simuler des consommations globales mensuelles ou annuelles (AFNOR 2013), des études s'intéressent à ses performances dynamiques. (Michalak 2019) compare, pour des cas d'études mono-zones issus de la méthode BESTEST, des courbes de 24 h de consommation horaire, simulées par le modèle ISO et des modèles SED. Les courbes de comparaison présentées apparaissent similaires. (Vivian et al. 2017) étudient le cas d'un bâtiment résidentiel simulé avec le logiciel de SED TRNSYS. Des indicateurs de performance dynamique sont calculés. Les CVRMSE calculés pour la demande de chauffage, en considérant une température de consigne constante et homogène dans tout le bâtiment, sont compris entre 5 % et 20 % selon le climat et l'enveloppe du bâtiment. Les RMSE calculés sur les courbes de température, lorsque le bâtiment est en régime libre (sans chauffage) sont compris entre 0,2 °C et 1 °C. (Roujol et al. 2003) ont

étudié la capacité du modèle à simuler les consommations de climatisation d'un bâtiment de bureaux réel entièrement climatisé à une température homogène. Les données de consommations horaires sont comprises dans l'intervalle de confiance des simulations évalué par échantillonnage de Monte-Carlo.

Le modèle ISO a très peu été étudié avec des méthodes de calibration. (Hedegaard et al. 2019) ont calibré le modèle ISO avec des données mensuelles de consommations d'énergie issues de maisons résidentielles d'un quartier réel. Les températures intérieures des maisons sont supposées fixes. Les indicateurs CVRMSE et NMBE, calculés sur les données de validation sont respectivement à moins de 30 % et 10 %.

Le modèle ISO semble ainsi performant pour simuler des consommations de chauffage sur des cas mono-zones et sur des bâtiments entiers, chauffés à une température homogène. Le cas de bâtiments regroupant plusieurs zones présentant des températures distinctes ne semble pas avoir été étudié. De manière générale, une faible diversité de cas d'étude a été étudiée.

3.4 Présentation de la méthode de construction d'un modèle RC identifiable

L'objectif de ce chapitre est d'identifier un modèle RC adapté à la calibration bayésienne à l'échelle du quartier et pouvant être intégré à une démarche de suivi et d'amélioration de la performance énergétique d'un quartier. L'état de l'art a permis de montrer l'importance de l'identifiabilité des modèles simplifiés, dits RC, dans cette approche, notamment pour la robustesse des modèles et pour pouvoir estimer des probabilités a priori. Ainsi, dans le but d'obtenir un modèle RC identifiable, une étude est réalisée sur la base du modèle ISO issu de la norme (AFNOR 2013) qui dispose d'une méthode complète de paramétrage à partir de caractéristiques physiques des bâtiments. L'identifiabilité du modèle est d'abord évaluée en comparant les résultats de simulation du modèle paramétré à partir des propriétés physiques, aux données de fonctionnement. Puis, une analyse plus fine est menée en étudiant la calibration bayésienne du modèle selon la fonction de vraisemblance $L_{i.i.d}$ rappelée en équation (53). La puissance thermique est considérée en entrée du modèle, les données de calibration sont les températures intérieures moyennes. Des

probabilités a priori uniformes sont considérées en entrée. L'objectif étant que les distributions a posteriori se resserrent autour des estimations physiques.

$$L_{i,i,d} = \prod_{k=1}^n \frac{1}{\sigma_g \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{D_k - \eta_k(\theta, \mathbf{x})}{\sigma_g} \right)^2} \quad (53)$$

- D_k : données de température intérieure moyenne
- $\eta_k(\theta, \mathbf{x})$: simulations de la température intérieure moyenne T_a

Pour une meilleure lisibilité, les paramètres calibrés sont normalisés par les estimations physiques. Ainsi, les paramètres calibrés pourront être considérés comme identifiables s'ils se rapprochent de 1. La calibration est réalisée au pas de temps horaire. Pour chaque calibration, trois chaînes d'échantillonnage de 1800 tirages, dont 300 tirages de « burn-in », sont exécutées en parallèle. La convergence des résultats de calibration est évaluée et des problèmes de corrélation entre les paramètres sont identifiés.

Des variantes du modèle ISO sont proposées pour améliorer son identifiabilité. La robustesse des modèles étudiés est aussi évaluée avec des données de validation. Un bâtiment de bureaux modélisé sur TRNSYS est utilisé comme cas d'étude. Ce choix d'un cas d'étude théorique permet de maîtriser les paramètres physiques du bâtiment et de les comparer plus facilement aux paramètres identifiés par calibration.

3.5 Description du modèle ISO (R5C1)

La norme (AFNOR 2013) présente différentes méthodes de calcul des besoins de chauffage et de refroidissement des bâtiments. Deux méthodes sont proposées : une méthode dite quasi-stationnaire permettant de calculer des consommations mensuelles, et une méthode basée sur un modèle RC permettant de mieux prendre en compte les effets thermiques dynamiques, et de considérer des scénarios horaires de température de consigne, de ventilation, et de protections solaires. L'objectif de cette seconde méthode reste cependant d'estimer des besoins globaux « *les résultats des heures individuelles ne sont pas validés et les valeurs horaires individuelles peuvent comporter des erreurs relatives importantes* » (AFNOR 2013).

3.5.1 La méthode horaire

La méthode horaire simple est généralement présentée sous la forme d'un modèle dit R5C1 (Figure 18). Ce modèle se compose de trois nœuds : un nœud lié à l'air θ_{air} , un nœud lié à la masse thermique du local étudié θ_m , et un nœud θ_s qui représente, selon la norme, « un mélange entre la température de l'air et la température radiante moyenne »

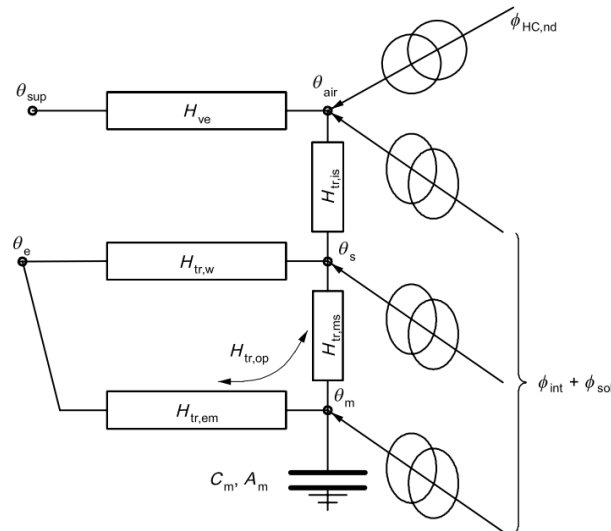


Figure 18 : Modèle R5C1 ISO

Pour mieux comprendre ce modèle, il est intéressant de s'intéresser à la méthode utilisée pour sa construction. Dans la suite du manuscrit, les modèles RC seront représentés selon un formalisme présenté en Figure 19. Les nœuds de températures sont notés Tx , et les résistances thermiques sont nommées selon les noms des nœuds de température auxquels elles sont reliées. De même, les flux thermiques imposés sont notés selon les noms des nœuds auxquels ils s'appliquent précédés du symbole φ . Pour alléger les figures, les noms des résistances et des flux ne seront pas systématiquement marqués.

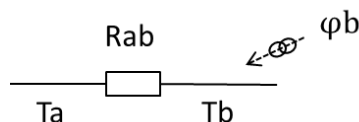


Figure 19 : Illustration de la notation utilisée pour les représentations de modèles RC

Le modèle ISO peut aussi être représenté sous forme du schéma présenté sur la Figure 20. Dans ce schéma, les trois nœuds caractéristiques sont : Ta : la température

de l'air intérieur, T_m : la température des surfaces intérieures liées à la masse inertielle du bâtiment, et T_l : la température des surfaces intérieures des parois légères. Les parois légères peuvent se décomposer en parois légères transparentes (vitrées) et non transparentes (façades rideaux). Les parois légères et les parois lourdes sont reliées au nœud d'air par des résistances convectives R_{ma} et R_{la} ; et sont reliées entre elles par une résistance radiative R_{ml} . De l'autre côté, les parois légères et lourdes sont reliées aux températures de surfaces extérieures par des résistances de conduction puis à la température d'air extérieur par des résistances de convection.

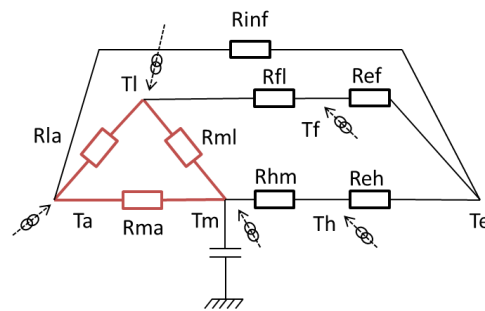


Figure 20 : Représentation détaillée du modèle R5C1 ISO

Pour arriver au modèle R5C1 présenté dans la norme ISO, le schéma ci-dessus doit être modifié. Le schéma présenté en Figure 21 est obtenu en passant d'une représentation triangle à une représentation étoile reliant les nœuds T_a , T_m et T_l . Cette transformation fait apparaître le nœud T_s .

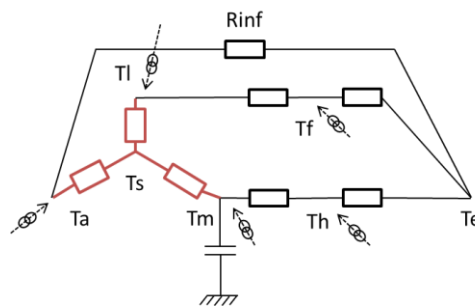


Figure 21 : Représentation étoile du modèle R5C1 ISO

Enfin, ce schéma est à nouveau modifié afin qu'il ne reste des flux thermiques que sur les trois nœuds T_s , T_a et T_m (Figure 22). Cette modification fait intervenir de nouvelles températures T_{em} et T_{el} calculées en fonction des résistances : R_{fl} , R_{ef} , R_{hm} , R_{eh} , et des flux : ϕ_f et ϕ_h . Le schéma ainsi obtenu correspond à celui présenté dans la norme (Figure 18).

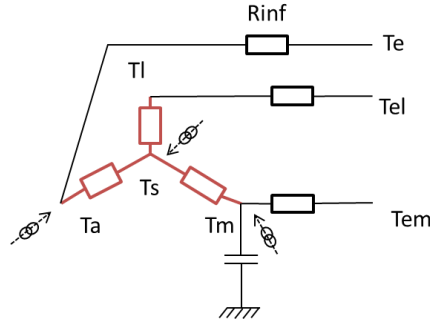


Figure 22 : Représentation finale du modèle R5C1 ISO

3.5.2 Le calcul de Cm et Am

Une particularité de ce modèle est le calcul des paramètres Cm et Am . La capacité Cm caractérise les effets d'inertie thermique liés aux parois lourdes. La surface Am dite « surface massique effective » représente la surface de contact entre les parois lourdes et l'air intérieur. Cm et Am sont définis dans la norme (AFNOR 2013), à partir des capacités surfaciques et des surfaces des parois lourdes du bâtiment : κ_j (kJ/m² K) et A_j (m²):

$$Cm = \Sigma (\kappa_j A_j) \quad (54)$$

$$Am = \frac{Cm^2}{\Sigma(\kappa_j^2 A_j)} \quad (55)$$

Les capacités surfaciques sont issues de la norme ISO 13786 (AFNOR 2010). Cette norme propose deux méthodes de calcul présentées en annexe 1 : une méthode détaillée et une méthode simplifiée. Il est à noter que ces méthodes sont peu décrites et peu argumentées dans la littérature. Il est possible d'obtenir des valeurs de Cm et Am avec des méthodes forfaitaires selon la classe d'inertie des bâtiments. Des valeurs par défaut sont proposées dans la norme (AFNOR 2013). Aussi, il existe dans la réglementation thermique (CSTB 2017) une méthode consistant à calculer des points d'inertie associés aux bâtiments en fonction des compositions de chaque paroi. Le total des points permet de déterminer la classe d'inertie et les valeurs de Am et Cm .

Une capacité supplémentaire peut être ajoutée à Cm pour prendre en compte la masse thermique liée au volume d'air et au mobilier (CSTB 2017).

Même si une seule capacité ne peut représenter à elle seule toutes les dynamiques d'un bâtiment en fonction des différentes sollicitations thermiques, il semble pertinent d'essayer de caractériser de manière globale les propriétés dynamiques d'un bâtiment.

Cette approche simplifiée est choisie dans le cadre de cette thèse. Dans la suite il sera entre autres cherché à évaluer la confiance que l'on peut accorder à cette méthode de calcul de la capacité Cm .

3.5.3 Formulation mathématique et informatique du modèle

Les modèles RC étudiés dans ce manuscrit sont codés selon une méthode illustrée dans ce paragraphe sur le cas du modèle ISO (Figure 20).

Premièrement, les équations de bilans d'énergie (loi des nœuds) sont écrites pour chaque nœud de température inconnue :

$$0 = \frac{Tm - Ta}{Rma} + \frac{Tl - Ta}{Rla} + \frac{Te - Ta}{Rinf} + \varphi a \quad (56)$$

$$0 = \frac{Ta - Tl}{Rla} + \frac{Tm - Tl}{Rml} + \frac{Tf - Tl}{Rfl} + \varphi l \quad (57)$$

$$0 = \frac{Tl - Tf}{Rfl} + \frac{Te - Tf}{Ref} + \varphi f \quad (58)$$

$$0 = \frac{Tm - Th}{Rhm} + \frac{Te - Th}{Reh} + \varphi h \quad (59)$$

$$Cm \frac{dTm}{dt} = \frac{Ta - Tm}{Rma} + \frac{Tl - Tm}{Rml} + \frac{Th - Tm}{Rhm} + \varphi m \quad (60)$$

Le système d'équations est ensuite écrit sous forme matricielle. Le vecteur X regroupe les températures d'état du système. Le vecteur X_2 contient les températures non incluses dans le vecteur d'état. Les sollicitations sont regroupées dans un vecteur u .

$$\dot{X} = A_1 X + B_1 X_2 + C_1 u \quad (61)$$

$$0 = A_2 X + B_2 X_2 + C_2 u \quad (62)$$

Dans le cas du modèle ISO :

$$X = (Tm) \quad X_2 = \begin{pmatrix} Ta \\ Tl \\ Tf \\ Th \end{pmatrix} \quad u = \begin{pmatrix} \varphi m \\ \varphi a \\ \varphi l \\ \varphi f \\ \varphi h \\ Te \end{pmatrix} \quad (63)$$

$$A_1 = \left(\frac{-1}{Cm} * \left(\frac{1}{Rma} + \frac{1}{Rml} + \frac{1}{Rhm} \right) \right) \quad (64)$$

$$B_1 = \begin{pmatrix} \frac{1}{Cm * Rma} & \frac{1}{Cm * Rml} & 0 & \frac{1}{Cm * Rhm} \end{pmatrix} \quad (65)$$

$$C_1 = \begin{pmatrix} \frac{1}{Cm} & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (66)$$

$$A_2 = \begin{pmatrix} \frac{1}{Rma} \\ \frac{1}{Rml} \\ 0 \\ \frac{1}{Rhm} \end{pmatrix} \quad (67)$$

$$B_2 = \begin{pmatrix} \frac{-1}{Rma} + \frac{-1}{Rla} + \frac{-1}{Rinf} & \frac{1}{Rla} & 0 & 0 \\ \frac{1}{Rla} & \frac{-1}{Rla} + \frac{-1}{Rml} + \frac{-1}{Rfl} & \frac{1}{Rfl} & 0 \\ 0 & \frac{1}{Rfl} & \frac{-1}{Rfl} + \frac{-1}{Ref} & 0 \\ 0 & 0 & 0 & \frac{-1}{Rhm} + \frac{-1}{Reh} \end{pmatrix} \quad (68)$$

$$C_2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \frac{1}{Rinf} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \frac{1}{Ref} \\ 0 & 0 & 0 & 0 & 1 & \frac{1}{Reh} \end{pmatrix} \quad (69)$$

Une manipulation matricielle permet ensuite de mettre le problème sous forme d'un modèle d'état continu :

$$X_2 = B_2^{inv}(-A_2X - C_2u) \quad (70)$$

$$\dot{X} = A_1X + B_1B_2^{inv}(-A_2X - C_2u) + C_1u \quad (71)$$

$$\dot{X} = (A_1 - B_1B_2^{inv}A_2)X + (C_1 - B_1B_2^{inv}C_2)u \quad (72)$$

$$\dot{X} = A_cX + B_cu \quad \text{avec} \quad A_c = (A_1 - B_1B_2^{inv}A_2) \quad \text{et} \quad B_c = (C_1 - B_1B_2^{inv}C_2) \quad (73)$$

Ce modèle d'état peut alors être discrétisé :

$$X^{k+1} = A_dX^k + B_du^k \quad \text{avec} \quad A_d = e^{A_c\Delta t} \quad \text{et} \quad B_d = \int_0^{\Delta t} e^{A_c\tau} d\tau B_c \quad (74)$$

Le système entier est alors modélisé par les deux équations suivantes :

$$X^{k+1} = A_dX^k + B_du^k \quad (75)$$

$$X_2^k = B_2^{inv}(-A_2 X^k - C_2 u^k) \quad (76)$$

Cette méthode permet de coder rapidement sous forme d'un modèle d'état les modèles RC étudiés dans cette thèse.

Afin de rendre la calibration plus rapide les modèles initialement codés en python ont été adaptés au langage Cython, puis compilés en langage C, avec la librairie Cython⁵. Le langage Cython est très proche du langage Python mais nécessite notamment de déclarer toutes les variables utilisées. Une attention particulière est portée sur la boucle temporelle du modèle qui représente la majeure partie du temps de calcul. Il s'agit de veiller à ce que le code compilé en C, ne fasse pas appel à des fonctions Python dans cette boucle temporelle ; car la communication entre les langages Python et C représente un important goulet d'étranglement pour la vitesse de calcul. Les temps de calcul sont ainsi divisés par un facteur situé entre 50 et 100. Avec un processeur i7-6820HQ de fréquence 2.7 GHz, et une mémoire RAM de 32GB, le modèle prend environ 0,003 s pour simuler 3 semaines de fonctionnement avec un pas de temps de 15mins.

3.5.4 Calcul des flux thermiques

Les flux φ_a , φ_m et φ_l sont calculés à partir des flux physiques suivants : la puissance du système de chauffage φ_{heat} , les gains internes φ_{int} (regroupant les apports dus à l'éclairage et les appareils électriques φ_{elec} , et les apports dus aux métabolismes des occupants φ_{meta}), et le flux solaire pénétrant dans le bâtiment $\varphi_{sol_{in}}$.

Les flux φ_{heat} et φ_{int} se divisent en une part radiative et une part convective selon les ratios $r_{cv_{int}}$ et $r_{cv_{heat}}$. Les parts convectives sont intégrées au flux φ_a . Les parts radiatives se divisent entre les flux φ_m et φ_l au prorata des surfaces des parois lourdes et légères : Am et Al . Dans les cas d'étude considérés dans cette thèse, les parois légères correspondent aux surfaces vitrées ($Al = S_{win}$). 90 % du flux solaire $\varphi_{sol_{in}}$ est dirigé sur le nœud des parois lourdes Tm , 10 % est supposé être directement transmis à l'air.

$$\varphi_a = \varphi_{int} r_{cv_{int}} + \varphi_{heat} r_{cv_{heat}} + \varphi_{sol_{in}} 0,1 \quad (77)$$

⁵ <https://cython.org/>

$$\varphi m = \varphi_{\text{int}}(1 - r_{\text{cvint}}) \frac{Am}{Am + Al} + \varphi_{\text{heat}}(1 - r_{\text{cvheat}}) \frac{Am}{Am + Al} + \varphi_{\text{solin}} 0,9 \quad (78)$$

$$\varphi l = \varphi_{\text{int}}(1 - r_{\text{cvint}}) \frac{Al}{Am + Al} + \varphi_{\text{heat}}(1 - r_{\text{cvheat}}) \frac{Al}{Am + Al} \quad (79)$$

Les flux φh et φf arrivent sur les surfaces externes du bâtiment, respectivement sur les parois opaques et sur les fenêtres. Ces flux considèrent les échanges radiatifs avec la voûte céleste φ_{sky_h} et φ_{sky_f} , définis selon les équations ci-dessous, en fonction de la température de ciel T_{sky} , et des températures de surface Th et Tf .

$$\varphi_{\text{sky}_h} = \sigma \varepsilon (S_{\text{wall}} * 0,5 + S_{\text{roof}}) (T_{\text{sky}}^4 - Th^4) \quad (80)$$

$$\varphi_{\text{sky}_f} = \sigma \varepsilon S_{\text{win}} 0,5 (T_{\text{sky}}^4 - Tf^4) \quad (81)$$

Le flux φh considère en plus le rayonnement solaire arrivant sur les parois extérieures opaques $\varphi_{\text{solopaque}}$:

$$\varphi h = \varphi_{\text{solopaque}} + \varphi_{\text{sky}_h} \quad (82)$$

$$\varphi f = \varphi_{\text{sky}_f} \quad (83)$$

Les températures Th et Tf sont estimées à chaque itération selon l'équation (76), et utilisées pour calculer les flux φh et φf .

3.5.5 Prise en compte de la ventilation

Les pertes thermiques liées à la ventilation sont prises en compte en ajoutant un terme $Rmeca$ à la résistance d'infiltration $Rinf$. La ventilation étant variable, les matrices B2 et C2 doivent être mises à jour à chaque pas de temps.

3.6 Description du cas d'étude

Le cas d'étude retenu est un bâtiment type de bureaux, issu de l'étude (Alessandrini et al. 2003). Il a été modélisé dans le logiciel de simulation énergétique détaillé TRNSYS⁶. Le bâtiment est décrit en détail dans le Tableau 2.

La période de calibration comprend trois semaines d'hiver (du 15 janvier au 5 février). Deux périodes de validation sont aussi considérées : les trois semaines restantes du mois de février, et le mois de mars. Pour les données de validation, la consigne de température est légèrement modifiée. Cette dernière est fixée à 23 °C

⁶ <http://www.trnsys.com/>

entre 3 h et 8 h du matin de sorte qu'une partie de la pointe de consommation du matin soit décalée. Cela est fait pour accentuer les différences entre les données d'apprentissage et les données de validation, et mieux éprouver la robustesse du modèle.

Géométrie	<p>Bâtiment de 3 étages identiques : Vue du dessus d'un étage :</p>
Murs	0,012 m de plâtre 0,040 m de laine de verre (isolation intérieure) 0,130 m de béton $U_{wall} : 0,93 \text{ W/m}^2.K$ (sans les résistances surfaciques)
Fenêtres	Double vitrage $U_{win} : 7,5 \text{ W/m}^2.K$ (sans les résistances surfaciques)
Toit	0,012 m de plâtre 0,06 m de laine de verre (isolation intérieure) 0,10 m de béton $U_{roof} : 0,65 \text{ W/m}^2.K$ (sans les résistances surfaciques)
Plancher intermédiaire	0,003 m de liège 0,10 m de béton $U : 7,57 \text{ W/m}^2.K$
Plancher bas	0,003 m de liège 0,10 m de béton U adiabatique
Parois intérieures	0,06 m de plâtre $U : 12 \text{ W/m}^2.K$
Capacité de la zone	$29\ 000 \text{ kJ/K}$ ($8 * \text{Air volume}$)
Capacités thermiques des parois	Plancher : $262\ 807 \text{ kJ/K}$ (0,1 m de béton) Murs : $143\ 591 \text{ kJ/K}$ (0,13 m de béton) Toit : $262\ 807 \text{ kJ/K}$ (0,1 m de béton)
Convection intérieure	$h_i = 3 \text{ W/m}^2.K$
Convection extérieure	$h_e = 18 \text{ W/m}^2.K$
Occupation	Zones de bureaux : $9 \text{ m}^2/\text{pers}$, 6h-18h, Zone de réunion : 80 W/pers , 9h-10h & 14h-17h
Gain d'appareils électriques	Zones de bureau : $r_{app} = 15 \text{ W/m}^2$: 00h-7h: 18% 7h-10h: linéaire 10h-17h : 80 % 17h-20h: linéaire 20h-24h: 18 %
Gain d'éclairage	$r_{light} = 15 \text{ W/m}^2$: fonction de l'occupation et de l'ensoleillement.
Infiltration	$0,2 V_{ch}/h$ (pendant les périodes de ventilation, le débit d'infiltration est supposé nul)
Ventilation	Zones de bureaux : $0,9 V_{ch_bureaux}/h$, fonction de l'occupation Zone de réunion : $0,9 V_{ch_reunion}/h$, fonction de l'occupation

Gain solaire	Transmittance : 0.77, Absorbance : 0.8 Meteonorm® (Paris)
Puissance de chauffage maximum	$\varphi_{max} = 300 \text{ kW}$
Température de consigne	Toutes les zones sont chauffées selon une même consigne de température : 21°C en période d'occupation, avec un réduit à 15°C pendant les nuits et les week-ends. Les relances sont programmées 3h avant l'occupation du mardi au vendredi et 6h avant l'occupation le lundi.

Tableau 2 : Description du cas d'étude : un bâtiment de bureau type.

3.7 Etude du modèle ISO sur un cas simplifié mono-zone

Dans un premier temps, le modèle ISO est étudié uniquement sur l'étage intermédiaire du bâtiment de bureaux présenté ci-dessus, dans lequel les parois intérieures ont été retirées afin d'obtenir un cas d'étude mono-zone. Cela permet dans un premier temps de ne pas inclure d'éventuels problèmes liés au cas multizones, tels que des écarts de températures entre zones. Les planchers intermédiaires inférieurs et supérieurs sont modélisés avec une condition limite adiabatique au niveau de leurs faces extérieures.

3.7.1 Estimation physique des paramètres du modèle ISO

Am et Cm sont calculés en utilisant chacune des trois méthodes présentées précédemment : la méthode détaillée, la méthode simplifiée et la méthode à points, les résultats sont affichés en Tableau 3 et Tableau 4. Les trois méthodes donnent des résultats similaires.

Parois	Surfaces m ²	Détaillée		Simplifiée		A
		Km	Cm	Km	Cm	points
		kJ/m ² K	kJ/K	kJ/ m ² K	kJ/K	Points
Plancher haut	1242	78,7	97770	76,4	94889	6
Parois extérieures	315	24,3	8310	20,6	6492	0
Plancher bas	1242	60,3	74855	65,6	81475	4
Total			180935		182856	10
Prise en compte d'une valeur forfaitaire pour le mobilier			205775		207696	11

Tableau 3 : Résumé du calcul de Cm selon les méthodes : détaillée, simplifiée et à point, pour le cas de l'étage de bureau mono-zone

	Détaillée	Simplifiée	A points
Cm (kJ/K)	205775	207696	204930
Am (m²)	2634	2627	3105
Am/S_{ch}	2,1	2,0	2,5
Cm/S_{ch} (kJ/ m² K)	166	167	165

Tableau 4 : Résultats du calcul de Cm et Am selon les méthodes : détaillée, simplifiée et à point, pour le cas de l'étage de bureau mono-zone

Les autres paramètres du modèle ISO sont calculés selon les formules suivantes :

Parois opaques	Convection extérieure	$Reh = \frac{1}{he(S_{wall} + S_{roof})} = \frac{1}{18((69 + 18) * 2 * 3 - 208)} = 0,18 \cdot 10^{-3} \quad (84)$
	Conduction	$Rhm = \frac{1}{U_{wall}(S_{wall} + S_{roof})} = \frac{1}{0,93((69 + 18) * 2 * 3 - 208)} = 3,4 \cdot 10^{-3} \quad (85)$
	Convection intérieure	$Rma = \frac{1}{hi Am} = \frac{1}{3 * 2634} = 0,13 \cdot 10^{-3} \quad (86)$
Fenêtres	Convection extérieure	$Ref = \frac{1}{he S_{win}} = \frac{1}{18 * 208} = 0,27 \cdot 10^{-3} \quad (87)$
	Conduction	$Rfl = \frac{1}{U_{win} S_{win}} = \frac{1}{7,5 * 208} = 0,64 \cdot 10^{-3} \quad (88)$
	Convection intérieure	$Rla = \frac{1}{hi S_{win}} = \frac{1}{3 * 208} = 1,6 \cdot 10^{-3} \quad (89)$
Infiltration	$Rinf = \frac{1}{Uinf} = \frac{3600}{\rho_{air} c_{air} \dot{V}_{inf} V_{ch}} = \frac{3600}{1,2 * 1004 * 0,2 * (69 * 18 * 3)} = 4 \cdot 10^{-3} \quad (90)$	
résistance radiative entre les parois lourdes et les parois légères	$Rml = \frac{1}{4\varepsilon\sigma S_{win}\theta_0^3} = \frac{1}{4 * 0,8 * (5,6703 * 10^{-8}) * 208 * 2634 * 293^3} = 1,05 \cdot 10^{-3} \quad (91)$	

Tableau 5 : Estimation physique des résistances du modèle ISO (en K/W)

Remarque :

Dans la norme, la résistance radiative entre les parois lourdes de surface Am , et les parois légères de surface S_{win} , est calculée selon l'équation suivante :

$$Rml = \frac{S_{win} + Am}{4\varepsilon\sigma S_{win}Am\theta_0^3} = \frac{208 + 2634}{4 * 0,8 * (5,6703 * 10^{-8}) * 208 * 2634 * 293^3} = 1,1 \cdot 10^{-3} \text{ (K/W)} \quad (92)$$

Cette formulation suppose que les surfaces S_{win} et Am sont équivalentes et que chacune des surfaces rayonne autant sur elle-même que sur l'autre surface. Cette hypothèse n'est pas adaptée au cas où les parois légères sont uniquement les fenêtres. Dans ce cas, il est préférable de faire l'hypothèse d'une petite surface dans un grand environnement. La résistance radiative dépend alors uniquement de la surface des fenêtres S_{win} (voir équation (91)).

3.7.2 Comparaison du modèle ISO avec le logiciel TRNSYS et ajout d'une seconde capacité Ca

Les températures intérieures issues des modèles TRNSYS et ISO sont comparées en Figure 23. Les sollicitations sont prises à l'identique dans les deux modèles : $T_{sky}, Te, \varphi_{sol_{opaque}}, \varphi_{sol_{in}}, \varphi_{int}, \varphi_{heat}$. Aussi dans les deux modèles, le chauffage est supposé être purement convectif ($r_{cv_{heat}} = 1$) ; les gains internes sont modélisés comme à 50 % convectifs ($r_{cv_{int}} = 0,5$).

La résistance globale paramétrée dans le modèle ISO apparaît légèrement trop faible, la courbe de température du modèle ISO est légèrement au-dessus de celle du modèle TRNSYS. Cela correspond à une erreur d'environ 5 % sur la résistance globale du modèle. Il est difficile de réduire cette erreur, les principales sources d'imprécision sont les résistances et les flux radiatifs. Aussi, des pics de température apparaissent au moment des relances de chauffage sur la courbe correspondant au modèle ISO. Ce phénomène provient du fait que le modèle ISO ne possède pas de capacité sur le nœud d'air. Ce phénomène a déjà été remarqué dans plusieurs études telles que (Hedegaard et al. 2019). Dans la suite, une capacité supplémentaire « Ca » sera donc ajoutée au nœud d'air ; on parlera alors du modèle ISO_ Ca présenté en Figure 24. La capacité Ca comprend les masses thermiques de l'air et du mobilier. Ces masses thermiques sont alors retirées de la capacité Cm . Les températures intérieures issues des modèles TRNSYS et ISO_ ca sont comparées en Figure 23. Les variations rapides de température au moment des relances de chauffage ont disparu.

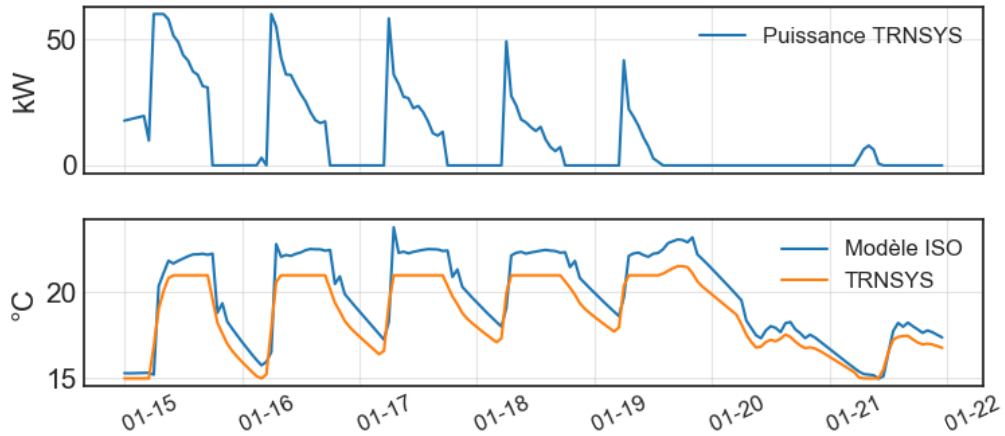


Figure 23 : Comparaison entre les températures simulées par les modèles TRNSYS et ISO sur le cas d'étude mono-zone

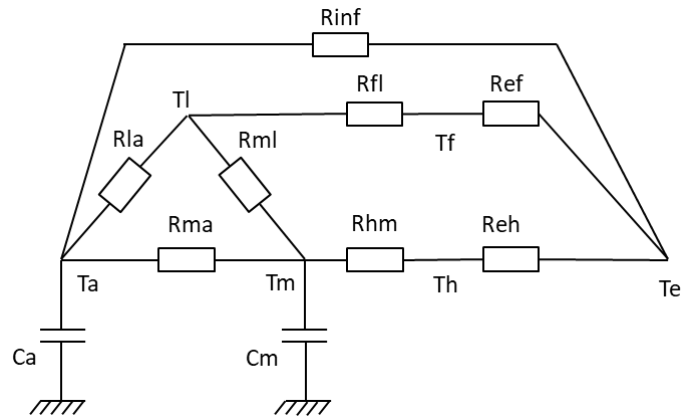


Figure 24 : Représentation du modèle ISO_Ca

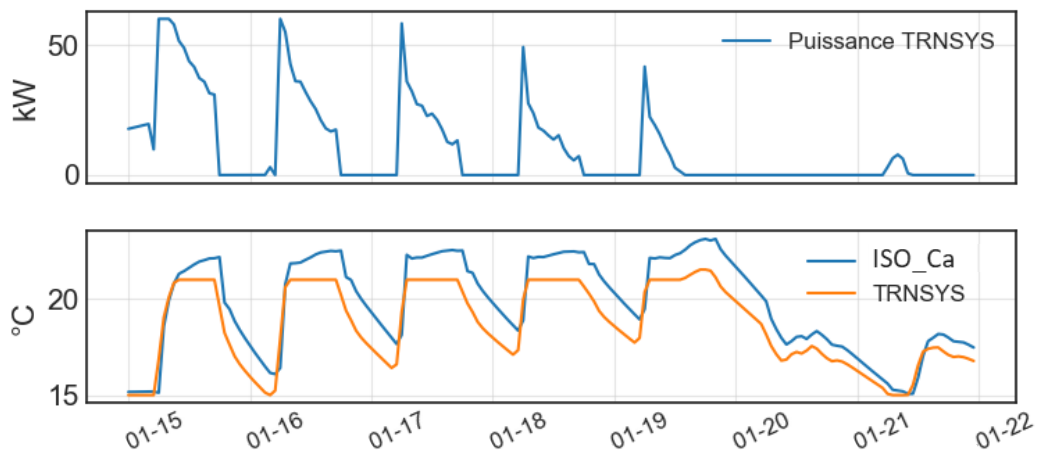


Figure 25 : Comparaison entre les températures simulées par les modèles TRNSYS et ISO_Ca sur le cas d'étude mono-zone

3.7.3 Première calibration du modèle et réduction du nombre de paramètres calibrés

Les chaînes d'échantillonnage issues de la calibration des 10 paramètres du modèle ISO_Ca sont présentées en Figure 26. Il apparaît difficile de calibrer l'ensemble des paramètres du modèle, seules les chaînes liées aux paramètres *Ram*, *Cm* et *Reh* convergent. Pour les autres paramètres, de fortes corrélations empêchent la convergence des chaînes. Ces corrélations peuvent aussi s'observer sur une matrice des graphes de corrélation (Figure 27). Il est alors pertinent d'écarter les paramètres peu influents et d'éviter de calibrer des paramètres fortement corrélés. Pour cela, une analyse de sensibilité de Sobol (Sobol 2001) a été menée, les résultats sont présentés en Figure 28. Pour cette analyse de Sobol, les paramètres normalisés varient de 0,5 à 2.

Les résistances *Rhm*, *Rfl*, et *Rinf* sont retenues comme paramètres à calibrer, car elles ressortent fortement de l'analyse de Sobol et appartiennent chacune à une branche distincte du modèle ISO, ce qui limite les risques de corrélation. Le paramètre *Cm* est également sélectionné pour la calibration ; cette capacité est non corrélée avec les autres paramètres (Figure 26, Figure 27), elle régit la dynamique du modèle. Enfin, la résistance *Rma* est aussi sélectionnée malgré un faible indice de Sobol. Une importante incertitude est liée à cette résistance qui est calculée à partir de la surface *Am* issue de la méthode (AFNOR 2010). De plus, cette résistance n'est pas corrélée avec les autres paramètres (Figure 26, Figure 27) ; en liant la capacité des parois lourdes au nœud d'air, elle joue aussi un rôle sur la dynamique du modèle.

Remarque : Dans l'étude (Hedegaard et al. 2019), les paramètres *Cm* et *Am* sont liés entre eux par une relation linéaire par morceaux. Une telle relation peut être construite à partir des jeux de paramètres proposés dans la norme (AFNOR 2013). Ainsi, seule la capacité *Cm* est calibrée, et la surface *Am* est calculée en fonction de *Cm*. Cette approche permet de réduire le nombre de paramètres calibrés ; elle n'a cependant pas été retenue dans cette thèse, car *Cm* et *Am* par l'intermédiaire de *Rma*, apparaissent décolérés (Figure 26 et Figure 27) et peuvent donc être calculés distinctement.

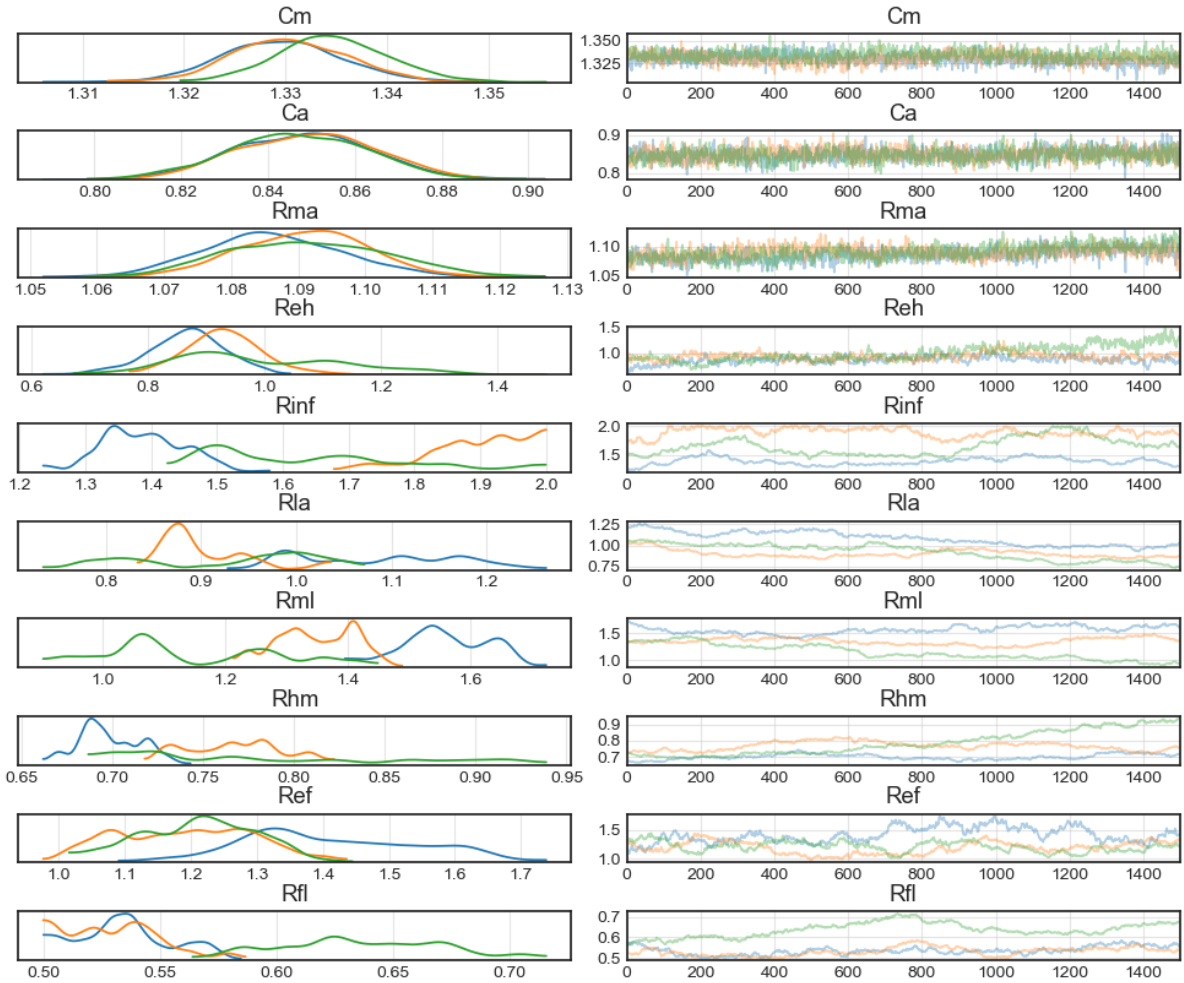


Figure 26 : Chaînes d'échantillonnage de la fonction de vraisemblance du modèle ISO_Ca appliqué au cas mono-zone

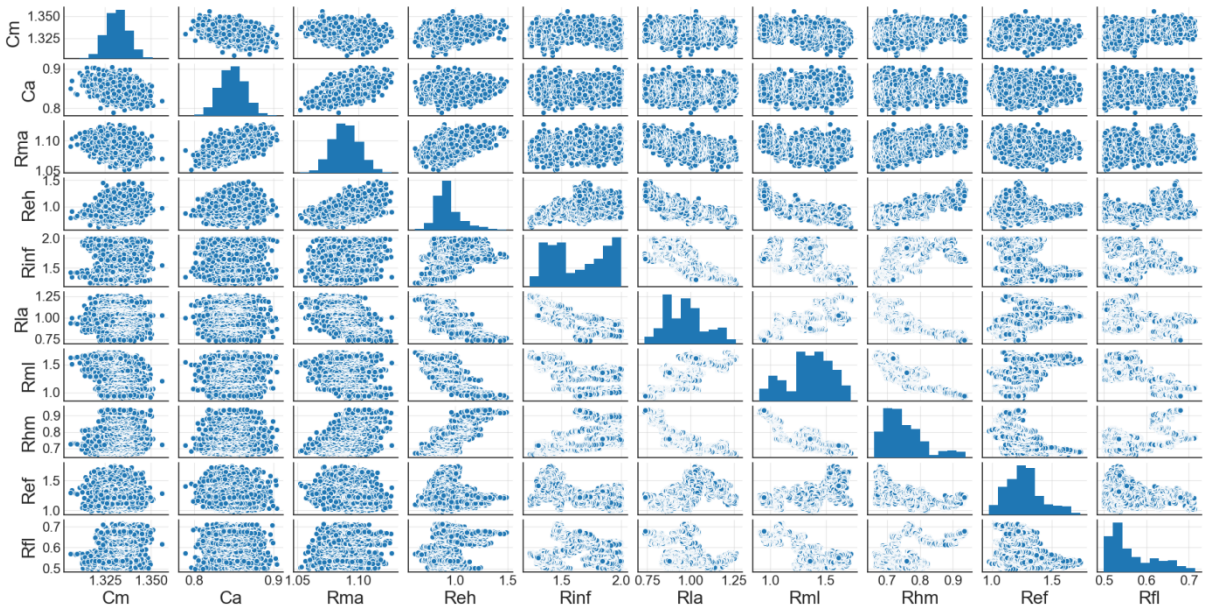


Figure 27 : Graphique des corrélations des valeurs échantillonnées du modèle ISO_Ca appliqué au cas mono-zone

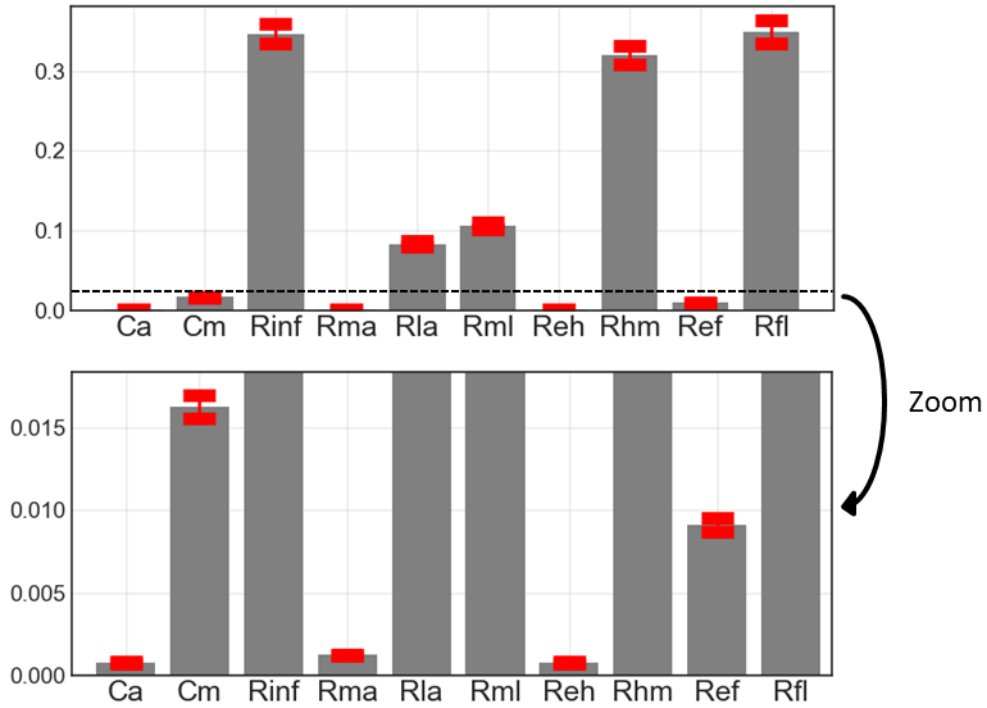


Figure 28 : Coefficients totaux de l'analyse de sensibilité de Sobol, sur les dix paramètres du modèle ISO_Ca, avec barres d'incertitude à 95 %

Le modèle est à nouveau calibré en considérant cette fois-ci uniquement les 5 paramètres sélectionnés. Les distributions échantillonnées présentées en Figure 29, montrent que les résistances *Rfl*, *Rhm* et *Rinf* restent fortement corrélées entre elles ; l'algorithme de calibration ne converge toujours pas.

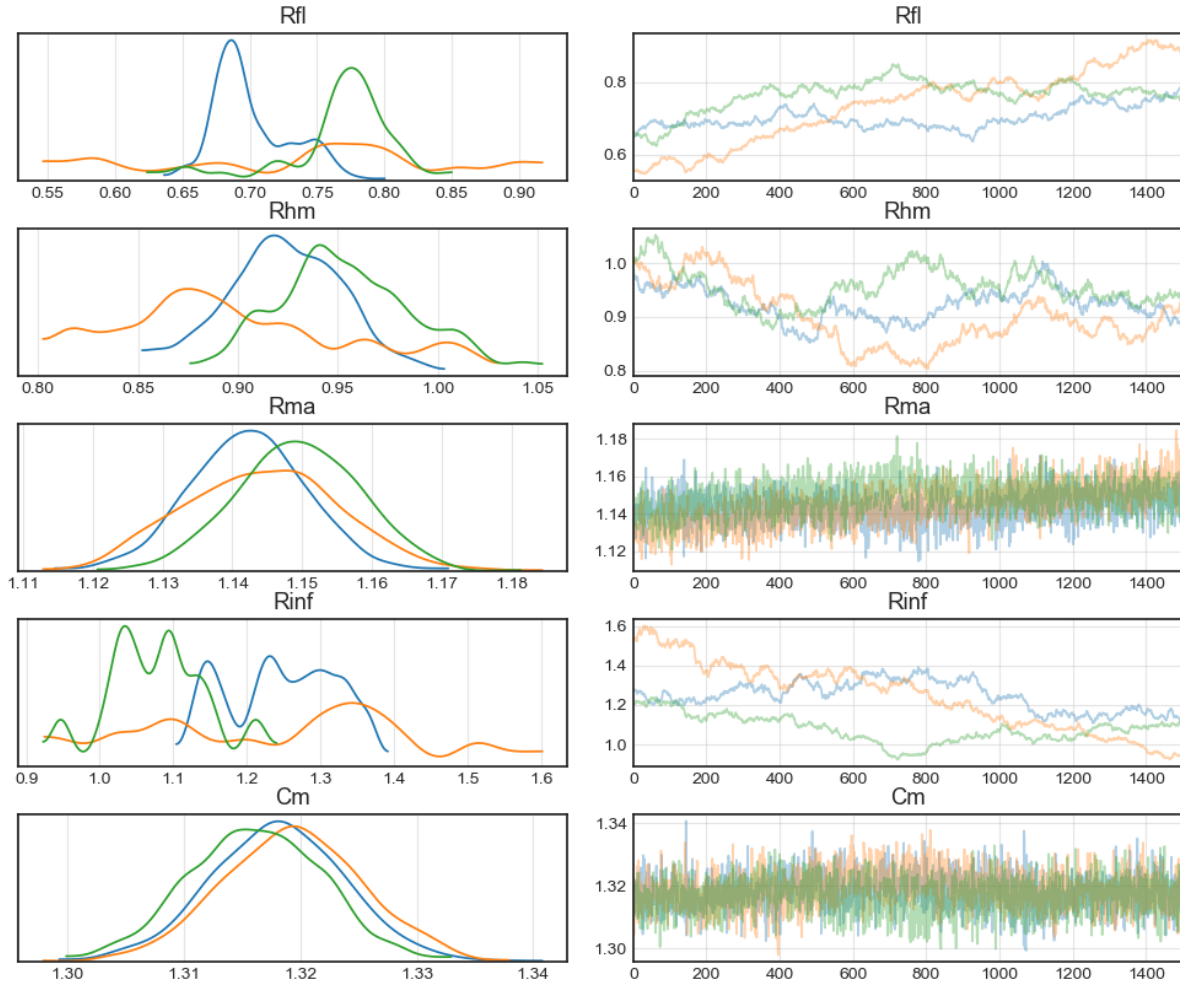


Figure 29 : Chaînes d'échantillonnage de la fonction de vraisemblance du modèle ISO_Ca appliqué au cas mono-zone, avec un jeu de paramètres réduit

3.7.4 Re-paramétrage du modèle ISO_Ca

Afin de résoudre le problème de corrélation mis en avant ci-dessus, un re-paramétrage du modèle est proposé. Les trois résistances $Rinf$, Rhm et Rfl se répartissent sur les trois branches du modèle et sont toutes fortement corrélées à la résistance globale du modèle. L'idée est de calibrer le flux thermique total entre l'intérieur et l'extérieur du bâtiment et la répartition de ce flux entre les différentes branches. Cependant, une difficulté est que le modèle ne correspond pas strictement à un flux qui se répartit en trois branches distinctes entre un point chaud et un point froid. Les branches lourdes et légères sont liées entre elles par la résistance radiative Rml , et le point le plus chaud peut varier entre le nœud d'air Ta et le nœud intérieur des parois lourdes Tm . Ainsi, deux méthodes ont été identifiées pour définir trois branches distinctes. La première considère la configuration étoile du modèle ISO; le

point S joue le rôle du point central à partir duquel le flux se répartit en trois branches distinctes (Figure 30 (a)). Le nœud S n'est pas toujours le point le plus chaud du système ; ce dernier variant entre les nœuds S , A et M , en fonction des sollicitations thermiques. Cependant les écarts entre les températures de ces trois nœuds sont faibles (moins de 1 degré). Ainsi cette approche conserve son sens physique malgré cette approximation. Dans la seconde approche la configuration triangle est conservée. La branche liée aux parois vitrées inclut alors les résistances Rml et Rla en parallèle (Figure 30 (b)). Ces deux approches ont été testées et donnent des résultats très proches. Dans la suite la seconde approche est retenue, car elle est plus facilement adaptable à d'autres variantes du modèle ISO_Ca.

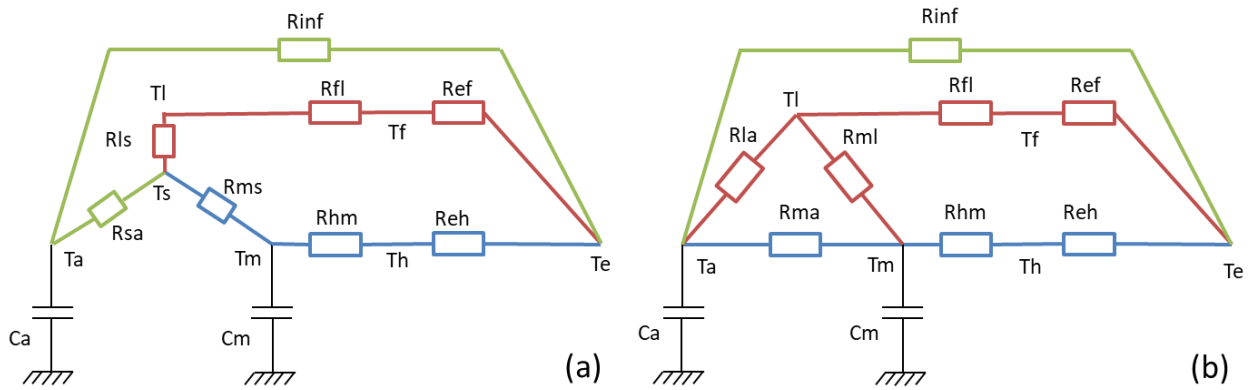


Figure 30 : Illustration de la division du modèle ISO_Ca en trois branches

Selon cette méthode, un coefficient de transfert global Ug peut être défini comme la somme des trois coefficients de transfert de chacune des branches définies ci-dessus.

$$U_{inf} = \frac{1}{Rinf} \quad (93)$$

$$U_{win} = \frac{1}{Ref + Rfl + \frac{1}{\frac{1}{Rml} + \frac{1}{Rla}}} \quad (94)$$

$$U_m = \frac{1}{Reh + Rhm + Rma} \quad (95)$$

$$Ug = U_{inf} + U_{win} + U_m \quad (96)$$

À partir de ces coefficients de transfert un nouveau jeu de paramètre est défini : Rg , $Rate_g$ et $Rate_l$. Rg correspond à la résistance globale du bâtiment. $Rate_m$ caractérise la répartition du flux total entre la branche liée à la capacité Cm et les

deux branches dépourvues de capacité. $Rate_l$ caractérise la répartition du flux entre les deux branches dépourvues de capacité thermique :

$$Rg = \frac{1}{Ug} \quad (97)$$

$$U_{win_inf} = U_{inf} + U_{win} \quad (98)$$

$$Rate_m = \frac{U_m}{Ug} \quad (99)$$

$$Rate_l = \frac{U_{inf}}{U_{win_inf}} \quad (100)$$

Il est donc proposé de calibrer les paramètres Rg , $Rate_g$ et $Rate_l$, à la place des résistances $Rinf$, Rel et Rhm qui seront alors calibrées indirectement selon les équations suivantes :

$$Rhm = \frac{1}{Um} - Rma - Reh = \frac{1}{Rate_m/Rg} - Rma - Reh \quad (101)$$

$$Rinf = \frac{1}{Uinf} = \frac{1}{U_{win_inf} * Rate_l} = \frac{1}{\frac{(1 - Rate_m)}{Rg} * Rate_l} \quad (102)$$

$$\begin{aligned} Rfl &= \frac{1}{Uwin} - Ref - \frac{1}{\frac{1}{Rml} + \frac{1}{Rla}} \\ &= \frac{1}{U_{win_inf} * (1 - Rate_l)} - Ref - \frac{1}{\frac{1}{Rml} + \frac{1}{Rla}} \\ &= \frac{1}{\frac{(1 - Rate_m)}{Rg} * (1 - Rate_l)} - Ref - \frac{1}{\frac{1}{Rml} + \frac{1}{Rla}} \end{aligned} \quad (103)$$

Une limite à cette approche est le risque de faire apparaître des résistances négatives lors de la calibration. Ce risque est faible pour le cas des parois lourdes, car la résistance de conduction Rhm est généralement bien supérieure aux résistances de convection et radiation. Cependant, pour la branche des parois vitrées, la résistance Rfl peut être très faible, notamment dans le cas d'une fenêtre à double vitrage; il y a alors un risque élevé de voir apparaître des valeurs négatives de résistances. Pour remédier à ce problème il est nécessaire de calibrer d'autres résistances de la branche. Le choix devrait se porter sur la résistance la plus influente et la plus incertaine. Cependant, aucune résistance ne se démarque particulièrement. Il n'est pas possible

non plus de calibrer toutes les résistances, car on aurait des problèmes de convergence. Une solution restante est de calibrer la surface de vitrage S_{win} qui influe proportionnellement sur l'ensemble des résistances de la branche. Le coefficient de transfert thermique surfacique de la branche des parois vitrées u_{win} est alors supposé fixe, et est initialement calculé selon l'équation :

$$u_{win} = \frac{1}{Ref * S_{win} + Rfl * S_{win} + \frac{1}{\frac{1}{Rml * S_{win}} + \frac{1}{Rla * S_{win}}}} \quad (104)$$

S_{win} est ainsi indirectement calibré à partir des trois paramètres définis précédemment Rg , $Rate_m$ et $Rate_l$:

$$S_{win} = \frac{U_{win}}{u_{win}} = \frac{\frac{1}{Rg} * (1 - Rate_m) * (1 - Rate_l)}{u_{win}} \quad (105)$$

Les résistances Ref , Rfl , Rml et Rla sont ensuite recalculées en fonction de S_{win} selon les équations (87), (88), (89), et (5).

Lorsque Rfl est suffisamment élevée (dans le cas de double ou triple vitrage), les deux approches présentées, consistant à calibrer indirectement Rfl ou S_{win} , présentent des résultats similaires. Pour généraliser la méthode, c'est donc l'approche passant par la calibration indirecte de la surface de vitrage S_{win} qui est retenue dans la suite.

Une calibration est réalisée avec le nouveau jeu de paramètre : Am , Cm , Rg , $Rate_m$ et $Rate_l$ (Figure 31, Figure 32). Les chaînes d'échantillonnage convergent, les paramètres ne sont plus corrélés. En plus de leur convergence, les valeurs identifiées sont proches de 1, tous les paramètres calibrés sont proches des paramètres physiques estimés. Le modèle apparaît ainsi identifiable. Aussi les courbes de température simulées par le modèle calibré sont très proches de celles issues du modèle TRNSYS (Figure 33, Figure 34, Figure 35). En Tableau 6, des RMSE sont donnés pour la période de calibration (janvier) et les deux mois suivants. Le modèle sera, dans la suite, confronté à d'autres cas d'études.

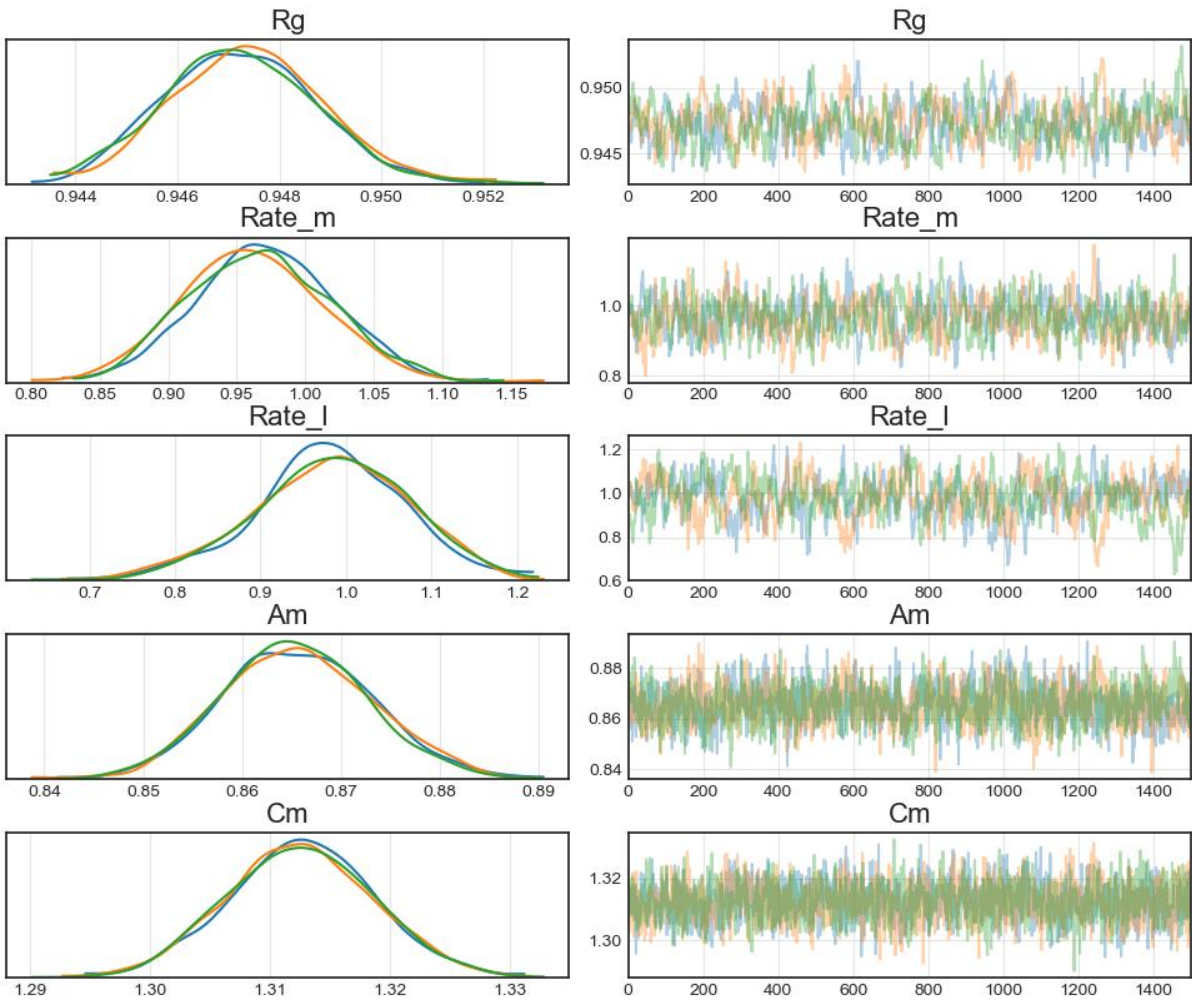


Figure 31 : Chaînes d'échantillonnage de la fonction de vraisemblance du modèle ISO_Ca appliqué au cas mono-zone, avec re-paramétrage

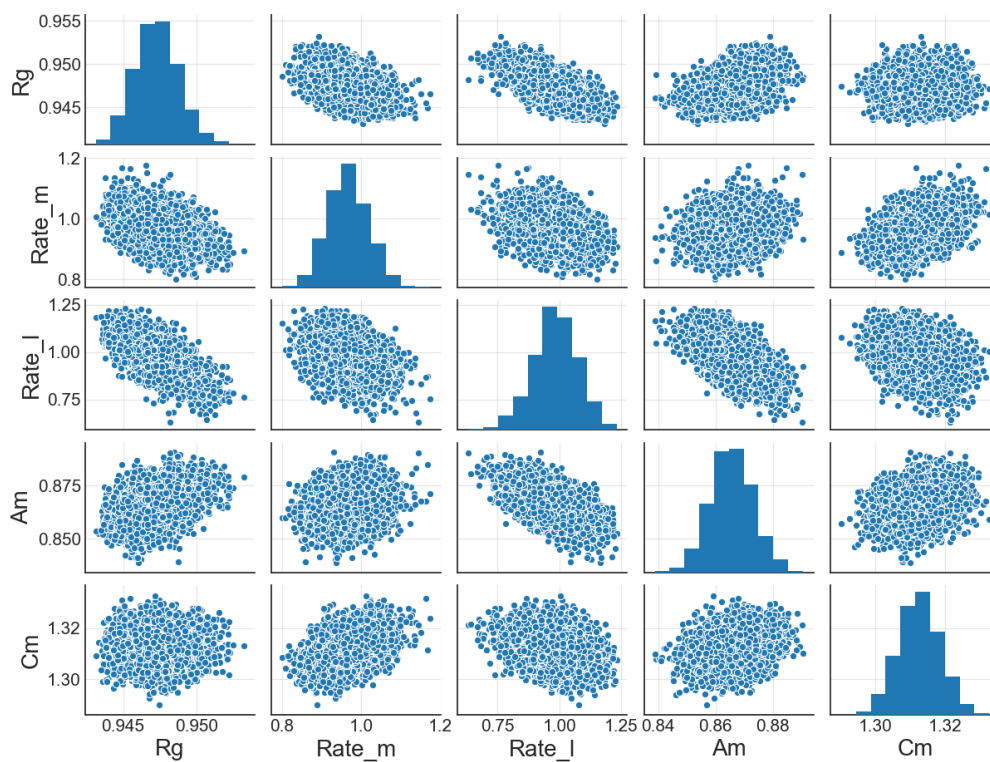


Figure 32 : Graphique des corrélations des valeurs échantillonnées du modèle ISO_Ca appliqué au cas mono-zone, avec re-paramétrage

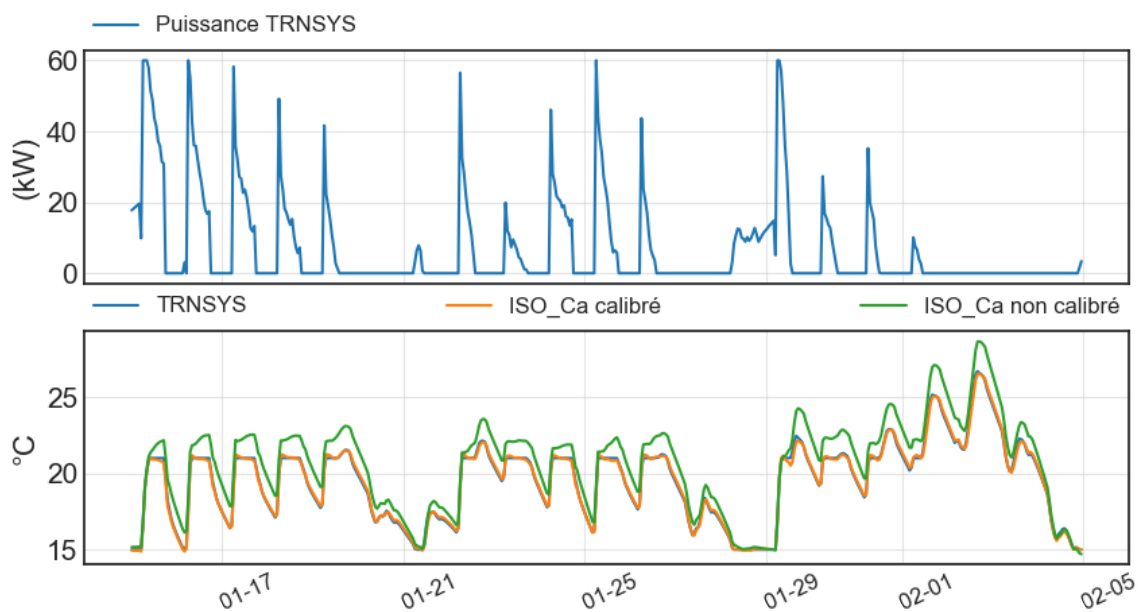


Figure 33 : Comparaison des courbes de températures simulées par les modèle TRNSYS et ISO_Ca, sur le jeu de données de calibration, dans le cas d'étude mono-zone (les courbes bleue et orange se superposent)

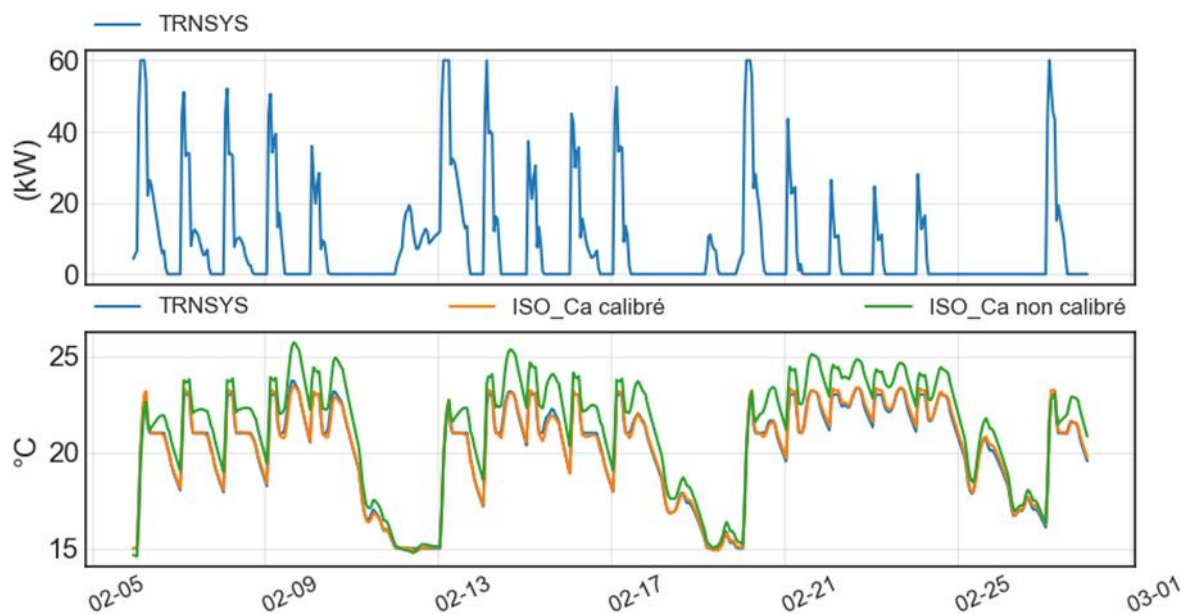


Figure 34 : Comparaison des courbes de températures simulées par les modèle TRNSYS et ISO_Ca, sur le jeu de données de validation de février, dans le cas d'étude mono-zone (les courbes bleue et orange se superposent)

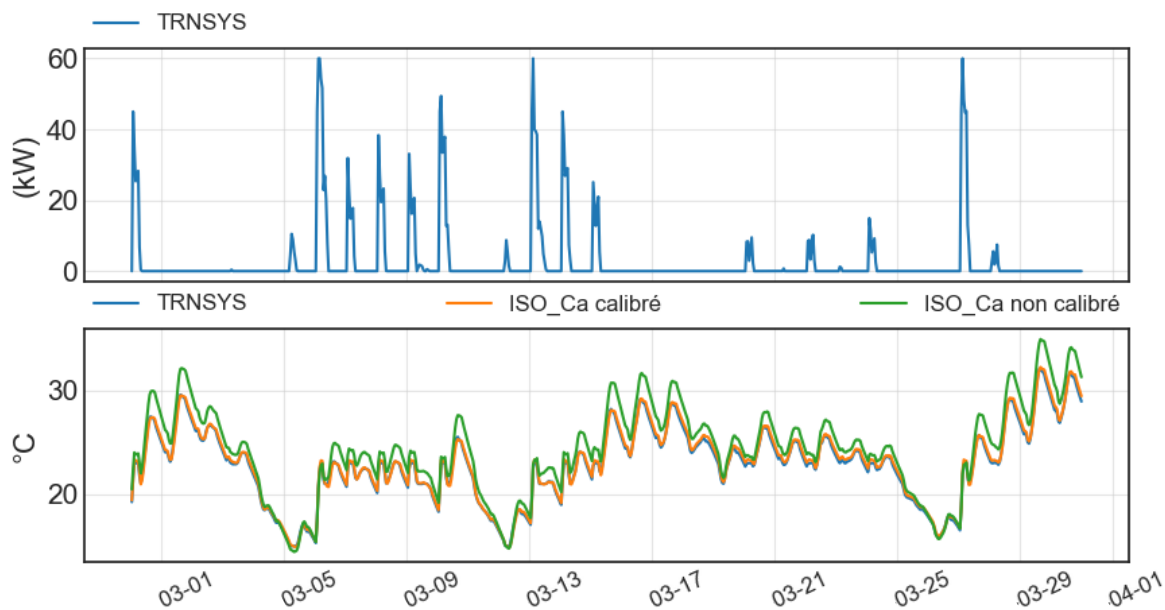


Figure 35 : Comparaison des courbes de températures simulées par les modèle TRNSYS et ISO_Ca, sur le jeu de données de validation de mars, dans le cas d'étude mono-zone (les courbes bleue et orange se superposent)

	Janvier (période de calibration)	Février	Mars
RMSE (°C)	0,15	0,17	0,21

Tableau 6 : RMSE entre les températures simulées par les modèle TRNSYS et ISO_Ca calibré, sur le cas d'étude mono-zone

Enfin, une nouvelle analyse de sensibilité de Sobol a été réalisée en considérant ces nouveaux paramètres (Figure 36). La résistance globale R_g est très largement dominante par rapport aux autres paramètres. Ce paramètre sera donc certainement plus facile à identifier, mais les erreurs sur ce dernier auront un fort impact sur les écarts de prédiction. Les paramètres les moins influents sont $Rate_m$ et $Rate_l$, ses paramètres pourront donc s'avérer difficiles à identifier. Cependant les influences relatives des paramètres C_m , A_m , $Rate_m$ et $Rate_l$ sont susceptibles de varier selon les cas d'étude.

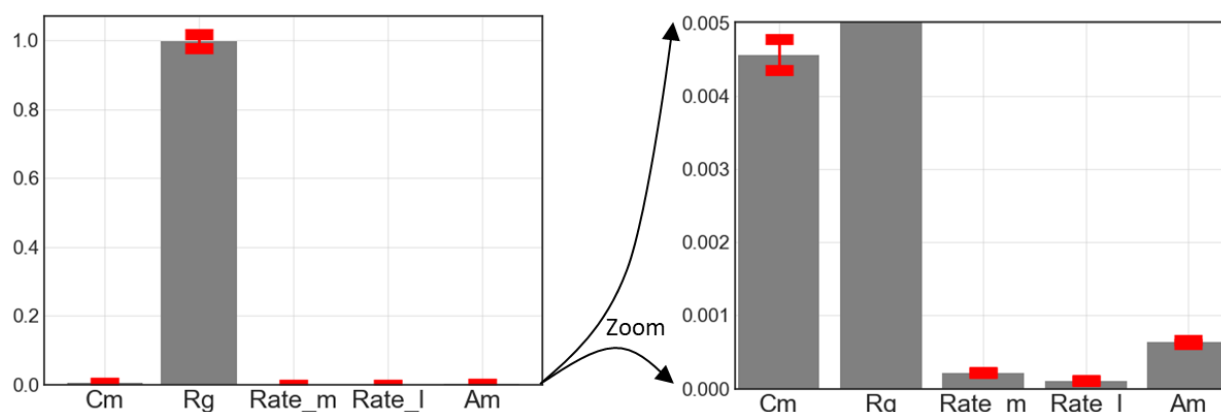


Figure 36 : Analyse de Sobol du modèle ISO_Ca re-paramétré

3.8 Etude du modèle ISO_Ca sur le bâtiment complet multi-zones

Le modèle est maintenant étudié sur le bâtiment entier présenté dans le paragraphe 3.6. La calibration du modèle RC nécessite une estimation globale de la température intérieure du bâtiment. La valeur retenue est la moyenne des températures de chaque zone, pondérée pas les surfaces au sol. En plus du cas de base, 11 scénarios sont étudiés pour tester la robustesse du modèle dans différentes configurations. Ces scénarios sont présentés dans le Tableau 7 sous forme de modifications apportées au cas de base.

Nom	Description																																																													
Cas de base (Base)	Ce scénario correspond au cas d'étude décrit en paragraphe 3.6 (un bâtiment de bureaux de 3 étages avec 5 zones par étage). Toutes les zones sont chauffées selon une même consigne de température. Les heures d'occupation sont identiques dans chaque étage. La consigne est fixée à 21°C avec un réduit à 15°C pendant les nuits et les week-ends. La température varie peu entre les zones, des écarts maximaux de 4°C sont observés les week-ends entre les sanitaires du dernier étage orientés au nord, et la salle de réunion du premier étage orientée au sud. L'inertie et isolation peuvent être qualifiées de moyennes.																																																													
Isolation extérieure (Isoext)	Pour les parois extérieures et le toit, les couches de laine de verre sont placées à l'extérieur des couches de béton, cela augmente l'inertie globale du bâtiment.																																																													
Inertie légère : (Léger)	La masse thermique du bâtiment est réduite. Les parois intérieures composées de 6 cm de plâtre sont remplacées par des parois constituées de deux plaques de plâtre de 1 cm chacune, séparées par 10 cm d'air (ce type de parois permet de faire passer entre autres les gaines électriques). L'épaisseur de la dalle de béton armé des planchers intermédiaires est divisée par 2 (de 10 cm, à 5 cm). Enfin, le bâtiment est supposé être constitué de façades rideaux, les parois extérieures en béton sont donc retirées.																																																													
Température de consigne fixe (Tfix)	La température de consigne du bâtiment est fixée à 21 °C en permanence.																																																													
Orientation Nord-Sud (NS)	A la place de l'orientation Est-Ouest, le bâtiment est orienté Nord-Sud (les grandes façades vitrées donnant sur les zones de bureaux sont orientées au Nord et au Sud). Ainsi d'importantes différences de température sont observées entre les zones, les bureaux orientés au Sud présentent des surchauffes de 10°C (31°C).																																																													
Zones non chauffées (Zones)	Les zones de sanitaires et de circulation ne sont pas chauffées, les différences de température entre les zones sont donc accrues. Des différences de température de 8 °C sont observées.																																																													
Modification de l'isolation 1 (ISO)	<p>La résistance globale de l'enveloppe est modifiée, 6 différentes modifications de l'enveloppe sont considérées. Cela permet de faire varier les ratios <i>Rate_m</i> et <i>Rate_l</i>. Ces scénarios sont présentés dans le tableau ci-dessous. Les fenêtres varient entre simple (s), double (d) et triple (t) vitrage. Les épaisseurs d'isolant utilisées dans les parois opaques sont divisées par 2, ou multipliées par 4.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th rowspan="2"></th> <th colspan="3">Vitrages</th> <th colspan="2">Parois opaques</th> <th rowspan="2"><i>Rate_m</i></th> <th rowspan="2"><i>Rate_l</i></th> </tr> <tr> <th>s</th> <th>d</th> <th>t</th> <th>/2</th> <th>* 4</th> </tr> </thead> <tbody> <tr> <td>ISO1</td> <td style="text-align: center;">X</td> <td></td> <td></td> <td style="text-align: center;">X</td> <td></td> <td>0,40</td> <td>0,18</td> </tr> <tr> <td>ISO2</td> <td style="text-align: center;">X</td> <td></td> <td></td> <td></td> <td style="text-align: center;">X</td> <td>0,10</td> <td>0,18</td> </tr> <tr> <td>ISO3</td> <td></td> <td style="text-align: center;">X</td> <td></td> <td style="text-align: center;">X</td> <td></td> <td>0,50</td> <td>0,28</td> </tr> <tr> <td>ISO4</td> <td></td> <td style="text-align: center;">X</td> <td></td> <td></td> <td style="text-align: center;">X</td> <td>0,14</td> <td>0,28</td> </tr> <tr> <td>ISO5</td> <td></td> <td></td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td></td> <td>0,58</td> <td>0,38</td> </tr> <tr> <td>ISO6</td> <td></td> <td></td> <td style="text-align: center;">X</td> <td></td> <td style="text-align: center;">X</td> <td>0,18</td> <td>0,38</td> </tr> </tbody> </table>		Vitrages			Parois opaques		<i>Rate_m</i>	<i>Rate_l</i>	s	d	t	/2	* 4	ISO1	X			X		0,40	0,18	ISO2	X				X	0,10	0,18	ISO3		X		X		0,50	0,28	ISO4		X			X	0,14	0,28	ISO5			X	X		0,58	0,38	ISO6			X		X	0,18	0,38
	Vitrages			Parois opaques		<i>Rate_m</i>	<i>Rate_l</i>																																																							
	s	d	t	/2	* 4																																																									
ISO1	X			X		0,40	0,18																																																							
ISO2	X				X	0,10	0,18																																																							
ISO3		X		X		0,50	0,28																																																							
ISO4		X			X	0,14	0,28																																																							
ISO5			X	X		0,58	0,38																																																							
ISO6			X		X	0,18	0,38																																																							

Tableau 7 : Description des scénarios étudiés

Pour chaque calibration différents indicateurs sont calculés :

- Un indicateur de convergence : le \hat{R} , défini au chapitre 2.
- Quatre indicateurs d'écarts:
 - RMSE_Init : entre les données de calibration et la simulation non calibrée
 - RMSE_Calib : entre les données de calibration et la simulation calibrée
 - RMSE_F : entre les données de validation de février et la simulation calibrée
 - RMSE_M : entre les données de validation de mars et la simulation calibrée.
- Des indicateurs d'identifiabilité : les valeurs des paramètres calibrés normalisés par les estimations physiques. Une valeur proche de 1 signifie que la valeur calibrée est proche de l'estimation physique du paramètre.

Le Tableau 8 regroupe l'ensemble de ces indicateurs pour chacun des scénarios étudiés :

Scenario	\hat{R}	RMSE_Init	RMSE_Calib	RMSE_F	RMSE_M	Paramètres calibrés normalisés par les estimations physiques				
						Rg	Cm	Am	Rate_m	Rate_l
Base	1,013	0,71	0,15	0,20	0,25	0,97	1,35	0,78	0,48	0,96
Isoext	1,002	1,02	0,13	0,16	0,22	0,96	1,69	0,81	0,65	0,28
Léger	1,001	0,61	0,24	0,34	0,50	0,96	1,10	0,75	0,25	1,71
Tfix	1,006	0,63	0,12	0,30	0,29	0,94	1,31	0,83	0,58	0,24
NS	1,004	0,78	0,19	0,24	0,26	0,96	1,33	0,76	0,25	1,15
Zones	1,002	0,64	0,14	0,23	0,27	0,98	1,32	0,70	0,39	0,95
ISO1	1,001	0,67	0,13	0,17	0,22	0,96	1,37	0,79	0,23	0,72
ISO2	1,001	1,66	0,19	0,28	0,65	0,89	1,33	0,81	0,22	1,28
ISO3	1,003	0,74	0,14	0,19	0,26	1,02	1,39	0,78	0,75	1,37
ISO4	1,004	1,25	0,16	0,23	0,37	0,92	1,32	0,78	0,23	1,00
ISO5	1,002	1,08	0,16	0,20	0,25	1,06	1,40	0,80	1,08	1,95
ISO6	1,007	0,74	0,15	0,21	0,29	0,97	1,31	0,78	1,37	1,71

Tableau 8 : Résultats de calibration du modèle ISO_Ca sur chacun des 12 scénarios étudiés

Pour l'ensemble des scénarios, le \hat{R} est inférieur à 1,1; les chaînes d'échantillonnage convergent. Dans l'ensemble des scénarios, les simulations sont proches des données d'apprentissage et de validation. L'indicateur RMSE_Calib varie entre de 0,12 °C et 0,24 °C. Sur les données de validation, les indicateurs RMSE_F et RMSE_M sont légèrement plus élevés, mais restent faibles (entre 0,16 °C et 0,65 °C). Pour le cas de base, les simulations issues des modèles TRNSYS et ISO_Ca sur les données de calibration et de validation sont affichées en Figure 37, Figure 38 et Figure 39.

En ce qui concerne l'identifiabilité physique des paramètres, la résistance globale apparaît correctement identifiée pour la majorité des scénarios (moins de 5 % d'écart avec la valeur physique). Des écarts plus élevés (10 %) sont observés lorsqu'un flux important passe par les parois légères (Iso2, Iso4); dans ces cas-là, les incertitudes liées aux résistances radiatives ont plus d'influence sur le modèle. Pour la plupart des scénarios, la capacité calibrée Cm est supérieure à l'estimation physique de 10 % à 30 %. L'écart est particulièrement important pour le scénario Isoext qui présente une inertie globale plus élevée. Plus la capacité augmente, plus les variations de cette capacité ont un faible impact sur le modèle, des erreurs plus élevées sont alors possibles. Les surfaces Am calibrées sont plus faibles que les estimations physiques (de 7 % à 30 %). Les écarts observés sur les paramètres Am et Cm , semblent raisonnables étant donné que ces paramètres caractérisant la dynamique globale du bâtiment sont comparés à des valeurs estimées par une méthode comportant de nombreuses hypothèses.

L'identification semble plus difficile pour les paramètres $Rate_m$ et $Rate_I$. Des écarts importants et variés sont observés pour les différents scénarios. Pour essayer de mieux analyser ces écarts, les valeurs calibrées non-normalisées, les valeurs estimées de manière physique, et les écarts relatifs entre ces valeurs, sont affichés dans le Tableau 9.

Scenario	Estimations physiques		Valeurs calibrées		Estimations Physique – Valeurs calibrées	
	<i>Rate_m</i>	<i>Rate_l</i>	<i>Rate_m</i>	<i>Rate_l</i>	<i>Rate_m</i>	<i>Rate_l</i>
Base	0,35	0,28	0,17	0,27	-0,18	-0,01
Isoext	0,35	0,28	0,23	0,08	-0,12	-0,20
Léger	0,35	0,28	0,09	0,48	-0,26	0,20
Tfix	0,35	0,28	0,20	0,07	-0,15	-0,21
NS	0,35	0,28	0,09	0,32	-0,26	0,04
Zones	0,35	0,28	0,14	0,27	-0,21	-0,01
ISO1	0,38	0,18	0,09	0,13	-0,29	-0,05
ISO2	0,09	0,18	0,02	0,23	-0,07	0,05
ISO3	0,48	0,28	0,36	0,38	-0,12	0,10
ISO4	0,14	0,28	0,03	0,28	-0,11	0,00
ISO5	0,55	0,38	0,59	0,74	0,04	0,36
ISO6	0,18	0,38	0,25	0,65	0,07	0,27

Tableau 9 : Comparaison entre les valeurs des paramètres *Rate_m* et *Rate_l* calibrés et non normalisés, avec les estimations physiques de ces derniers

Pour le ratio *Rate_m* qui représente la répartition du flux entre la branche lourde et les branches légères (fenêtres et infiltration), les écarts entre les valeurs estimées et calibrées varient de 0,04 à -0,29. Des écarts importants sont notamment observés dans les scénarios NS et Zones. Pour ces scénarios, les écarts de température entre zones peuvent s'expliquer par des difficultés d'identifiabilité. Un fort écart est aussi observé dans le scénario ayant la plus faible résistance d'enveloppe (Iso1); dans ce cas, les erreurs liées aux résistances radiatives peuvent avoir un impact plus important sur l'identifiabilité de *Rate_m*. Enfin, dans la majorité des cas, les valeurs calibrées sont plus faibles que les estimations physiques. Cela peut être dû à la position de la capacité *Cm*. Etant donné la structure du modèle ISO_Ca, le flux qui traverse les parois opaques, passe par la capacité *Cm*. Cependant, en pratique une bonne partie de cette capacité se situe au niveau des planchers et parois intérieures. En faisant passer une part plus importante du flux par les parois légères, il est possible que le modèle cherche à se rapprocher de ce comportement correspondant à une capacité intérieure non directement liée à l'extérieur. Ce phénomène peut aussi expliquer le fait que l'erreur soit particulièrement élevée pour le scénario « Léger » qui considère des façades rideaux sans masse thermique. Une proposition sera faite dans le paragraphe suivant pour essayer de répondre à ce problème. Enfin les écarts observés restent finalement cohérents avec le fait que le paramètre *Rate_m* n'a pas un

fort impact sur le modèle et est ainsi fortement soumis aux potentielles erreurs de simplification.

Pour le ratio *Rate_I*, qui représente la répartition du flux entre les deux branches légères (fenêtres et infiltration), les écarts varient de manière plus importante : de -0,21 à 0,36 et sont difficiles à interpréter. Cela peut s'expliquer par le fait que ce paramètre a très peu d'influence sur le modèle (Figure 36) et est donc particulièrement difficile à identifier.

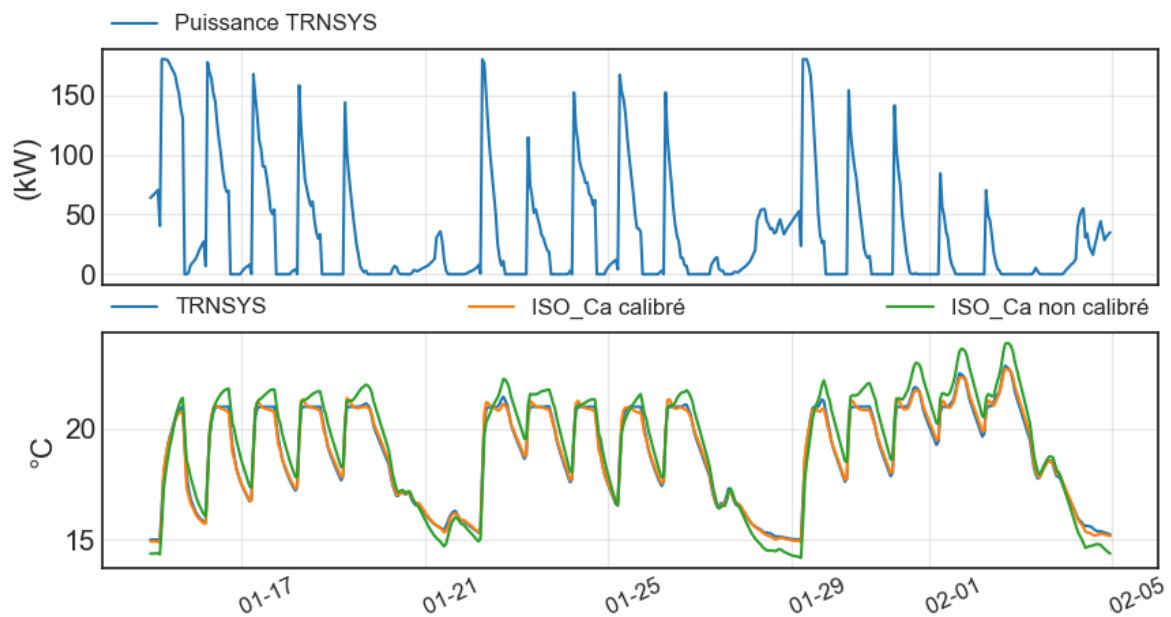


Figure 37 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de calibration, dans le cas de la configuration de base : 'Base' (les courbes bleue et orange se superposent)

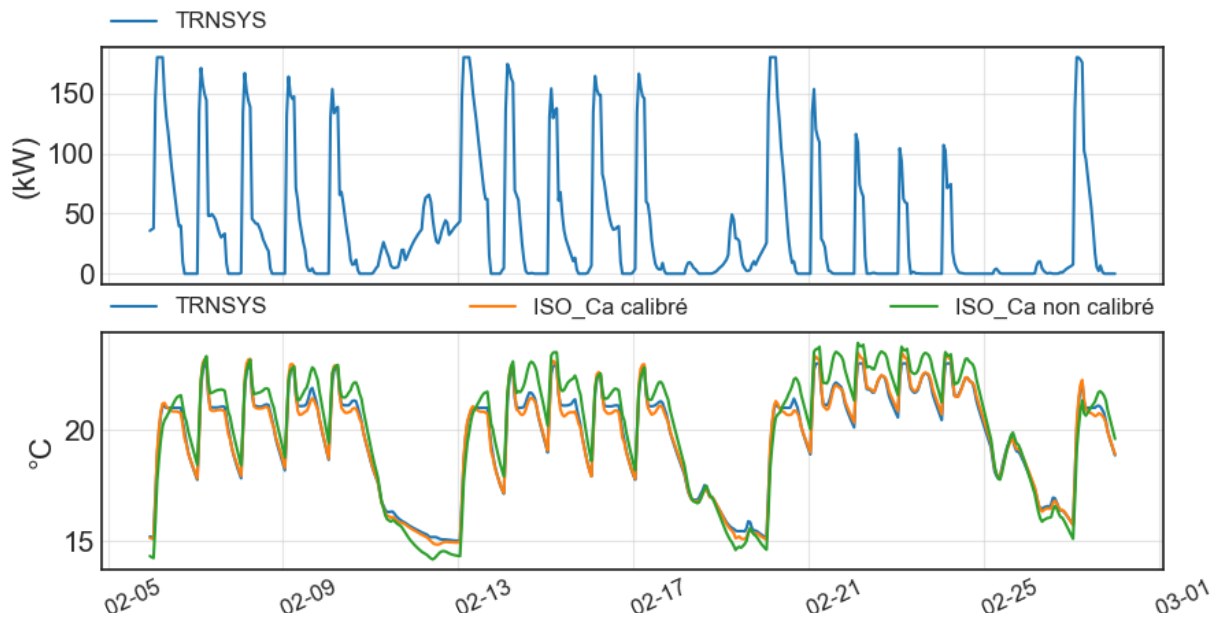


Figure 38 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de validation de février, dans le cas de la configuration de base : ‘Base’

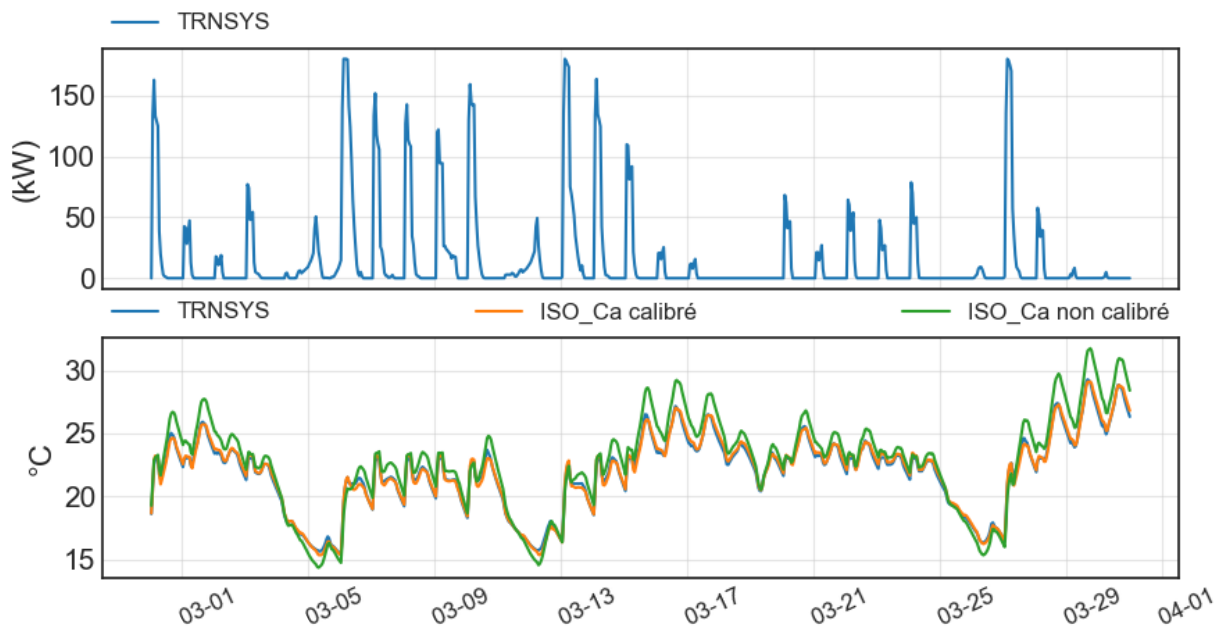


Figure 39 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de validation de mars, dans le cas de la configuration de base : ‘Base’

3.9 Etude d’une variante du modèle ISO_Ca, le modèle ISO_CpCa

Un potentiel problème mis en avant dans le paragraphe précédent réside dans la position de la capacité C_m dans le modèle ISO_Ca. Située au niveau des surfaces intérieures des parois lourdes, la position de cette capacité semble amener le modèle à faire passer une part plus importante du flux thermique par les parois légères. Pour

tenter de résoudre ce problème une nouvelle structure de modèle est proposée (Figure 40). Ce modèle, que l'on nommera ISO_CpCa, possède un nouveau nœud T_p relié à une capacité C_p représentant la masse thermique interne du bâtiment. La masse thermique située au niveau des parois lourdes en contact avec l'extérieur n'est pas modélisée. Dans le cas d'une isolation par l'intérieur, cette masse thermique est peu influente sur le modèle. Des problèmes peuvent cependant se révéler dans le cas d'une isolation par l'extérieur.

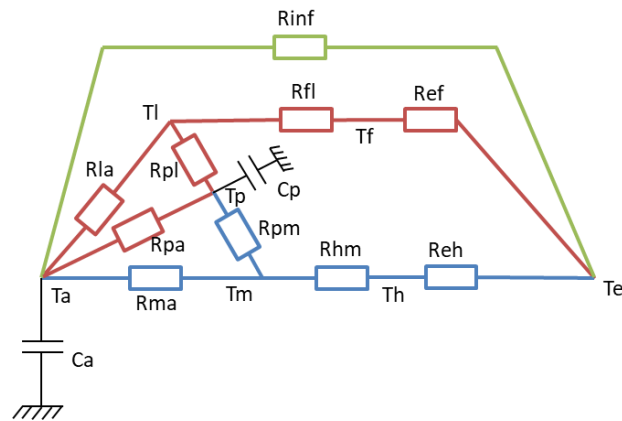


Figure 40 : Modèle ISO_CpCa

La capacité C_p est liée aux surfaces intérieures des parois lourdes et des fenêtres par des résistances radiatives R_{pl} et R_{pm} estimées en faisant l'hypothèse d'une petite surface dans un grand environnement :

$$R_{pm} = \frac{1}{4\varepsilon\sigma A_m \theta_0^3} \quad (106)$$

$$R_{pl} = \frac{1}{4\varepsilon\sigma A_l \theta_0^3} \quad (107)$$

La capacité C_p est également liée au nœud d'air par une résistance convective R_{ap} calculée en fonction de la surface de la masse thermique interne A_p :

$$R_{ap} = \frac{1}{A_p * h_i} \quad (108)$$

Les paramètres A_p et C_p prennent les mêmes valeurs que A_m et C_m dans le modèle ISO_Ca, calculés selon la méthode détaillée (AFNOR 2010). A_m correspond dans ce modèle à la surface intérieure des parois lourdes (toit et murs extérieurs). Le flux radiatif entre les surfaces intérieures des fenêtres et les parois lourdes n'est pas modélisé. Cette hypothèse est faite car sur un grand bâtiment à plusieurs étages, les

surfaces de fenêtres et de parois lourdes échangent peu entre elles, mais échangent principalement avec les surfaces intérieures liées au nœud Tp .

Le modèle ainsi défini est calibré sur les 12 scénarios définis au paragraphe précédent. Les paramètres calibrés sont à nouveau : Rg , Ap , Cp , $Rate_m$ et $Rate_l$. Les résultats de calibration sont affichés dans le Tableau 10. Comme précédemment les valeurs calibrées non-normalisées des paramètres $Rate_m$ et $Rate_l$ sont affichées dans le Tableau 11.

Scenario	\hat{R}	RMSE _Init	RMSE _Calib	RMSE _F	RMSE _M	Paramètres calibrés normalisés par les estimations physiques				
						Rg	Cp	Ap	$Rate_m$	$Rate_l$
Base	1,002	0,89	0,13	0,17	0,28	0,95	1,36	0,70	0,71	0,29
Isoext	1,006	1,16	0,13	0,13	0,21	0,95	1,69	0,71	0,67	0,21
Léger	1,002	0,81	0,22	0,31	0,64	0,95	1,12	0,65	0,64	1,37
Tfix	1,004	0,92	0,09	0,33	0,28	0,94	1,26	0,60	0,84	0,22
NS	1,014	0,96	0,17	0,20	0,25	0,95	1,35	0,68	0,63	0,70
Zones	1,007	0,79	0,12	0,21	0,33	0,96	1,34	0,63	0,64	0,35
ISO1	1,003	0,91	0,12	0,15	0,22	0,93	1,38	0,71	0,30	0,26
ISO2	1,001	1,68	0,19	0,27	0,65	0,89	1,34	0,72	0,22	1,18
ISO3	1,004	0,72	0,12	0,14	0,22	0,97	1,38	0,70	0,77	0,22
ISO4	1,001	1,29	0,15	0,21	0,37	0,92	1,33	0,69	0,27	0,99
ISO5	1,000	0,65	0,13	0,15	0,26	1,01	1,38	0,71	1,23	0,23
ISO6	1,003	0,78	0,13	0,21	0,41	0,98	1,32	0,69	2,09	1,32

Tableau 10 : Résultats de calibration du modèle ISO_CpCa sur chacun des 12 scénarios étudiés

Scenario	Estimations physiques		Valeurs calibrées		(Estimations Physiques – Valeurs calibrées)	
	<i>Rate_m</i>	<i>Rate_l</i>	<i>Rate_m</i>	<i>Rate_l</i>	<i>Rate_m</i>	<i>Rate_l</i>
Base	0,35	0,28	0,25	0,08	-0,10	-0,20
Isoext	0,35	0,28	0,23	0,06	-0,12	-0,22
Léger	0,35	0,28	0,22	0,38	-0,13	0,10
Tfix	0,35	0,28	0,29	0,06	-0,06	-0,22
NS	0,35	0,28	0,22	0,19	-0,13	-0,09
Zones	0,35	0,28	0,23	0,10	-0,12	-0,18
ISO1	0,38	0,18	0,11	0,05	-0,27	-0,13
ISO2	0,09	0,18	0,02	0,21	-0,07	0,03
ISO3	0,48	0,28	0,37	0,06	-0,11	-0,22
ISO4	0,14	0,28	0,04	0,28	-0,10	0,00
ISO5	0,55	0,38	0,68	0,09	0,13	-0,29
ISO6	0,18	0,38	0,38	0,50	0,20	0,12

Tableau 11 : Comparaison entre les valeurs des paramètres *Rate_m* et *Rate_l* calibrés et non normalisés, avec les estimations physiques de ces derniers

Aucun problème de corrélation n'est observé ($\hat{R} < 1,1$). Les indicateurs d'écarts sont similaires à ceux observés avec le modèle ISO_Ca. Les valeurs identifiées des paramètres *Rg*, *Cp* et *Ap* sont également très proches de celles identifiées dans le cas du modèle ISO_Ca. Une légère amélioration de l'identification du paramètre *Rate_m* semble cependant apparaître. Les écarts absolus entre les valeurs calibrées et estimées sont compris entre 0,13 et 0,06, excepté dans le cas du bâtiment très faiblement isolé (ISO1) qui présente toujours une erreur importante. La dérive observée précédemment sur le scénario 'léger' est en partie corrigée. Étonnamment, aucune dérive particulière n'est observée pour le scénario Isoext. Les erreurs restantes sont certainement dues aux imprécisions de simplification. C'est donc cette variante du modèle ISO qui sera privilégiée dans la méthode développée dans cette thèse.

3.10 Conclusion

L'objectif de ce chapitre était de sélectionner un modèle RC ayant une bonne capacité prédictive du comportement thermique des bâtiments d'un quartier, et de forts liens avec les caractéristiques physiques des bâtiments.

Après analyse bibliographique, le modèle ISO est apparu pertinent pour répondre à ces objectifs. Il dispose notamment d'une méthode complète de paramétrage, incluant l'estimation des paramètres dynamiques Am et Cm . Une capacité supplémentaire Ca a été ajoutée au nœud d'air donnant lieu au modèle ISO_Ca. Ce modèle a été étudié sur le cas d'étude d'un bâtiment de bureau modélisé sur TRNSYS. Pour analyser en détail le modèle, ce dernier a été calibré sur les données de températures issues du modèle TRNSYS. Un reparamétrage du modèle a été proposé avec succès afin d'éliminer les problèmes de convergence lors de la calibration. 12 configurations différentes du bâtiment ont été étudiées. Des jeux de données de validation, fortement différents des données d'entraînement, ont aussi été utilisés pour éprouver la robustesse du modèle

Le paramètre le plus influent du modèle est de loin la résistance globale Rg . Cela permet une identification robuste, les erreurs de paramétrage sont faibles pour l'ensemble des scénarios (environ 5 %). Ces faibles écarts restent cependant la principale origine des erreurs de prédiction des températures, ils sont probablement dus aux imprécisions de simplification, notamment au niveau des flux radiatifs. Les autres paramètres Cm , Am , $Rate_m$ et $Rate_l$ régissent plus en détail le comportement dynamique du modèle. Ces paramètres sont plus difficiles à estimer, des erreurs plus importantes sont observées ; mais ces erreurs ont moins d'influence sur les prédictions de température. Bien qu'impliquant de fortes hypothèses les estimations des paramètres Am et Cm parviennent à s'approcher des valeurs optimales (généralement 10 % à 30 % d'écart). Les écarts sont plus élevés pour les paramètres $Rate_m$ et $Rate_l$. Pour le paramètre $Rate_m$, les erreurs ont pu en partie être expliquées, et le modèle ISO_CpCa a permis de réduire significativement ces écarts. C'est donc ce dernier modèle qui sera retenu dans la suite de cette thèse. Le paramètre $Rate_l$ est le moins influant, les écarts observés sont encore plus importants que pour $Rate_m$ et sont difficiles à interpréter; il peut être décidé de ne pas calibrer ce paramètre.

L'étude présentée dans ce chapitre aurait pu être prolongée, par exemple en considérant d'autres jeux de données, d'autres cas d'étude, ou d'autres modèles. La

méthode d'estimation des paramètres A_m et C_m pourrait probablement être améliorée. Il aurait également pu être intéressant d'étudier plus en détail l'identification des paramètres *Rate_m*, et *Rate_l* en essayant de discerner des cas dans lesquels ces paramètres sont plus influents et donc plus facilement identifiables. La prise en compte d'un débit d'infiltration dynamique (dépendant de la vitesse du vent) pourrait par exemple rendre le paramètre *Rate_l* plus facilement identifiable. La difficulté reste de disposer des données de vent locales fiables.

Le comportement du modèle ainsi observé est compatible avec la méthode développée dans cette thèse. Il pourra être utilisé pour simuler les comportements thermiques des bâtiments d'un quartier à partir des données des constructions disponibles. Aussi, ce modèle peut prendre place dans la méthode de calibration bayésienne développée. Les liens entre les paramètres du modèle et les paramètres physiques des bâtiments sont maîtrisés, des probabilités a priori sur les paramètres calibrés peuvent être définis. Le re-paramétrage permettra aussi une convergence plus facile de la méthode de calibration bayésienne.

Chapitre 4

Paramétrage a priori d'un modèle
de quartier : cas du quartier de
Clichy-Batignolles

Table des matières

Chapitre 4.....	105
4.1 Introduction.....	108
4.2 Construction d'une table SIG.....	109
4.2.1 Utilisation de la base de données BD_TOPO®.....	110
4.2.2 Corrections manuelles de la table SIG.....	111
4.2.3 Calcul des surfaces de déperdition.....	113
4.2.4 Calculs des apports solaires.....	114
4.3 Identification et positionnement des locaux associés à des surfaces chauffées et des usages.....	116
4.3.1 Les différentes définitions de surfaces.....	117
4.3.2 Utilisation de la base de données MAJIC (zone ENS).....	119
4.3.3 Utilisation des documents de construction (zone Ouest).....	122
4.3.4 Lien entre les locaux et la table SIG.....	123
4.4 Enrichissement statistique : application aux bâtiments de bureaux.....	124
4.4.1 Classification des bâtiments de bureaux par types.....	125
4.4.2 Classe d'inertie.....	126
4.4.3 Coefficients de déperditions thermiques.....	126
4.4.4 Infiltration.....	127
4.4.5 Profils d'occupation.....	129
4.4.6 Apports thermiques métaboliques liés aux occupants.....	133
4.4.7 Apports thermiques liés aux consommations d'électricité spécifique.....	134
4.4.8 Ventilation mécanique.....	134
4.4.9 Scénario de stores extérieurs (apports solaires).....	134
4.4.10 Scénario de températures de consigne.....	135

4.5	Exemples de modèle a priori sur le quartier de Clichy-Batignolles	136
4.6	Conclusion	138

4.1 Introduction

Ce chapitre présente une méthode de construction, à l'échelle du quartier, d'une base de données pour le paramétrage a priori de modèles de bâtiment de type RC, dont la structure a été sélectionnée dans le chapitre précédent : ISO_CpCa. Les paramètres de ces modèles a priori pourront, par la suite, être mis à jour par calibration bayésienne, avec des données d'exploitation. La construction d'une telle base de données à l'échelle d'un quartier peut s'avérer chronophage. Pour limiter ce temps de travail, la méthode cherche à exploiter autant que possible des bases de données territoriales. Cette méthode est présentée sur le cas d'étude du quartier de Clichy-Batignolles (Figure 41), qui permet d'illustrer l'hétérogénéité des bâtiments et des données disponibles. Les données nécessaires au paramétrage a priori du modèle ISO_CpCa sont rappelées ci-dessous :

- Les surfaces de déperdition (S_{wall} , S_{floor} , S_{roof} , S_{win})
- Les apports solaires ($\varphi_{sol_{opaque}}$, $\varphi_{sol_{in}}$)
- Les propriétés thermiques du bâtiment : les coefficients de déperdition thermique (U_{wall} , U_{floor} , U_{roof} , U_{win}), la résistance d'infiltration R_{inf} , et les paramètres liés à la capacité thermique: Cm et Am (selon la norme (AFNOR 2013))
- Les apports thermiques liés au métabolisme des occupants : φ_{meta}
- Les apports thermiques liés aux consommations d'électricité spécifique : φ_{elec}
- La résistance dynamique de ventilation mécanique : R_{meca}
- Des scénarios d'utilisation des stores extérieurs
- Des scénarios de température de consigne

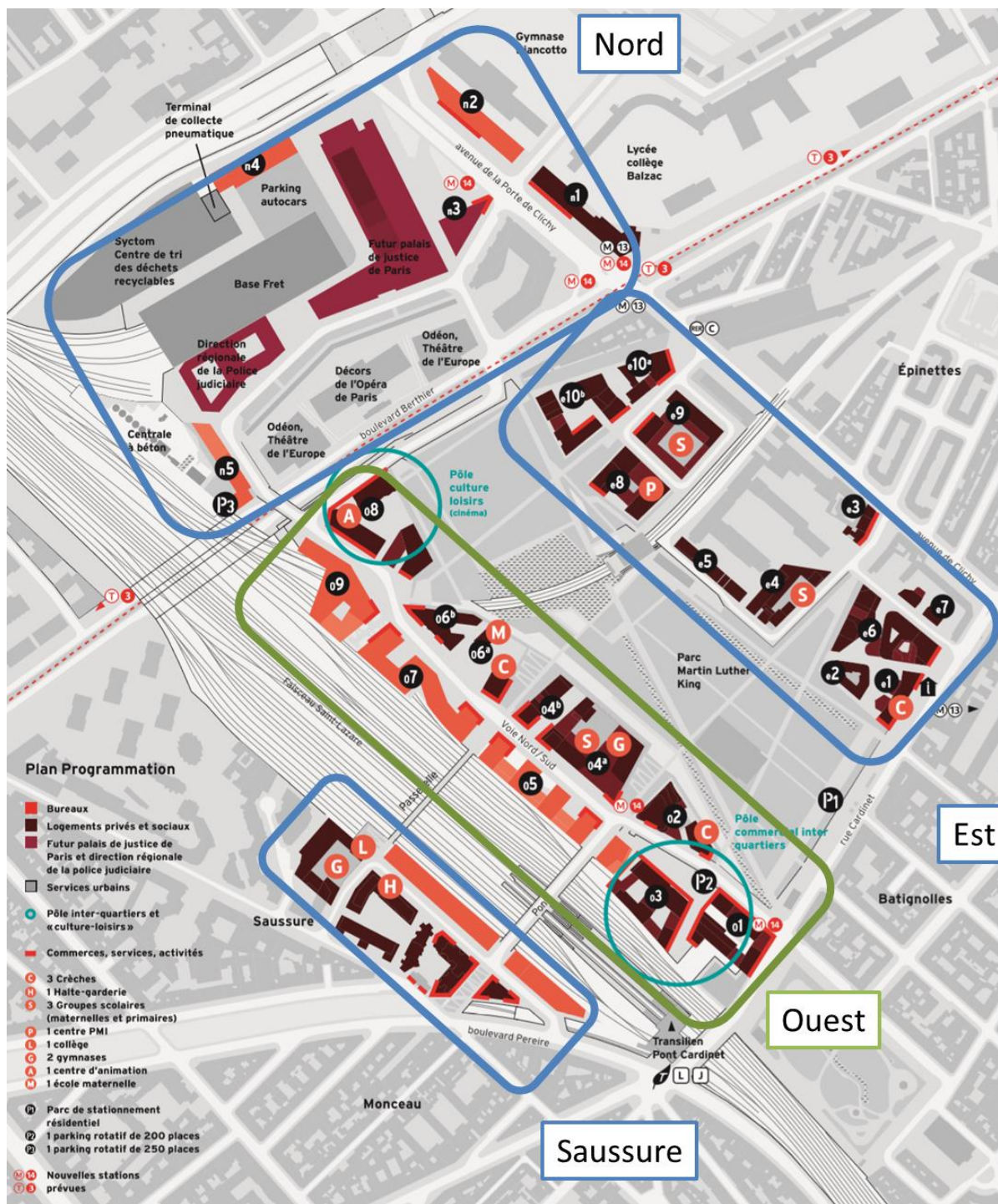


Figure 41 : Plan du quartier de Clichy-Batignolles

4.2 Construction d'une table SIG

Afin de prendre en compte les formes géométriques du quartier et d'estimer les surfaces de déperdition thermique et les gains solaires, une table SIG (Système d'Information Géographique) est créée. Une table SIG est caractérisée par la présence d'une colonne (attribut) correspond à une forme géométrique géolocalisée selon un

Système de Coordonnées de Référence (SCR). Cette colonne peut être lue et affichée par un logiciel SIG tel que QGIS largement connu et utilisé par les acteurs du territoire, sous licence libre et téléchargeable sur le site <http://www.qgis.org/>. Dans le cas présent, les formes géométriques géolocalisées sont les emprises au sol des bâtiments. La complexité intermédiaire de la table SIG créée est en accord avec l'approche de modélisation choisie, elle permet de capter les différences majeures de surfaces de déperdition thermique et gains solaires en fonction des types de bâtiments et de tissus urbains, tout en limitant le travail nécessaire à la création de telles tables à l'échelle de quartier.

4.2.1 Utilisation de la base de données BD_TOPO®

La méthode de création de la table SIG proposée s'appuie d'abord sur la table BATI_INDIFERENCIE® de la base de données BD_TOPO® publiée par l'Institut National de Géographie (IGN). On observe sur la Figure 42 que le quartier de Clichy-Batignolles étant récent, la base de données BD_TOPO® a progressivement été mise à jour ces dernières années et semble aujourd'hui relativement complète. La BD_TOPO® est également utilisée pour les formes des bâtiments frontaliers aux quartiers. En effet, ces derniers jouent un rôle de masque solaire proche pour les bâtiments du quartier, ils seront donc pris en compte dans les calculs de l'ensoleillement reçu par les bâtiments.

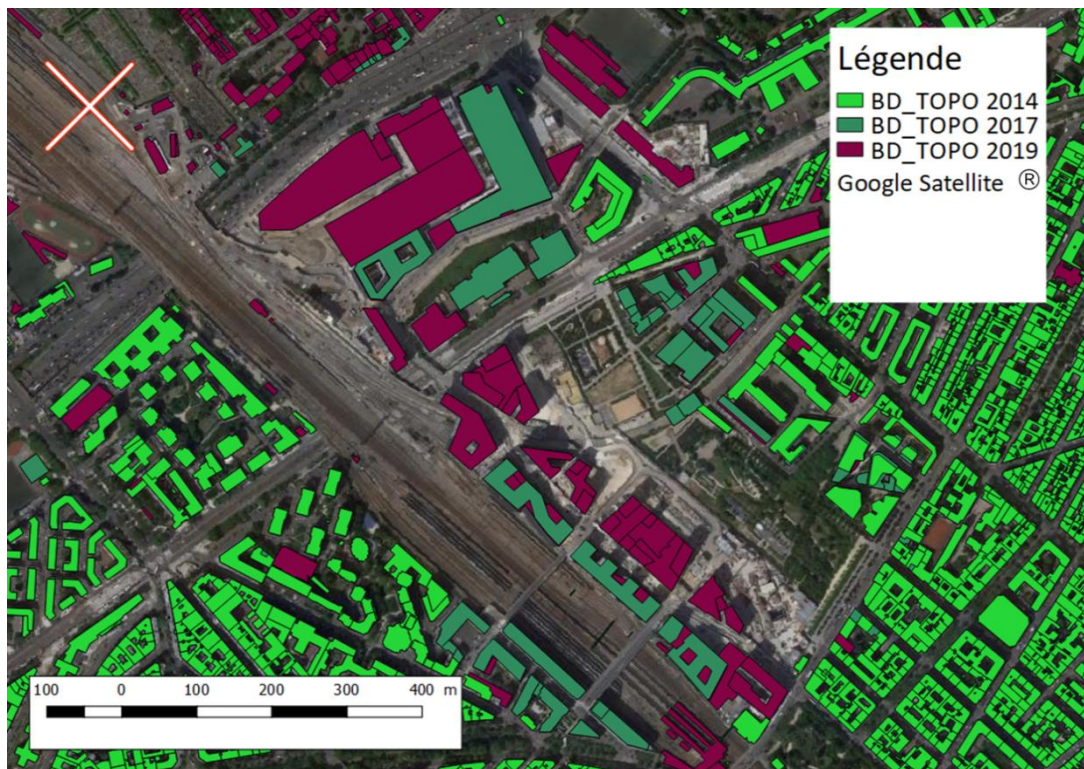


Figure 42 : Superposition des données de la BD_TOPO®, avec le quartier de Clichy Batignolles

Cependant l'utilisation de cette base semble insuffisante pour la description des bâtiments dans le périmètre du quartier étudié pour les raisons suivantes :

- L'architecture des bâtiments de ce quartier est souvent complexe et ainsi mal modélisée par la base de données ;
- La base de données permet difficilement de considérer des zones thermiques distinctes pour les bâtiments multi usages (exemple : bâtiments de logements avec commerce de bas d'immeuble) ;
- Cette base ne dispose pas d'information concernant les taux de vitrage.

4.2.2 Corrections manuelles de la table SIG

Afin de pallier les manques de la BD_TOPO®, il est proposé d'améliorer manuellement la table de données SIG. Le logiciel QGIS associé à des images aériennes Google Satellite®, permet de retracer plus finement les contours des bâtiments. Les bâtiments de hauteurs variables sont découpés en plusieurs formes de hauteur homogène. Les valeurs de ces hauteurs sont majoritairement déduites du nombre d'étages.

Aussi les formes géométriques associées à des usages distincts, répartis sur différents étages d'un même bâtiment (exemple : commerces de bas d'immeuble),

sont considérés distinctement. L'identification des formes bâties associées à des usages distincts peut être faite via Google Street Map®, ou en utilisant, si disponibles, des plans de construction. En outre, Google Street Map® est utilisé pour renseigner les nombres d'étages, les taux de vitrage, et la présence de stores extérieurs.

Sur la Figure 43 est présentée une illustration de la construction manuelle de la table SIG, comparée aux données issues de la BD_TOPO® dans le cas du quartier de Clichy-Batignolles. Le modèle SIG complet du quartier est présenté sur la Figure 44.



Figure 43 : Comparaison entre la BD_TOPO® et la table SIG construite manuellement



Figure 44 : Table SIG du quartier de Clichy-Batignolles

Pour chaque forme géométrique identifiée dans la table SIG, les surfaces de déperdition et les apports solaires sont estimés selon les méthodes suivantes.

4.2.3 Calcul des surfaces de déperdition

Un algorithme a été développé en langage SQL (Structured Query Language), afin de calculer les surfaces de déperdition (S_{wall} , S_{floor} , S_{roof}) associées aux formes géométriques de la table SIG. La librairie PostGIS, extension du système de gestion de bases de données relationnelles PostgreSQL, permet la manipulation d'informations géographiques sous forme de géométries. Cet algorithme permet notamment d'identifier les surfaces mitoyennes horizontales et verticales (voir Figure 45). L'ensemble des surfaces mitoyennes identifiées sur le quartier de Clichy-Batignolles sont présentées en Figure 46. Les surfaces de vitrage (S_{win}) sont calculées à partir des taux de vitrage estimés dans le paragraphe précédent.

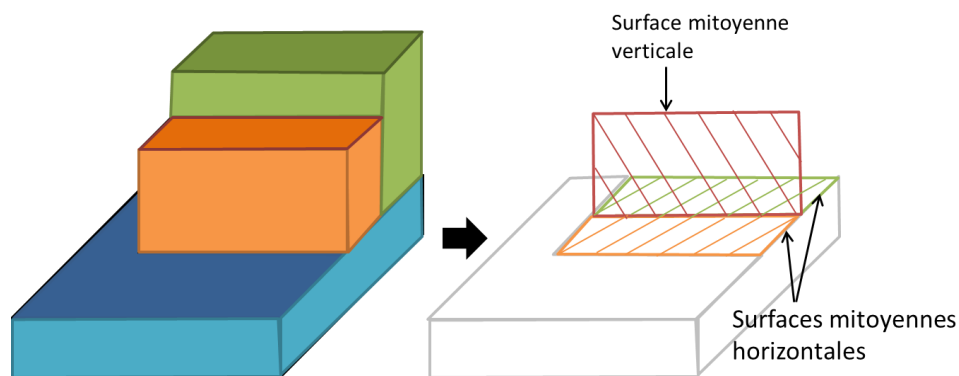


Figure 45 : Exemple d'identification de surfaces mitoyennes horizontales et verticales

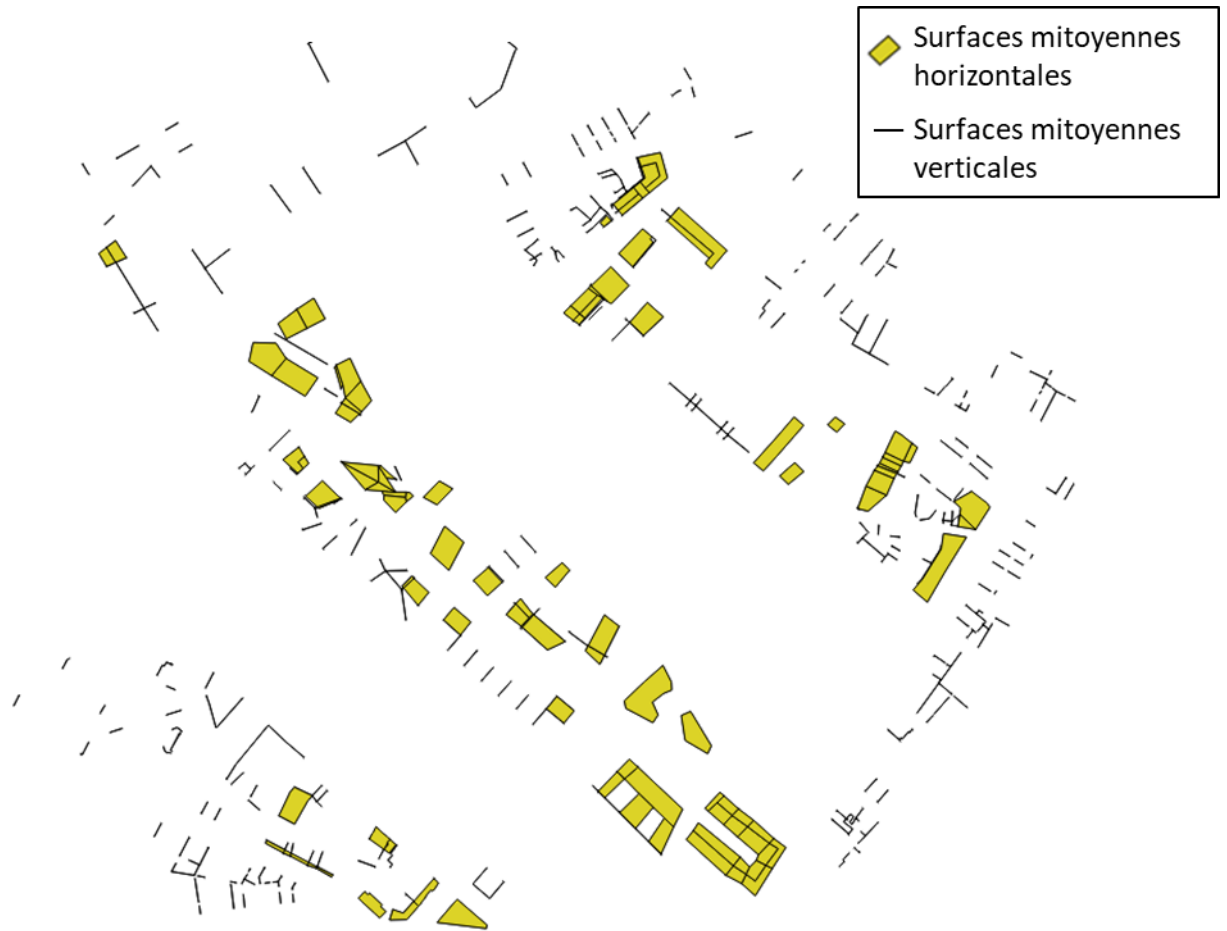


Figure 46 : Surfaces mitoyennes horizontales et verticales identifiées sur le quartier de Clichy-Batignolles

4.2.4 Calculs des apports solaires

Les apports solaires radiatifs ($\varphi_{\text{sol}_{\text{in}}}$ et $\varphi_{\text{sol}_{\text{opaque}}}$), reçus respectivement par les surfaces vitrées (S_{win}), et les surfaces opaques ($S_{\text{wall}} + S_{\text{roof}}$), sont calculés par le logiciel libre Daysim (Reinhart s. d.), via l'interface de la plateforme City Energy Analyst (Fonseca et al. 2016). Daysim a fait l'objet de plusieurs validations et publications (Reinhart 2004; Reinhart et al. s. d.; Reinhart et Andersen 2006; Reinhart et Herkel 2000; Reinhart et Walkenhorst 2001). Sur chaque façade et toiture, des points de calcul sont positionnés (le nombre de points de calcul considérés est paramétrable). Pour ces points de calcul, les rayons du soleil sont projetés à chaque pas de temps. Le logiciel considère les réflexions entre bâtiments qui dépendent notamment des taux de vitrage des parois verticales. L'interface développée pour la plateforme CEA (City Energy Analyst), permet d'utiliser

directement en entrée la table SIG du quartier (Figure 48). De plus, l'interface permet un paramétrage du logiciel Daysim adapté à la modélisation de quartiers.

En plus de la table SIG, Daysim nécessite en entrée un modèle numérique de terrain (MNT). Cela correspond à une matrice de pixels, où chaque pixel correspond à une valeur d'élévation du terrain. La donnée MNT RGE ALTI® 2018 fournie par l'IGN est utilisée (Figure 47).

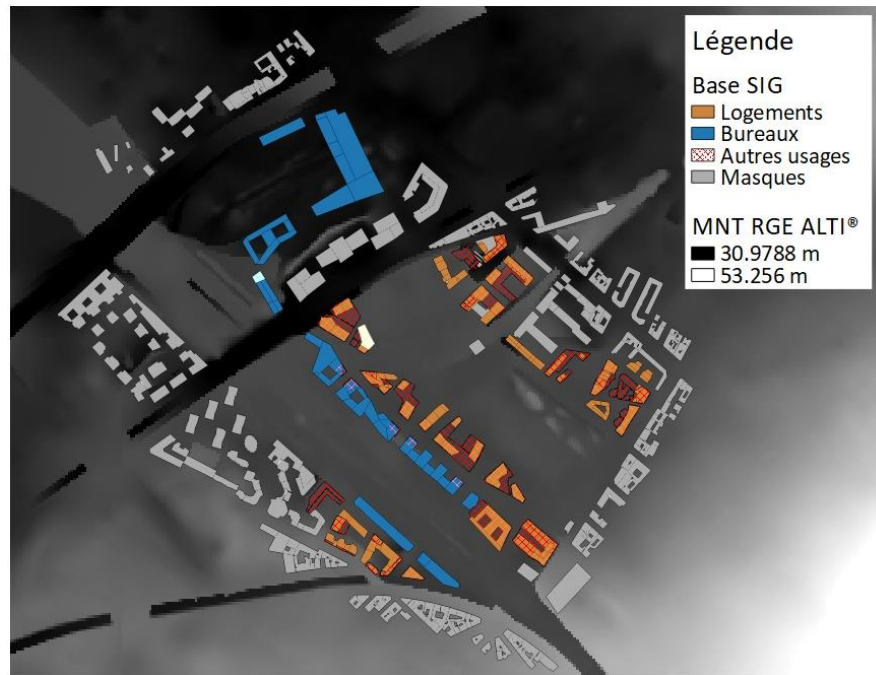


Figure 47 : Modèle numérique de terrain utilisé en entrée de Daysim (extrait de MNT RGE ALTI® 2018)

Dans le cas de la modélisation du quartier de Clichy-Batignolles, des données météo récentes issues de la station de Montsouris sont utilisées. Ces données sont définies au pas de temps horaire (UT0), et renseignent la température extérieure et la puissance solaire globale reçue sur une paroi horizontale. Afin de correspondre aux données d'entrée nécessaires à Daysim, l'ensoleillement global est séparé en ensoleillement direct et diffus selon la méthode de (Orgill et Hollands 1977).

Les données de sortie regroupent, pour chaque bâtiment, l'ensoleillement horaire reçu par chaque point de calcul et la surface associée à ces points de calcul en m^2 . Ces données sont utilisées en entrée du modèle de quartier développé dans cette thèse.

Remarque : L'atténuation des apports solaires radiatifs ($\varphi_{sol_{in}}$) par les stores extérieurs sera présentée plus loin, dans le paragraphe 3 : Enrichissement statistique.

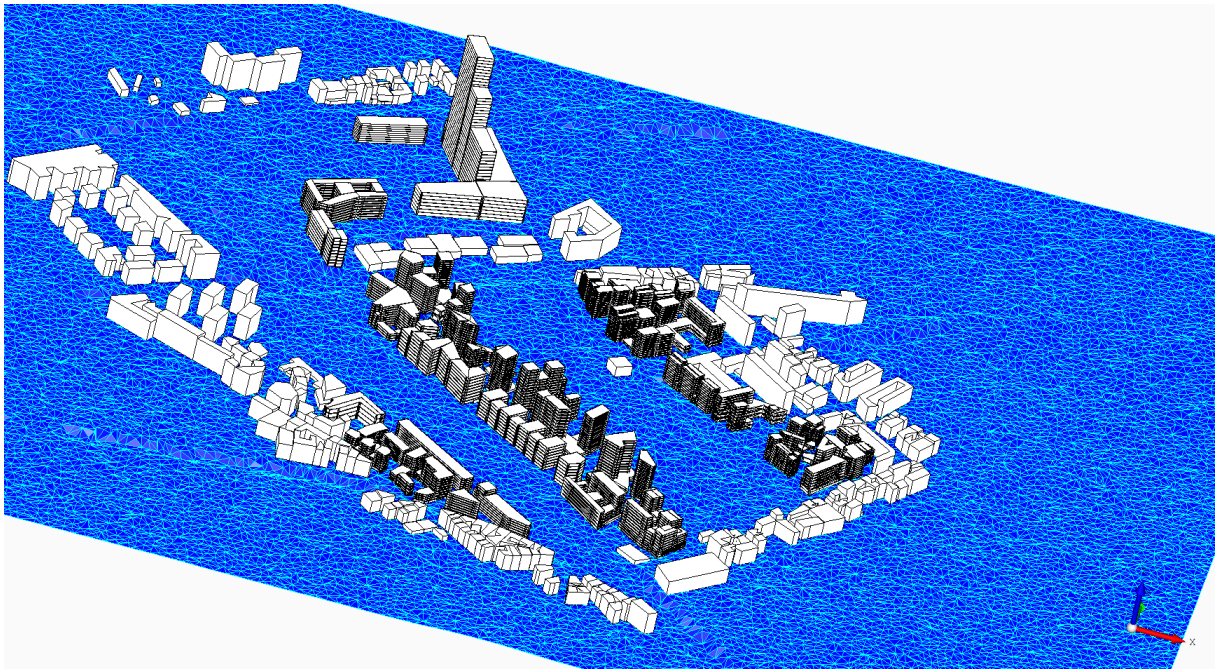


Figure 48 : Modèle du quartier en entrée du logiciel Daysim, créé par la plateforme CEA (Fonseca et al. 2016) à partir de la table SIG

4.3 Identification et positionnement des locaux associés à des surfaces chauffées et des usages

Bien qu'il soit possible d'estimer les surfaces et les volumes chauffés à partir de la table SIG (surface au sol multipliée par le nombre d'étages), cette information est entachée de beaucoup d'incertitudes : les bâtiments ne sont pas complètement occupés et les architectures complexes sont mal représentées. De plus, la table SIG seule ne nous renseigne pas précisément sur les usages des bâtiments. Ainsi, après avoir rappelé les différentes définitions de surface utilisées dans le secteur du bâtiment, il est fait appel à d'autres sources d'informations pour identifier et positionner les surfaces chauffées et les usages associées. La méthode proposée utilise d'abord la base des fichiers fonciers MAJIC ; cela permet de ne pas dépendre des données de construction fournis par l'aménageur, qui se sont avérées être fortement hétérogènes, demandent un travail de traitement important, et sont difficiles à obtenir, notamment sur des quartiers non récents. Cette base de données est utilisée pour la zone ENS du quartier de Clichy-Batignolles. L'utilisation des données de construction est cependant retenue pour la zone Ouest du quartier de Clichy-Batignolles, plus récente, sur laquelle la base MAJIC n'est pas à jour.

Dans la suite, on utilisera l'entité du « local ». Cette entité est définie dans la base MAJIC : « Le local est un ensemble de « différentes constructions qui, en raison de leur agencement, forment un ensemble immobilier homogène d'un même compte-proprétaire ». Les locaux peuvent prendre plusieurs formes : appartement, maison, ou local commercial (établissement de bureaux, commerce). Il sera alors cherché à identifier les locaux du quartier, caractérisés par des surfaces chauffées et des usages.

4.3.1 Les différentes définitions de surfaces

Dans le domaine, du bâtiment il existe de nombreuses définitions de la surface : SHON, SHAB, SHOB, SRT, surface Carrez, surface utile, surface de plancher (Tableau 12). Depuis 2012, la SHON et la SHOB ont été remplacées par la surface de plancher. Cependant la SHON continue à apparaître dans des documents de construction et des rapports de simulation énergétique dynamique (SED).

En plus des surfaces présentées dans le Tableau 12, il y a la surface utile pour laquelle plusieurs définitions coexistent. Selon le Code de la construction et de l'habitation, dans l'article R. 353-161, la surface utile est égale « à la surface habitable du logement, telle qu'elle est définie à l'article R. 111-2, augmentée de la moitié de la surface des annexes définies par un arrêté du Ministre chargé du logement ». Sont notamment inclus dans les annexes : les balcons, les garages et les places de parking. Ne sont en revanche pas compris : les surfaces de circulations verticales et horizontales, les paliers d'étage et l'encombrement des ouvrages construits (murs, voiles, cloisons, poteaux, etc.). Aussi, dans le domaine de l'immobilier d'entreprise, il existe aussi la SUB (surface utile brute), utilisée pour déterminer les loyers. Il s'agit de la SHON moins les éléments structuraux (poteaux, murs extérieurs, refends gaines techniques, circulations verticales...) et les parties communes. Enfin il y a la SUN (surface utile nette), relative à la surface réservée aux espaces de travail : bureaux, ateliers, laboratoires, salles de réunion.

Dans le modèle développé, la surface doit être liée au volume d'air chauffé. La surface qui se rapproche le plus de cette définition est la surface RT, mais cette surface est rarement renseignée dans les données exploitées. La SHAB semble également appropriée ; cependant, les espaces situés sous les escaliers, devraient être comptabilisés. De même, les surfaces de plancher dont la hauteur sous plafond est inférieure à 1,80 m devraient être considérées. On peut cependant noter que ces

spécificités ont peu d'impact sur les valeurs de surfaces, notamment dans le cas de grands bâtiments collectifs. Aussi, une incertitude demeure sur la prise en compte ou non, dans les surfaces SHAB renseignées pour les bâtiments collectifs, des parties communes (zone de circulation). Enfin, les locaux techniques non chauffés, non pris en compte peuvent avoir un impact sur la consommation.

TABLEAU COMPARATIF DES SURFACES RÉGLEMENTAIRES (2017)

		Surface hors œuvre brute	Surface hors œuvre nette	Emprise au sol	Surface taxable	Surface de plancher	Surface thermique RT	Surface loi Carrez	Surface habitable
		SHOB ¹	SHON ¹	ES	ST	SDP	SRT	Carrez	SHAB
Principe général	Surface de plancher de hauteur ≤1,80m	OUI	OUI	OUI	NON	NON	OUI	NON	NON
	Épaisseur des murs extérieurs (compris embrasures)	OUI	OUI	OUI	NON	NON	OUI	NON	NON
	Vides et trémies d'escaliers et ascenseurs	NON	NON	OUI	NON	NON	NON	NON	NON
	Marches et paliers intermédiaires d'escaliers	NON	NON	OUI	NON	NON	NON	NON	NON
	Rampes d'accès intérieures	NON	NON	OUI	NON	NON	NON	NON	NON
	Gaines et conduits de cheminées (hauteur ≤1,80m)	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON
	Surface des cloisons (compris embrasures)	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON
	Surface des placards et foyers de cheminée	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI
Surface de placard avec cumulus à hauteur ≤1,80m	OUI	OUI	OUI	OUI	OUI	OUI	NON	OUI	
Niveaux habitables	Local technique chauffé	OUI	OUI	OUI	OUI	OUI ²	OUI	OUI	OUI
	Local technique non chauffé	OUI	OUI	OUI	OUI	OUI ²	NON	OUI	OUI
	Cellier chauffé	OUI	OUI	OUI	OUI	OUI ²	OUI	OUI	OUI
	Cellier non chauffé	OUI	OUI	OUI	OUI	OUI ²	NON	OUI	OUI
Combles	Comble non aménageable ³ (hauteur ≤1,80m)	OUI	NON	OUI	NON	NON	NON	NON	NON
	Comble non aménageable ³ (hauteur >1,80m)	OUI	NON	OUI	OUI	NON	NON	NON	NON
	Comble aménageable et non aménagé (h >1,80m)	OUI	OUI	OUI	OUI	OUI	NON	OUI	NON
	Comble aménagé (hauteur >1,80m)	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI
	Local technique en comble (hauteur >1,80m)	OUI	NON	OUI	OUI	OUI	NON	OUI	NON
Sous-sol	Sous-sol non aménageable (hauteur ≤1,80m)	OUI	NON	OUI ⁴	NON	NON	NON	NON	NON
	Sous-sol non aménageable (hauteur >1,80m)	OUI	NON	OUI ⁴	OUI	OUI	NON	NON	NON
	Sous-sol aménageable et non aménagé (h >1,80m)	OUI	OUI	OUI ⁴	OUI	OUI	NON	NON	NON
	Sous-sol aménagé (hauteur >1,80m)	OUI	OUI	OUI ⁴	OUI	OUI	OUI	OUI	NON
	Cave	OUI	NON	OUI ⁴	OUI	OUI ²	NON	NON	NON
Garage	Local technique en sous-sol (hauteur >1,80m)	OUI	NON	OUI ⁴	OUI	OUI	NON	OUI	NON
	Surface de stationnement close et couverte	OUI	NON	OUI	OUI	NON	NON	NON	NON
	Surface de stationnement non close (carport)	OUI	NON	OUI	NON ⁵	NON	NON	NON	NON
Véranda	Surface de stationnement extérieure	NON	NON	NON	NON ⁵	NON	NON	NON	NON
	Véranda chauffée	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON
Construction non close	Véranda non chauffée	OUI	OUI	OUI	OUI	OUI	NON	OUI	NON
	Auvent et casquette non soutenus par des poteaux	NON	NON	NON	NON	NON	NON	NON	NON
	Auvent et casquette soutenus par des poteaux	OUI	NON	OUI	NON	NON	NON	NON	NON
	Construction non close (hangar, abri, préau, etc)	OUI	NON	OUI	NON	NON	NON	NON	NON
	Loggia, balcon	OUI	NON	OUI	NON	NON	NON	NON	NON
	Toiture-terrasse	OUI	NON	OUI	NON	NON	NON	NON	NON
	Débords de toiture et modénatures	NON	NON	NON	NON	NON	NON	NON	NON
Piscines	Terrasse surélevée ≥60cm du sol naturel	OUI	NON	OUI	NON	NON	NON	NON	NON
	Terrasse de plain-pied	NON	NON	NON	NON	NON	NON	NON	NON
	Bassin sans couverture	NON	NON	OUI	NON ⁵	NON	NON	NON	NON
	Bassin avec couverture de hauteur ≤1,80m	NON	NON	OUI	NON ⁵	NON	NON	NON	NON
	Bassin avec couverture de hauteur >1,80m	OUI	NON	OUI	NON ⁵	NON	NON	NON	NON
Pourtour bassin avec couverture de hauteur >1,80m	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	

¹ La SHOB et la SHON n'existent plus depuis 2012. Elles sont remplacées par la surface de plancher.

² Sauf dans les bâtiments autres que maisons individuelles.

³ Non aménageable = encombré par charpente ou avec plancher non porteur.

⁴ Sauf si le sous-sol est totalement enterré, sans construction au dessus.

⁵ Les stationnements extérieurs et piscines sont soumis à la TA, bien que hors surface taxable.



Tableau 12 : Récapitulatif des différentes définitions de surfaces de bâtiments

4.3.2 Utilisation de la base de données MAJIC (zone ENS)

Sur la zone ENS, on dispose de la base de données des fichiers fonciers MAJIC (Mise À Jour des Informations Cadastrales), disponible dans le cadre de ce projet de recherche. Cette base de données se décompose en plusieurs tables liées par des identifiants (Figure 49).

Dans le but d'identifier et localiser les surfaces chauffées du quartier, on s'intéressera aux tables de la section « Propriété Bâtis ». Dans cette section, la table 'local' dispose d'informations à l'échelle des locaux d'habitation (logements) ou locaux professionnels. Pour le résidentiel, l'attribut 'Stoth', issu de cette dernière table, est utilisé. Cet attribut représente une surface qui varie légèrement de la surface habitable SHAB, car elle prend en compte les surfaces des espaces dont la hauteur sous plafond est inférieure à 1,80 (voir Figure 50). La fiabilité de cet attribut est jugée « bonne » par le CEREMA⁷ : « Ces données sont par construction très fiables, mais peuvent parfois présenter des décalages par rapport à une confrontation terrain (données déclaratives). Ces décalages peuvent être considérés comme non-significatifs à une échelle communale ». En ce qui concerne les locaux professionnels, l'attribut 'sprobati' de la table 'local' est utilisé. Cet attribut correspond à la « Surface des parties principales et secondaires couvertes du local professionnel (en m²) » ; il est disponible depuis 2017 et sa fiabilité n'a pas encore été évaluée par le CEREMA. Aussi, l'attribut 'ccoape' est utilisé, lequel correspond au code NAF et permet d'identifier l'activité du local professionnel (restaurant, petit commerce...). La fiabilité de cet attribut est jugée bonne par le CEREMA.

Les locaux renseignés dans la base MAJIC, sont géographiquement localisables en les liant à la parcelle à laquelle ils appartiennent, via l'identifiant id_par. En effet, la table 'parcelle' à l'avantage de posséder une colonne de géométries SIG qui permet d'afficher les contours des parcelles. Ainsi, il est possible de superposer les contours des parcelles à la table SIG des formes bâties. La superposition de ces contours avec la table SIG du quartier et la base BD_TOPO®, est présentée en Figure 51. Cette dernière figure montre que pour le tissu urbain dense situé autour du quartier d'étude, le découpage en parcelles est proche du découpage en bâtiments de la BD_TOPO®. Cela permet potentiellement de coupler de manière automatique et

⁷ <http://doc-datafoncier.cerema.fr/ff/docff/>

relativement précise les informations MAJIC aux informations BD_TOPO®. Cependant, pour des quartiers neufs tels que la zone Ouest du quartier de Clichy-Batignolles, on observe que la table parcelles n'est pas à jour (Figure 51). Les parcelles ne correspondent pas aux bâtiments et les données foncières correspondant à ces parcelles sont non renseignées. Dans de tels cas, les bases de données ne permettent pas de caractériser le quartier. Il faudra utiliser d'autres sources d'information telles que les données de l'aménageur (voir paragraphe suivant).

Dans le cas de la zone ENS, le découpage des parcelles correspond aux formes des bâtiments. Cependant, les parcelles englobent souvent plusieurs bâtiments. Ainsi pour localiser plus précisément les surfaces issues des données foncières, il est nécessaire d'utiliser les adresses postales fournies dans la table 'local' de la base de données. En effet dans cette table, sont disponibles les attributs : numéro de voie, libellé de voie, indicatif (bis, ter) et numéro de bâtiment. Il est donc possible de concaténer ces attributs et de grouper les informations par adresse. Les adresses sont ensuite utilisées pour localiser plus précisément les locaux à l'aide de Google Map®. Cependant, cette opération peut encore s'avérer difficile, car sur un quartier récent, les adresses peuvent être non actualisées.

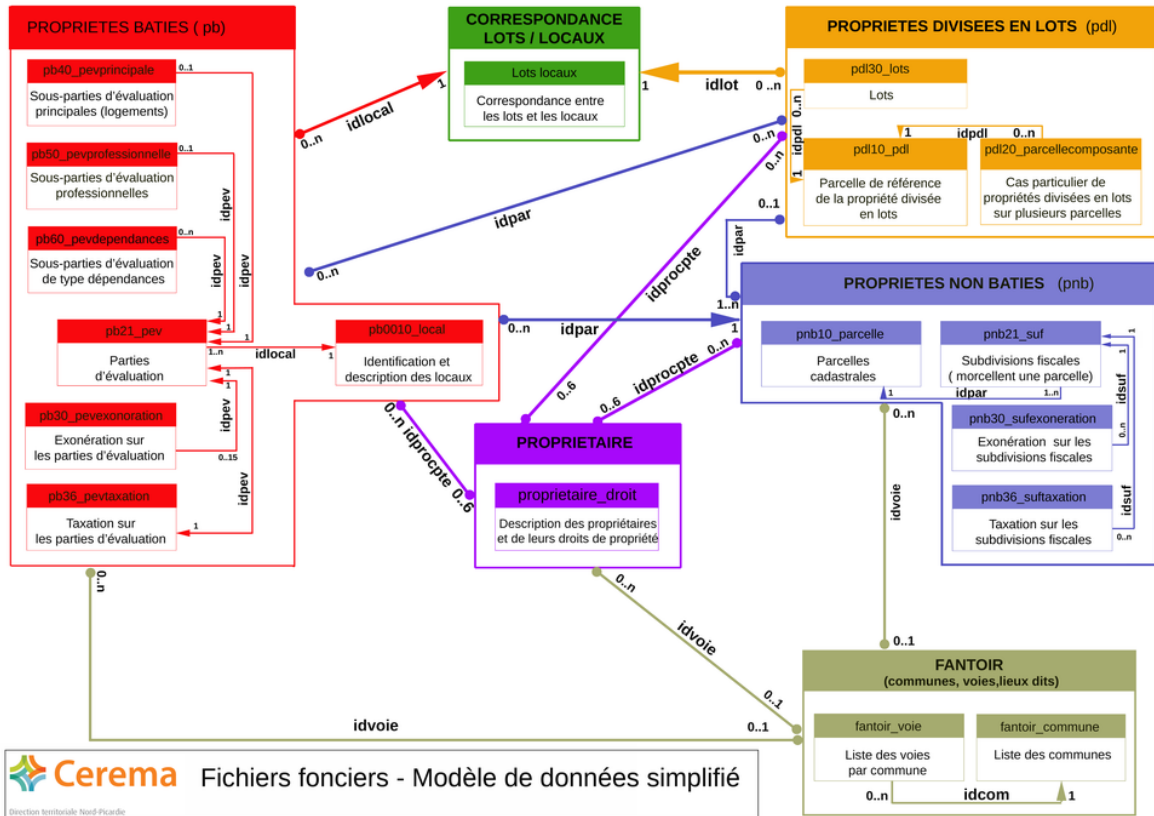


Figure 49 : Représentation de la base de données des fichiers fonciers (Cerema 2017)

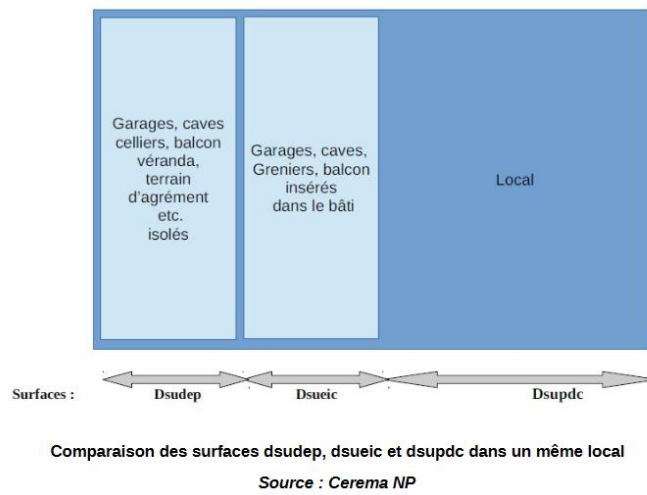


Figure 50 : Comparaison pour un même local, des surfaces : Carrez, habitable et surface réelle (stoth) (Cerema s. d.)

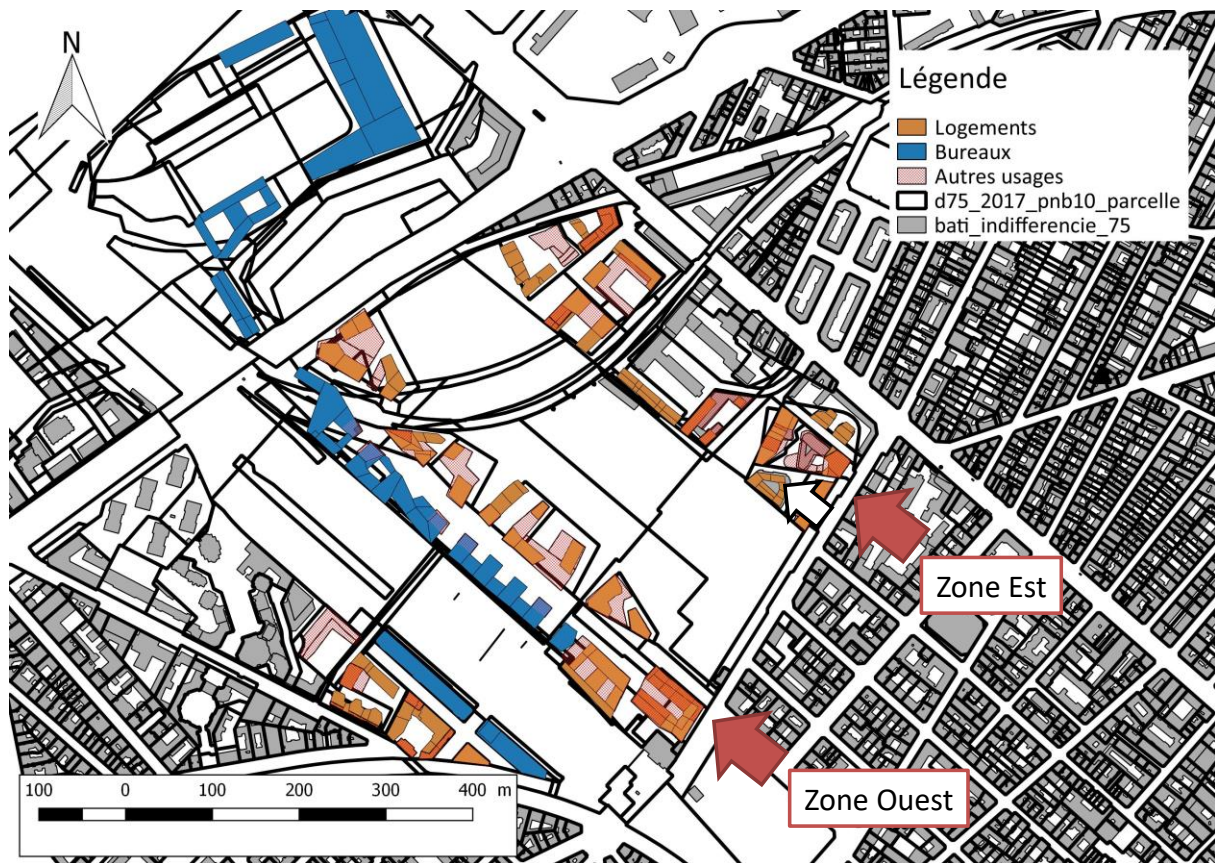


Figure 51 : Superposition des couches SIG : contours de parcelles issus de la base de données MAJIC, base de données BD_TOPO® et modèle SIG du quartier

4.3.3 Utilisation des documents de construction (zone Ouest)

Sur la zone Ouest, les bâtiments sont très récents, il n'est ainsi pas possible de s'appuyer sur des bases de données pour paramétrer le modèle de quartier. Cependant, dans le cadre du projet CoRDEES, il a été possible de récupérer des documents de construction, fournis par l'aménageur :

- rapports de simulations thermiques dynamiques STD,
- plans de bâtiments,
- documents d'aménageur décrivant des logements (surfaces, nombre de pièces)
- rapports de délibération communale sur les programmes de logements sociaux.

En recoupant l'ensemble de ces documents, il a été possible d'identifier et localiser l'ensemble des locaux de la zone, caractérisés par des surfaces chauffées et des usages. Parfois, ce n'est pas la surface SHAB qui est renseignée mais, la surface utile, ou la surface SHON ; dans ce cas, la surface SHAB est estimée à partir de la surface renseignée, multipliée par un coefficient de correction calculé sur les bâtiments

où les deux types de surface sont disponibles. Il arrive que le type de surface renseigné dans un document ne soit pas précisé. Dans ces cas, ce dernier est estimé au mieux, en fonction des indices disponibles.

4.3.4 Lien entre les locaux et la table SIG

Les locaux, caractérisés par des surfaces chauffées et des usages, sont liés aux informations contenues dans la table SIG, selon la méthode illustrée en Figure 52. Comme vu dans les paragraphes précédents, la localisation des locaux n'est pas toujours précise. Pour lier ces informations, une échelle intermédiaire est créée en regroupant des formes géométriques de la table SIG. Ces regroupements, appelés « zones », sont créés à la fois en fonction des informations issues de la table SIG, et des informations issues de la table des surfaces chauffées. Les contraintes menant à la création de ces zones sont illustrées en Figure 52:

- Les locaux de logements 1, 2, 3 et le local de bureaux diffus 4, sont localisés quelque part dans les formes 1, 2, 3, 4. Selon les formes géométriques, il est préférable de créer au moins deux zones : 1 et 2. Ces deux zones ont des formes, orientations et donc comportements thermiques différents. Les locaux sont répartis entre les deux zones au prorata des surfaces et des volumes bâtis.
- Le local de bureaux 5 est localisé dans la forme SIG 5.
- Les locaux de commerce 6 et 7 sont localisés dans la forme 6.

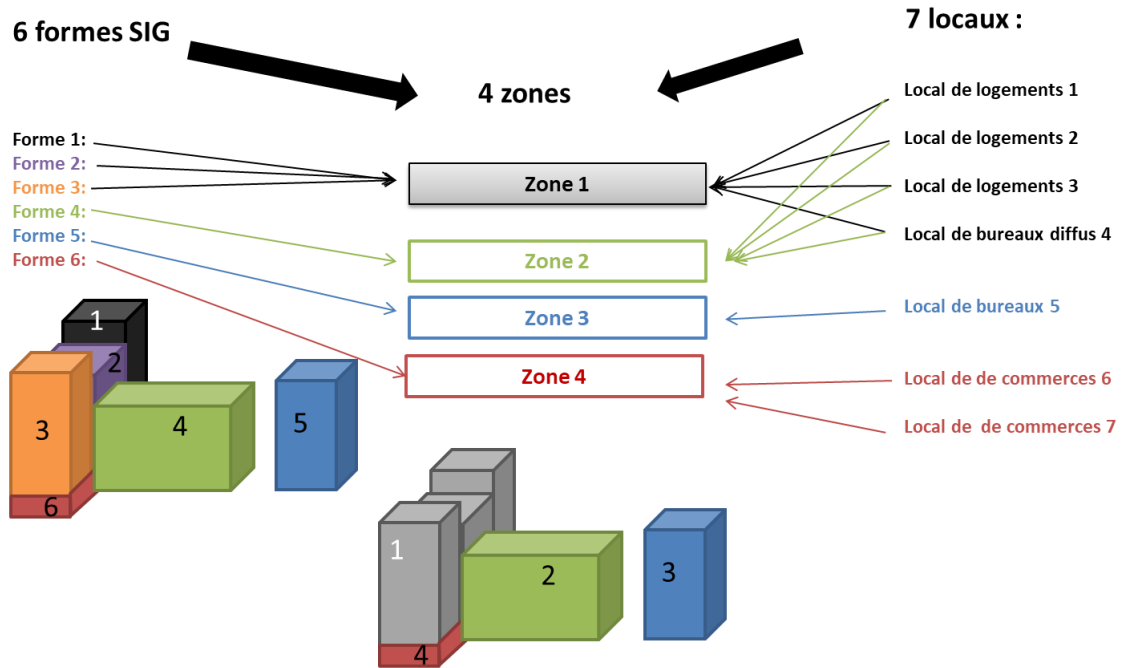


Figure 52 : Illustration de la méthode liant les surfaces chauffées aux informations contenues dans la table SIG

Pour chaque zone, les surfaces de déperditions et les apports solaires sont calculés selon la méthode présentée dans les paragraphes 4.2.3 et 4.2.4. Ces zones pourront par la suite être modélisées par des modèles ISO_CpC, mais de nombreuses données d'entrée restent à estimer.

4.4 Enrichissement statistique : application aux bâtiments de bureaux

Une fois qu'ont été identifiées les surfaces chauffées par usage, les gains solaires, et les surfaces de déperdition, il reste à estimer :

- Les propriétés thermiques : coefficient de déperdition thermique U , Résistance d'infiltration R_{inf} , et paramètres liés à la capacité thermique : C_m et A_m (selon la norme (AFNOR 2013))
- Les apports thermiques liés au métabolisme aux occupants : Φ_{meta}
- Les apports thermiques liés aux consommations d'électricité spécifique : Φ_{elec}
- La résistance dynamique de ventilation mécanique : R_{meca}
- Des scénarios d'utilisation des stores extérieurs
- Des scénarios de température de consigne

Ces dernières années, des bases de données, dites d'enrichissement, ont été développées pour estimer statistiquement ces données d'entrée et alimenter des logiciels de simulation thermique de bâtiments à l'échelle territoriale. La plateforme SMART-E (Berthou et al. 2015) permet d'estimer ces entrées pour le secteur résidentiel, à partir d'informations telles que les profils des habitants et les années de construction ou de rénovation des bâtiments. De même, le logiciel CEA (City Energy Analyst) dispose d'une base de données open source, permettant pour chaque usage (résidentiel, bureaux, restaurant, hôtel, hôpitaux, école...) d'estimer ces données d'entrée à partir des années de construction et de rénovation.

Cependant, ces bases d'enrichissement sont à considérer avec précaution : peu de références et de justifications sont disponibles. De plus, la base CEA a été réalisée pour le cas de la Suisse ; il sera alors nécessaire de faire évoluer cette base à partir d'études adaptées au cas de la France. Les bases de données et les modèles de la plateforme CEA ont été mis en libre accès sur GitHub dans le but d'un développement collaboratif. Une thèse visant au développement d'une méthode de validation de ces bases de données d'enrichissement est actuellement en cours (Lefort et al. 2021).

Malgré ces limites, la méthode développée dans cette thèse propose de s'appuyer sur ces bases de données pour l'enrichissement statistique de la base SIG. Le logiciel SMART-E sera privilégié pour le secteur résidentiel ; pour les autres secteurs, la base CEA pourra être utilisée. Pour les bâtiments de bureaux, une méthode spécifique est proposée dans les paragraphes suivants.

4.4.1 Classification des bâtiments de bureaux par types

Afin de mieux caractériser les bâtiments de bureaux, une classification de ces derniers par type est proposée. Dans le cadre du projet « Objectif 100 » (Alessandrini et al. 2003), trois types de bâtiments de bureaux, correspondant à des comportements thermiques différents, sont proposés (Tableau 13). La première catégorie correspond à de grands bâtiments de bureaux, fortement vitrés. La deuxième catégorie correspond à des bâtiments de tailles et taux de vitrage plus faibles, rencontrés principalement en zone urbaine et pouvant faire l'objet de rénovations, tels que les bâtiments dits « néo haussmannien ». Enfin, la dernière catégorie correspond à de petits bâtiments, majoritairement rencontrés en zones péri-urbaines. Les différents types de bureaux

peuvent être identifiés visuellement ou éventuellement à partir de leur catégorie de surface.

Type de bâtiment	1	2	3
Descriptif	Grande Surface, mur rideau, plateau large et salle de réunion aveugle, bureaux paysagés	Immeuble rénové, taille intermédiaire, structure plutôt lourde et bureaux cloisonnés	Immeuble de petite taille (péri urbain...)
Surface totale utile	15000 m ²	5000 m ²	1000 m ²
Hauteur sous plafond	3 m	3 m	2,7m

Tableau 13 : Typologies des bâtiments de bureaux (Alessandrini et al. 2003)

4.4.2 Classe d'inertie

Une classe d'inertie peut être attribuée à chaque bâtiment de bureaux selon leurs catégories (Tableau 14). Ces classes d'inertie font référence à la norme (AFNOR 2013) et permettent d'estimer des valeurs par défaut de C_m et A_m (Tableau 15).

Type de bâtiment	1	2	3
Classe d'inertie	Légère	Moyenne	Moyenne

Tableau 14 : Classe d'inertie des bâtiments de bureaux selon leurs typologies (Alessandrini et al. 2003)

Classe	A_m (m ²)	C_m (J/K)
Très léger	$2,5 S_{ch}$	$80000 S_{ch}$
Léger	$2,5 S_{ch}$	$110000 S_{ch}$
Moyen	$2,5 S_{ch}$	$165000 S_{ch}$
Lourd	$3 S_{ch}$	$260000 S_{ch}$
Très lourd	$3,5 S_{ch}$	$370000 S_{ch}$

Tableau 15 : Valeurs par défaut de C_m et A_m selon la classe d'inertie du bâtiment (AFNOR 2013)

4.4.3 Coefficients de déperditions thermiques

L'estimation des coefficients de déperditions thermiques (U) se fait selon la méthode 3CL-DPE. Cette méthode est déjà utilisée dans le logiciel SMART-E ; elle a été développée pour le résidentiel, mais s'appuie sur l'évolution des réglementations thermiques qui s'appliquent à tout type de bâtiment sans distinction d'usage. Il a

donc été choisi d'étendre cette méthode aux bâtiments de bureaux. Les coefficients de déperditions thermiques des murs, des planchers, des toits et des fenêtres sont estimés à partir des informations suivantes :

- années de construction,
- années de rénovation (si l'isolation du bâtiment a été améliorée),
- type de toiture (comble ou terrasse),
- type de plancher (terre-plein, local non chauffé).

Les ponts thermiques sont pour le moment pris en compte de manière forfaitaire selon la norme (AFNOR 2017b). Cette méthode pourra, par la suite, être affinée en utilisant des coefficients linéiques de déperdition, et en calculant, à partir de la table SIG, des longueurs de lignes de déperditions thermiques.

La singularité des bâtiments de type 1 est prise en compte. Ces bâtiments sont, pour la quasi-totalité, réalisés en façades rideaux, qui correspondent à des façades légères, non-porteuses, souvent composées de panneaux vitrés, ajoutées sous forme de panneaux préfabriqués. La plupart du temps, une partie des panneaux est opaque, cette partie peut être isolée de l'intérieur. Autant pour la partie opaque que pour la partie transparente, les réglementations thermiques s'appliquent de la même manière que pour les logements. Par contre, un niveau technologique et un taux de rénovation plus élevé que pour les logements, ainsi que de fortes contraintes liées à la sécurité, impliquent l'utilisation de vitrages spécifiques. Les coefficients de déperdition des parois vitrées sont alors fixés à 2 W/K.m^2 pour les bureaux de type 1.

Les établissements de bureaux, de types 2 et 3, sont supposés présenter des architectures similaires aux bâtiments de logements. Ainsi, les réglementations thermiques et techniques disponibles sont supposées être les mêmes que pour les logements.

4.4.4 Infiltration

Pour rappel, la résistance d'infiltration est calculée à partir du débit d'infiltration \dot{V}_{inf} , suivant l'équation suivante :

$$R_{inf} = \frac{1}{U_{inf}} = \frac{3600}{\rho_{air} c_{air} \dot{V}_{inf} V_{ch}} \quad (109)$$

Le débit d'infiltration d'air est estimé par une méthode simplifiée issue de la norme (AFNOR 2004). Cette méthode repose sur l'équation (110). Le débit

d'infiltration est donné en fonction du volume d'air total (V_{ch}), du taux horaire d'infiltration d'air à 50 Pa (n_{50}), et des coefficients 'e' et 'ε', permettant la prise en compte de la hauteur des bâtiments et à l'exposition au vent du bâtiment (Tableau 16 et Tableau 17). Dans la norme (AFNOR 2017a), des valeurs par défaut du taux de renouvellement d'air (n_{50}) sont données selon l'âge des bâtiments (Tableau 18). Pour les établissements de type 1, afin de représenter la perméabilité des façades rideau, le n_{50} est fixé à $2 h^{-1}$ ⁸.

$$\dot{V}_{inf} = 2 \cdot V_{ch} \cdot n_{50} \cdot e \cdot \varepsilon \quad (110)$$

- n_{50} : taux horaire de renouvellement d'air (h^{-1}), pour une pression de référence 50 Pa
- e : coefficient d'exposition au vent;
- ε : facteur correctif de hauteur, prenant en compte l'augmentation de la vitesse du vent avec la hauteur de l'espace considéré au-dessus du niveau du sol.

Hauteur	ε
0->10m	1
10->30m	1,2
>30	1,5

Tableau 16 : Facteur correctif ε (AFNOR 2004)

Exposition	e
Site non abrité (bâtiments en zone ventée, bâtiments de grande hauteur en centre-ville)	0,05
Site modérément abrité (bâtiments en campagne protégés par des arbres ou par d'autres bâtiments, banlieues)	0,03
Site très abrité (bâtiments de taille moyenne en centre-ville, bâtiments en forêt)	0,02

Tableau 17 : Facteur correctif e (AFNOR 2004)

⁸ <http://www.fenetrealu.com/uploads/documents/influence-fenêtres-façades-rideaux-aluminium-permeabilite-des-batiments-158.pdf>: « Le classement moyen des parties fixes des façades rideaux est au minimum 4 fois moins déperditif que le A*4 des fenêtres, soit moins de $0,10 \text{ m}^3 / \text{h} / \text{m}^2$ sous 4 Pa (environ 2 Vol/h à 50 Pa). Ainsi, les débits de fuite de l'enveloppe de bâtiments tertiaires largement vitrés sont extrêmement faibles. »

Année de construction ou rénovation	n50
<i>année</i> ≥ 1995	$n50 \leq 3 \text{ h}^{-1}$
1977 ≤ <i>année</i> ≤ 1995	$3 \text{ h}^{-1} \leq n50 \leq 6 \text{ h}^{-1}$
<i>année</i> ≤ 1977	$n50 \geq 6 \text{ h}^{-1}$

Tableau 18 : Taux de renouvellement de l'air par default (AFNOR 2017a)

L'estimation des débits d'infiltration pourra par la suite être améliorée, notamment grâce à la méthode disponible dans (ASHRAE 2013). Celle-ci permet de prendre en compte la vitesse du vent qui peut varier au cours du temps. En plus des infiltrations par les murs verticaux, les établissements de bureaux peuvent être soumis à l'infiltration issue des cages d'escalier, cages d'ascenseur et portes d'entrée coulissantes. Ces infiltrations sont pour le moment négligées, elles pourront par la suite être étudiées et ajoutées à l'infiltration des murs. Des informations sont disponibles dans (ASHRAE 2013).

4.4.5 Profils d'occupation

Les profils d'occupation des bâtiments de bureaux, sont estimés à partir de « L'enquête Emploi du temps » réalisée par l'Insee. Parmi les questionnaires distribués, se trouve un semainier de travail à destination des personnes considérées comme actives (Figure 53). Dans ce questionnaire, sont renseignées les plages horaires travaillées au cours d'une semaine. 4726 semainiers ont été collectés. Ces données constituent la table « Carnet » de l'enquête Emplois du temps. Les répondants remplissent, sur sept jours consécutifs, leurs horaires de travail, au pas de temps 15 min, hors temps de repas, pauses et temps de trajets du domicile au lieu de travail. La collecte s'est déroulée sur un an, de septembre 2009 à septembre 2010, afin de prendre en compte les variations saisonnières des emplois du temps.

Exemple de semainier :

Jour	0h	1h	2h	3h	4h	5h	6h	7h	8h	9h	10h	11h	12h	13h	14h	15h	16h	17h	18h	19h	20h	21h	22h	23h
Jour 1																								
Jour 2																								
Jour 3																								
Jour 4																								
Jour 5																								
Jour 6																								
Jour 7																								

Figure 53 : Exemple d'un semainier rempli dans le cadre de l'Enquête emplois du temps de l'INSEE

La table « carnet » peut être liée à la table « individu » du fichier Emplois du temps, renseignant notamment le type de la journée enquêtée (Tableau 19). Les emplois du temps des individus ayant renseigné une semaine dite « normale » de travail, sont moyennés ; il en résulte un profil hebdomadaire d'occupation. Ce profil hebdomadaire est répété plusieurs fois, de manière à construire un profil d'occupation annuel : Occ_0 (congrés, et arrêts maladie, non pris en compte).

Remarque : Il est possible de sélectionner des semainiers liés à des individus travaillant dans un secteur d'activité spécifique (au sens du code NAF A38).

Pour chaque mois, un facteur correctif de prise en compte des congés est calculé. A partir de l'information sur le type de la journée enquêtée, disponible dans la table « individu ». La répartition mensuelle, moyenne, des jours de congés pris sur une année, est calculée en divisant les nombres de jours de congés pris pendant chaque mois, par le nombre total des jours de congés pris pendant l'année (Tableau 20). Les 9 derniers jours du mois de décembre sont traités séparément afin de prendre en compte les congés de fin d'année. Ces clés de répartition sont ensuite utilisées pour calculer des facteurs correctifs de prise en compte des jours de congés :

$$r_{congés_m} = \frac{(1 - \frac{f_{c_m}}{100}) * n_c}{n_{j_m}} \quad (111)$$

- f_{c_m} : clé de répartition mensuelle des jours de congés (Tableau 20)
- n_c : nombre de jours de congés moyen sur une année (fixée à 40)
- n_{j_m} : nombre de jours ouvrés dans le mois m
- $r_{congés_m}$: facteur correctif de prise en compte des congés

Comme pour les jours de congés, un facteur correctif de prise en compte des jours d'arrêt maladie est calculé selon l'équation (112):

$$r_{maladie_m} = \frac{(1 - \frac{f_{m_m}}{100}) * n_m}{n_{j_m}} \quad (112)$$

- f_{m_m} : clés de répartition mensuelle des jours de congés (Tableau 21)
- n_m : nombre de jours d'arrêt maladie moyen sur une année (fixé à 11)⁹
- n_{j_m} : nombre de jours ouvrés dans le mois m

Le profil d'occupation moyen, prenant en compte les jours de congés et d'arrêt maladie, est obtenu selon l'équation (113), et est présenté sur les Figure 54 et Figure 55.

$$Occ = r_{maladie_m} r_{congés_m} Occ_0 \quad (113)$$

Remarque : Ce travail a été réalisé avant l'explosion du télétravail dû à la crise sanitaire du COVID19.

1	Journée normale de travail
2	Journée semi-travaillée
3	Jour de repos légal ou férié non travaillé
4	Jour de repos légal ou férié travaillé
5	Jour de congé maladie
6	Journée de congé (vacances)
7	Autres cas (pour les personnes n'exerçant pas d'activité professionnelle)

Tableau 19 : Attributs concernant le type de la journée enquêtée, issus de la table individu de l'enquête emplois du temps

⁹http://www.malakoffmederic.com/groupe/blobs/medias/s/2bdbcc541820004d/2015-09-08-CP_Etude-Absenteisme-.pdf: « Le nombre de jours d'absence par salarié est de 11,4 jours en 2015 »

Mois	Clés de répartition mensuelle des jours de congés	Mois	Clés de répartition mensuelle des jours de congés
1	3,9 %	7	10,3 %
2	6,3 %	8	28,8 %
3	5,6 %	9	3,7 %
4	11,8 %	10	8,9 %
5	4,2 %	11	3,8 %
6	4,4 %	12.1	3,5 %
		12.2	4,7 %

Tableau 20 : Répartition mensuelle des jours de congés pris par les actifs

Mois	Répartition des jours d'arrêts maladie	Mois	Répartition des jours d'arrêts maladie
1	4,9 %	7	10,3 %
2	5,1 %	8	8,1 %
3	9,3 %	9	6,3 %
4	6,0 %	10	9,3 %
5	6,8 %	11	10,7 %
6	10,4 %	12	12,9 %

Tableau 21 : Répartition mensuelle des jours d'arrêt maladie pris par les actifs.

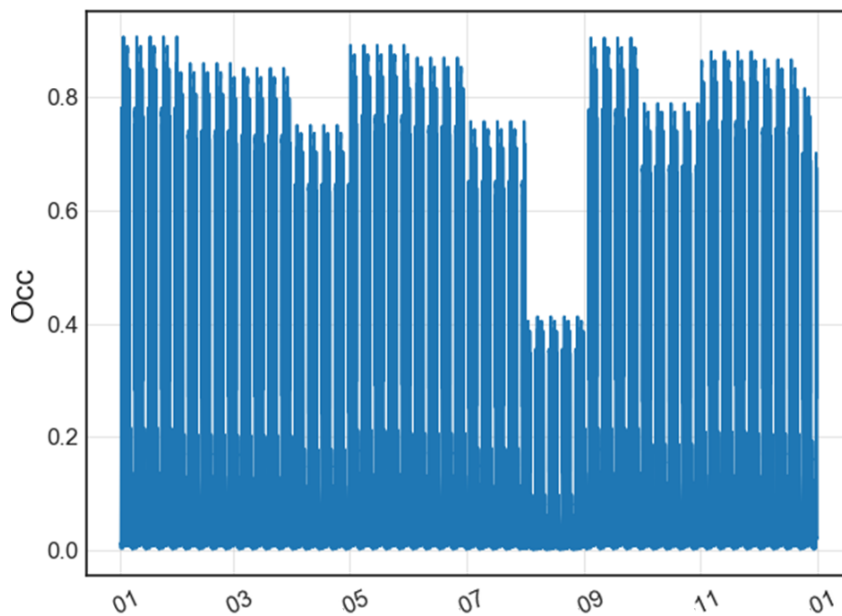


Figure 54 : Profil d'occupation des bureaux (*Occ*) sur une année

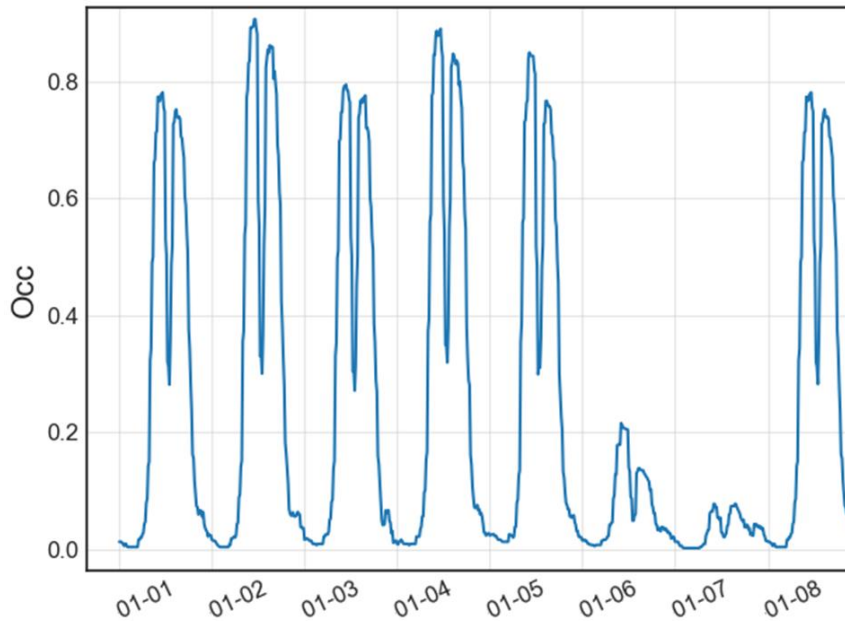


Figure 55 : Profil d'occupation des bureaux (Occ) sur une semaine

4.4.6 Apports thermiques métaboliques liés aux occupants

Les apports thermiques engendrés par le métabolisme des occupants (φ_{meta}) sont modélisés selon l'équation (114), à partir du profil d'occupation Occ , et d'un ratio de m^2 de surface chauffée par occupant : r_{pers} . Chaque personne présente dans le bâtiment est supposée émettre une puissance thermique sensible de 80 W :

$$\varphi_{meta} = Occ * 80 * \frac{S_{ch}}{r_{pers}} \quad (114)$$

L'étude (ARENE 2009), réalisée sur l'Île-de-France, fournit des estimations du ratio r_{pers} selon la taille des établissements et la localisation (voir Tableau 22).

	Moins de 1000 m ²	Entre 1000 et 5000 m ²	Plus de 5000 m ²
Paris	27	39	34
1 ^{ère} couronne	25	34	40
2 ^{ème} couronne	24	29	47
Ensemble	25	34	41

Tableau 22 : Répartition selon la taille et le lieu géographique des ratios de surfaces chauffées par personne : r_{pers} (ARENE 2009)

4.4.7 Apports thermiques liés aux consommations d'électricité spécifique

Les scénarios décrivant les apports thermiques engendrés par les consommations d'électricité spécifique sont modélisés selon l'équation (115). Un paramètre issu de la base de données CEA (r_{elec}), représentant le pic des apports thermiques liés aux consommations d'électricité spécifique, par m^2 de surface chauffée, est multiplié par le profil d'occupation du bâtiment. Un talon de consommation, indépendant de l'occupation, et égale à 10 % du pic de consommation, est considéré :

$$\varphi_{elec} = \left(\frac{Occ}{\max(Occ)} S_{ch} \cdot r_{elec} \cdot 0,9 + r_{elec} \cdot S_{ch} \cdot 0,1 \right) \quad (115)$$

4.4.8 Ventilation mécanique

Lorsqu'elles ne sont pas connues, les pertes thermiques par ventilation sont modélisées selon l'équation (116). La résistance thermique de ventilation est calculée en fonction du débit de ventilation nominal \dot{V}_{meca} et des heures de mise en marche et d'arrêt du système de ventilation : $h1$ et $h2$.

$$R_{meca} = \begin{cases} \frac{3600}{\dot{V}_{meca} V_{ch} \rho_{air} c_{air}}, & t \in [h1 : h2] \text{ et } t \in [lundi : vendredi] \\ 0, & \text{sinon} \end{cases} \quad (116)$$

Le débit de ventilation a priori \dot{V}_{meca} (V_{ch}/h) est issu de la base de données CEA. Une ventilation mécanique est présente dans la plupart des établissements de bureaux. Par défaut, dans les bureaux, la ventilation mécanique est supposée fonctionner, les jours ouvrés, de 6h30 ($h1$) à 19h30 ($h2$). Dans le cas où le système de régulation de la ventilation n'est pas connu, la méthode simplifiée disponible dans la norme (AFNOR 2004) suggère de prendre comme débit d'air moyen total, le maximum entre le débit d'infiltration moyen et le débit de ventilation.

4.4.9 Scénario de stores extérieurs (apports solaires)

Les apports solaires internes (pénétrant dans le bâtiment par les surfaces vitrées) dépendent fortement des protections solaires utilisées. La présence ou non de stores extérieurs peut être connu, mais le scénario d'utilisation dépend généralement du comportement des occupants et reste ainsi incertain. Afin d'estimer un scénario a

priori d'utilisation des stores extérieurs, il est fait appel à l'étude (Alessandrini et al. 2006) qui présente une corrélation empirique entre l'éclairement reçu sur la façade d'un bâtiment de bureaux (E), et le pourcentage d'obstruction des parois vitrées (indice de protection solaire : IPS). Cette corrélation est définie par les deux équations suivantes :

$$IPS = \begin{cases} \frac{0,25 - 0,07}{28000} \cdot E + 0,07, & \text{pour } E < 28000 \text{ lux} \\ \frac{0,45 - 0,25}{100000 - 28000} \cdot E + 0,17, & \text{pour } E > 28000 \text{ lux} \end{cases} \quad (117)$$

4.4.10 Scénario de températures de consigne

Lorsque la température intérieure moyenne du bâtiment n'est pas connue, une étape de régulation est ajoutée au modèle. De nombreux modèles de régulation sont possibles et les paramètres du système de régulation pourraient être calibrés. Cependant, pour simplifier le problème, il est fait l'hypothèse d'une régulation parfaite, avec prise en compte d'une saturation à une puissance maximale : Φ_{max} , qui peut être estimée en observant le profil des données consommation. La puissance Φ_{heat}^k nécessaire au pas de temps k pour atteindre la température demandée T_{cons}^{k+1} au pas de temps $k+1$, est calculée à chaque itération selon les équations suivantes :

$$\begin{pmatrix} Ta = T_{cons} \\ Tp \end{pmatrix}^{k+1} = A_d \begin{pmatrix} Ta \\ Tp \end{pmatrix}^k + B_d \begin{pmatrix} \varphi p = \varphi_{int} + \Phi_{heat_opt} + \varphi_{sol_in} \\ \varphi a \\ \varphi l \\ \varphi f \\ \varphi h \\ Te \end{pmatrix}^k \quad (118)$$

$$\Phi_{heat}^k = \min(\Phi_{heat_opt}^k, \Phi_{max}) \quad (119)$$

Des scénarios a priori de température de consigne sont modélisés par une température nominale Th_0 et d'une température de réduit Thr_0 :

$$T_{cons} = \begin{cases} Th_0, & \text{en période d'occupation } (occ > 0,1) \\ Thr_0, & \text{hors occupation } (occ < 0,1) \end{cases} \quad (120)$$

Pour le cas d'un bâtiment de bureau en période hivernale, les changements de température de consigne peuvent être observés sur la courbe des consommations de chauffage.

4.5 Exemples de modèle a priori sur le quartier de Clichy-Batignolles

La méthode de paramétrage a priori des modèles RC présentée dans ce chapitre est testée sur deux bâtiments du quartier de Clichy-Batignolles, pour lesquels des données consommation de chauffage, issues d'une sous-station du réseau de chaleur, ont pu être récupérées : un bâtiment résidentiel et un bâtiment de bureaux. Pour cause de confidentialité des données, ces bâtiments doivent rester anonymes.

Pour ces deux cas d'études, les données de puissance de chauffage sont comparées aux résultats de simulation issus des modèles a priori; d'abord à l'échelle mensuelle (Figure 56 et Figure 58), puis à l'échelle horaire (Figure 57 et Figure 59)

Il ressort de cette comparaison que les erreurs mensuelles sont relativement faibles (de l'ordre de 10 %). Ces modèles a priori semblent alors en mesure d'être utilisés pour estimer des ordres de grandeur de gains potentiels d'énergie, associés à des scénarios de rénovation des bâtiments d'un quartier.

Les profils horaires présentent en revanche des différences importantes. Cela est dû aux fortes incertitudes sur les paramètres dynamiques du modèle RC (Cp et Ap), et sur les entrées dynamiques, telles que le profil de température de consigne. Ces modèles a priori ne semblent donc pas permettre l'étude de scénarios de flexibilité ; pour cela de modèles STD de bâtiments types seront probablement préférables car plus en mesure de représenter les effets d'inertie thermique.

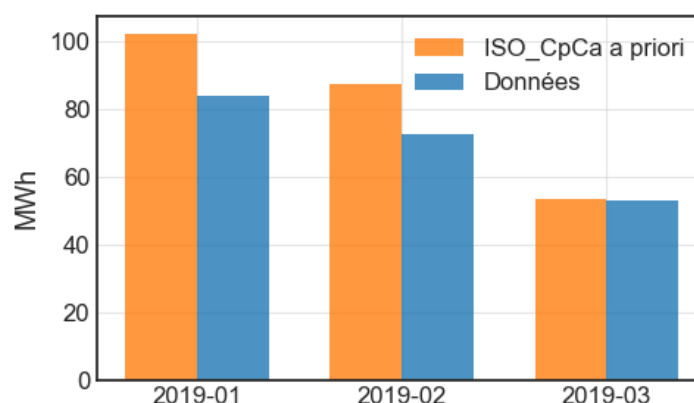


Figure 56 : Comparaison, sur un bâtiment de bureaux du quartier de Clichy-Batignolles, de données de consommation mensuelles avec les résultats de simulation du modèle ISO_CpCa a priori

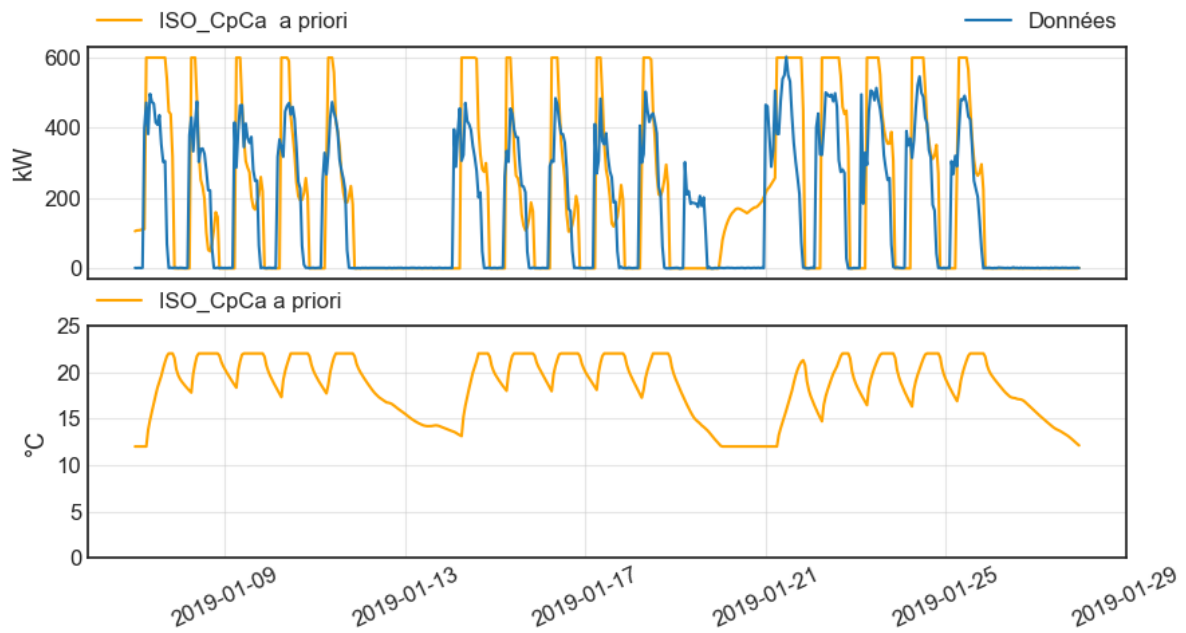


Figure 57: Comparaison, sur un bâtiment de bureaux du quartier de Clichy-Batignolles, de données de consommation horaires avec les résultats de simulation du modèle ISO_CpCa a priori

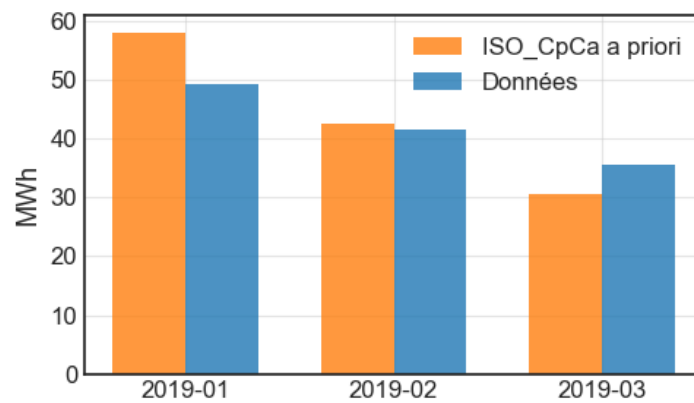


Figure 58 : Comparaison, sur un bâtiment résidentiel du quartier de Clichy-Batignolles, de données de consommation mensuelles avec les résultats de simulation du modèle ISO_CpCa a priori

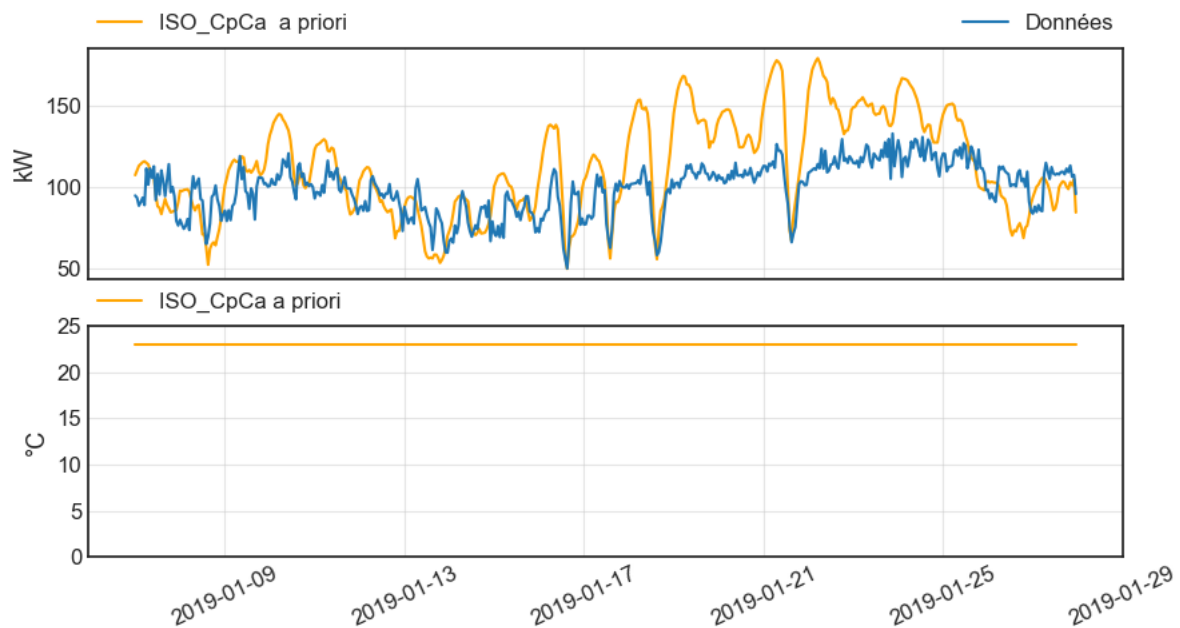


Figure 59 : Comparaison, sur un bâtiment résidentiel du quartier de Clichy-Batignolles, de données de consommation horaires avec les résultats de simulation du modèle ISO_CpCa a priori

4.6 Conclusion

Dans ce chapitre une méthode de construction d'une base de données pour le paramétrage a priori d'un modèle de quartier est proposée. Cette méthode se décompose en 3 parties : construction d'une table de données SIG, estimation et localisation des surfaces chauffées par usages, et enrichissement statistique. La méthode s'appuie sur l'utilisation de bases de données existantes (MAGIC, BD_TOPO) et de données issues de plateformes de simulation de quartiers (CEA, SMART-E). Les limites à l'utilisation de ces bases de données sont mises en avant, et la méthode est complétée par l'utilisation d'autres sources d'informations, telles que des documents de construction, des images aériennes, et des études statistiques supplémentaires.

La diversité des quartiers et des sources d'informations disponibles, rendent le développement d'une telle méthode particulièrement difficile ; il existe encore une forte marge d'amélioration, notamment pour des usages moins fréquents et moins étudiés que le résidentiel, tels que les bâtiments académiques ou les commerces de bas d'immeuble.

Pour les bâtiments de bureaux et de logements, le modèle ISO_CpCa a priori paramétrés selon cette méthode, apparaît capable de simuler des consommations mensuelles de chauffage avec une incertitude de l'ordre de 10 %. Ce modèle a priori

semble en mesure d'être utilisés pour estimer des ordres de grandeur des gains potentiels d'énergie, associés à des scénarios de rénovation. Ce dernier pourrait aussi être utilisé pour simuler des scénarios de sobriété tels que la réduction des températures de consigne ; et des modèles de système tels que des pompes à chaleur ou des panneaux solaires, pourront être ajoutés afin d'étendre la diversité des scénarios de performance énergétique modélisables. En revanche, les modèles RC paramétrés a priori semblent peu adaptés pour l'étude des scénarios de flexibilité étant donné les problèmes d'identifiabilité des paramètres A_p et C_p et les fortes incertitudes sur les données d'entrée dynamiques.

Le paramétrage des modèles pourrait être amélioré par calibration à partir de données d'exploitation afin d'évaluer des actions de performance énergétique et de faire du suivi de performance. L'apport de la calibration du modèle sera étudié dans le chapitre suivant.

Chapitre 5

Application de la méthode de calibration bayésienne

Table des matières

Chapitre 5.....	141
5.1 Introduction.....	143
5.2 Discussion sur le choix de la fonction de vraisemblance.....	143
5.2.1 Fonction de vraisemblance <i>Li. i. d</i>	144
5.2.2 Fonction de vraisemblance <i>LGP</i>	147
5.2.3 Fonction de vraisemblance <i>LKal</i>	149
5.2.4 Conclusion sur le choix de la fonction de vraisemblance.....	153
5.3 Discussion sur le nombre de capteurs de température	153
5.4 Identification des capacités et des limites de la méthode de calibration de modèles RC selon les données disponibles.....	156
5.4.1 Estimation des incertitudes a priori	158
5.4.2 Analyse des modèles a posteriori.....	163
5.4.3 Identification de dérives de consommation.....	166
5.4.4 Identification des causes des dérives de consommation	169
5.5 Conclusion	172

5.1 Introduction

L'objectif de ce chapitre est d'étudier les possibilités et les limites de l'utilisation de la méthode de calibration bayésienne de modèles RC pour répondre aux différents objectifs de la thèse : étude de scénarios de performance énergétique, évaluation d'actions de performance énergétique mises en place (IPMVP), identification de dérives de consommation et des causes de ces dérives. En plus des paramètres du modèle RC, la méthode de calibration bayésienne nécessite la prise en compte des incertitudes liées aux données d'entrée dynamiques, telles les apports de chaleur liés à l'ensoleillement et aux appareils électriques. Après avoir discuté du choix de la fonction de vraisemblance et du nombre de capteurs de température utilisés, des cas d'étude dans lesquels les données disponibles varient sont analysés dans le but d'identifier ce qu'il est possible d'attendre de la méthode de calibration, selon les données disponibles. Les cas d'études sont issus du modèle TRNSYS d'un bâtiment de bureaux, présenté dans le chapitre 3.

5.2 Discussion sur le choix de la fonction de vraisemblance

Dans le chapitre 2, trois fonctions de vraisemblance ont été envisagées. Dans les paragraphes suivants, le modèle ISO_CaCp est calibré en utilisant successivement ces différentes fonctions de vraisemblance. La puissance thermique est considérée en entrée du modèle, les données de calibration sont les températures intérieures. Le cas d'étude considéré correspond au scénario « Base » présenté dans le chapitre 3. Pour cette étude, il est fait l'hypothèse d'une très bonne connaissance du bâtiment. L'ensemble des sollicitations sont supposées connues, seules les incertitudes liées aux paramètres du modèle RC sont à prendre en compte.

Dans le chapitre 3, les paramètres $Rate_m$ et $Rate_l$ sont apparus difficiles à identifier et peu influants sur le modèle, ces derniers seront donc fixés. L'incertitude sur les paramètres Rg , Cp et Ap est prise en compte via des probabilités a priori selon les équations suivantes :

$$\begin{aligned} Rg &= Rg_0 * Rg^* & Rg^* &\sim N(1 ; 0,3) \text{ (sans unité)} \\ Cp &= Cp_0 * Cp^* & Cp^* &\sim N(1 ; 0,3) \text{ (sans unité)} \\ Ap &= Ap_0 * Ap^* & Ap^* &\sim N(1 ; 0,3) \text{ (sans unité)} \end{aligned} \tag{121}$$

Rg_0, Cp_0, Ap_0 : estimations physiques a priori

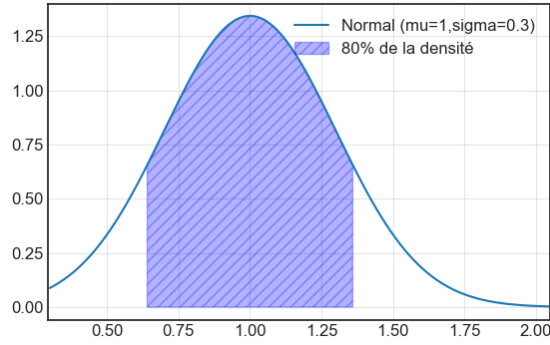


Figure 60 : Loi de probabilité normale $N(1; 0,3)$

5.2.1 Fonction de vraisemblance $L_{i.i.d}$

La fonction de vraisemblance $L_{i.i.d}$ est rappelée en équation (110). Les erreurs entre les données observées et les résultats de simulation sont modélisées par des lois normales indépendantes, centrées en zéro et d'écart-type σ_g . Une loi de probabilité a priori Gamma est associée au paramètre σ_g (Figure 61). Les distributions a priori utilisées sont résumées en Tableau 23.

$$L_{i.i.d} = \prod_{k=1}^n \frac{1}{\sigma_g \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{D_k - \eta_k(\theta, x)}{\sigma_g} \right)^2} \quad (122)$$

- D_k : données de température intérieure moyenne
- $\eta_k(\theta)$: simulations de la température intérieure moyenne T_a

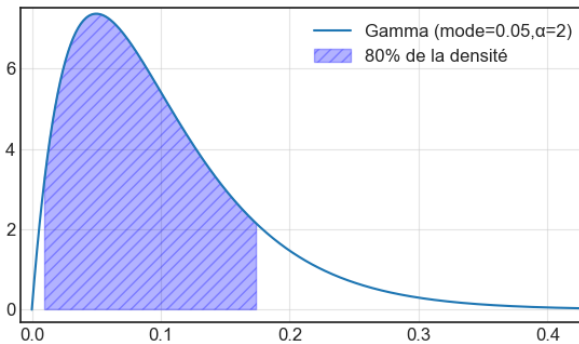


Figure 61 : Loi de probabilité gamma $G(0,05; 2)$

Distribution a priori	
Cp^*	Normale (1; 0,3) (sans dimension)
Rg^*	Normale (1; 0,3) (sans dimension)
Ap^*	Normale (1; 0,3) (sans dimension)
σ_g	Gamma (0,05; 2) (°C)

Tableau 23 : Résumé des distributions a priori

Les distributions a posteriori des paramètres calibrés sont présentées en Figure 62 et résumées statistiquement dans le Tableau 24. Ces distributions a posteriori apparaissent très resserrées. Sachant le modèle RC sélectionné et sachant les vecteurs de sollicitation, alors les paramètres du modèle RC sont identifiés avec une très faible incertitude. Les estimations physiques des paramètres ne sont pas comprises dans les intervalles HPD_95 associés à ces distributions a posteriori (Tableau 24) ; mais, selon les résultats du chapitre 3, à l'exception de la résistance globale Rg , les paramètres identifiés du modèle ISO_CpCa ne peuvent pas être directement associés aux estimations physiques correspondantes. Aussi, l'hypothèse des erreurs indépendantes, peut être à l'origine d'une sous-estimation des intervalles de crédence HPD.

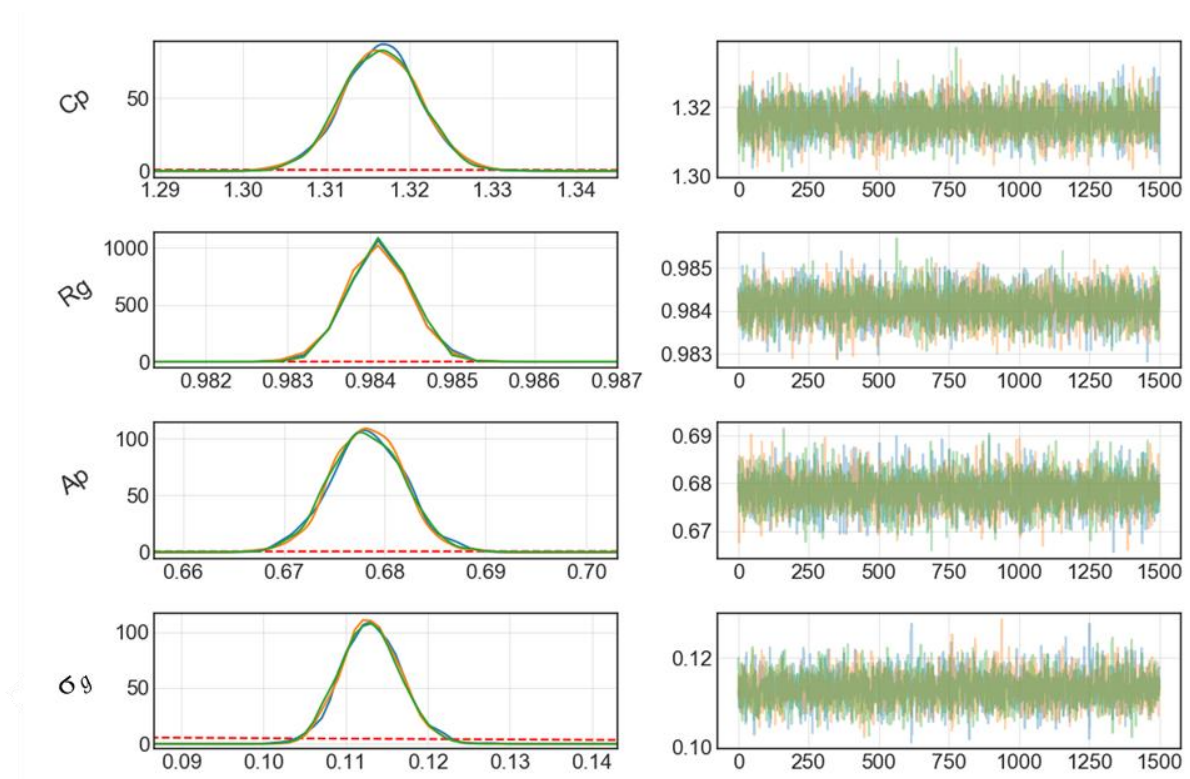


Figure 62 : Distribution des probabilités a posteriori échantillonnées en considérant la fonction de vraisemblance $L_{i.i.d}$

	Moyenne	Ecart type	HPD95	\hat{R}
Cp^*	1,317	0,0046	1,308 - 1,325	1,001
Rg^*	0,984	0,0004	0,983 - 0,985	1,000
Ap^*	0,678	0,0037	0,671 - 0,686	1,000
σ_g	0,113	0,0036	0,106 - 0,12	1,001

Tableau 24 : Résumé statistique des distributions a posteriori échantillonnées en considérant la fonction de vraisemblance $L_{i.i.d}$

Sur la Figure 63, la distribution des prédictions de température a posteriori est représentée via les intervalles de crédence HPD95 et HPD50. Ces intervalles sont particulièrement resserrés (environ 0,4 °C d'amplitude pour les intervalles HPD95). Pour chaque période, les RMSE a posteriori sont inférieures à 0,25 °C (Tableau 25). Le modèle calibré présente donc de bonnes capacités prédictives.

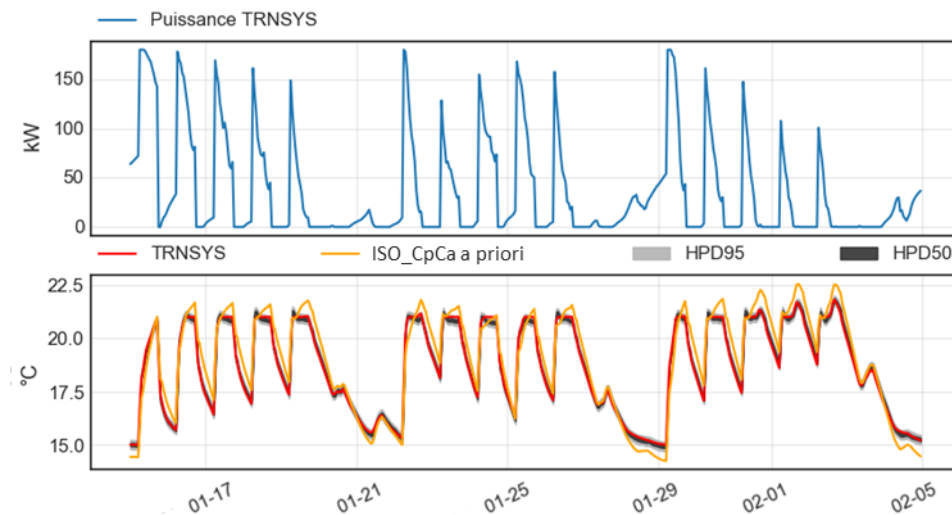


Figure 63 : Prédictions de températures a priori et a posteriori sur la période de calibration dans le cas de la fonction de vraisemblance $L_{i.i.d}$

	RMSE a priori	RMSE a posteriori
Janvier (période de calibration)	0,62	0,11
Février	0,65	0,18
Mars	0,86	0,33

Tableau 25 : RMSE a priori et a posteriori pour les périodes de calibration et de validation dans le cas de la fonction de vraisemblance $L_{i.i.d}$

Sur la Figure 64 sont analysés les résidus obtenus sur la période de calibration. La distribution des résidus est proche d'une loi normale centrée en zéro et d'écart

type σ_g . L'autocorrélation des résidus présente des pics de corrélation pour tout intervalle multiple de 24 h; cela correspond au profil journalier des sollicitations.

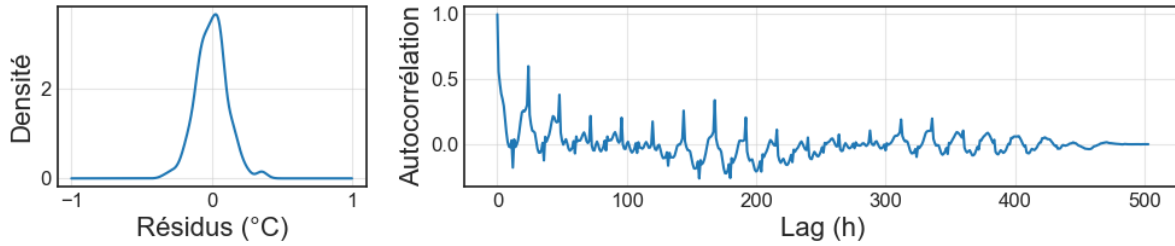


Figure 64 : Analyse des résidus, pour le cas d'étude de la fonction de vraisemblance $L_{i.i.d}$

5.2.2 Fonction de vraisemblance L_{GP}

La fonction de vraisemblance L_{GP} est rappelée en équation (123). Cette fonction de vraisemblance modélise les corrélations temporelles via un processus gaussien de fonction de covariance K (124). Des probabilités a priori sont attribuées à chacun des paramètres de la fonction de vraisemblance (Tableau 26).

$$L_{GP} = \frac{1}{(2\pi)^{n/2} |\Sigma_{LGP}|^{1/2}} e^{-\frac{1}{2} (D - \eta(\theta))^T \Sigma_{LGP}^{-1} (D - \eta(\theta))} \quad (123)$$

$$\Sigma_{LGP,ij} = \sigma_g^2 e^{-\frac{(t_i - t_j)^2}{2l_1^2}} e^{-\frac{2\sin^2\left(\frac{\pi(t_i - t_j)}{24h}\right)}{l_2^2}} e^{-\frac{(P_i - P_j)^2}{2l_3^2}} \quad (124)$$

- D_k : données de température intérieure moyenne
- $\eta_k(\theta)$: simulations de la température intérieure moyenne Ta

	Distribution a priori
Cp^*	Normale (0;0,3) (normalisée)
Rg^*	Normale (1;0,3) (normalisée)
Ap^*	Normale (1;0,3) (normalisée)
σ_g	Gamma (0,05;2) (°C)
l_1	Gamma (25;3) (h)
l_2	Gamma (1;3) (sans unité)
l_3	Gamma (50;3) (kW)

Tableau 26 : Résumé des distributions a priori pour le cas d'étude de la fonction de vraisemblance L_{GP}

Remarque : Pour assurer la stabilité numérique, un faible bruit : $0,001 * I_n$ est ajouté à la matrice Σ_{LGP} .

Les résultats de calibration sont présentés dans les Tableau 27 et Tableau 28. Les résultats issus de l'utilisation de la fonction de vraisemblance $L_{i.i.d}$ sont ajoutés à ces tableaux pour faciliter la comparaison. Les valeurs moyennes a posteriori des paramètres Cp , Ap , Rg et σ_g , obtenues avec les fonctions de vraisemblance $L_{i.i.d}$ et L_{GP} sont similaires. Cependant, les écart-types correspondant à la fonction L_{GP} sont plus élevés; la prise en compte des corrélations entre les données rend ces dernières moins informatives. La distribution a posteriori des courbes de température est similaire à celle obtenue avec la vraisemblance $L_{i.i.d}$ (Tableau 31).

	$L_{i.i.d}$		L_{GP}				
	Moyenne	Ecart type	Moyenne	Ecart type	hpd_5	hpd_95	\hat{R}
Cp^*	1,317	0,0046	1,329	0,019	1,294	1,366	1,000
Rg^*	0,984	0,0004	0,981	0,0034	0,973	0,987	1,000
Ap^*	0,678	0,0037	0,689	0,010	0,669	0,710	1,000
σ_g	0,12	0,0036	0,21	0,026	0,16	0,26	1,000
l_1			28	1,0	25	29	1,000
l_2			1,07	0,083	0,91	1,2	1,000
l_3			56	7,3	42	70	1,002

Tableau 27 : Résumé statistique des distributions a posteriori pour le cas d'étude de la fonction de vraisemblance L_{GP}

	$L_{i.i.d}$		L_{GP}
	RMSE a priori	RMSE a posteriori	RMSE a posteriori
Janvier (période de calibration)	0,62	0,11	0,12
Février	0,65	0,18	0,16
Mars	0,86	0,33	0,32

Tableau 28 : RMSE a priori et a posteriori pour les périodes de calibration et de validation, pour le cas d'étude de la fonction de vraisemblance L_{GP}

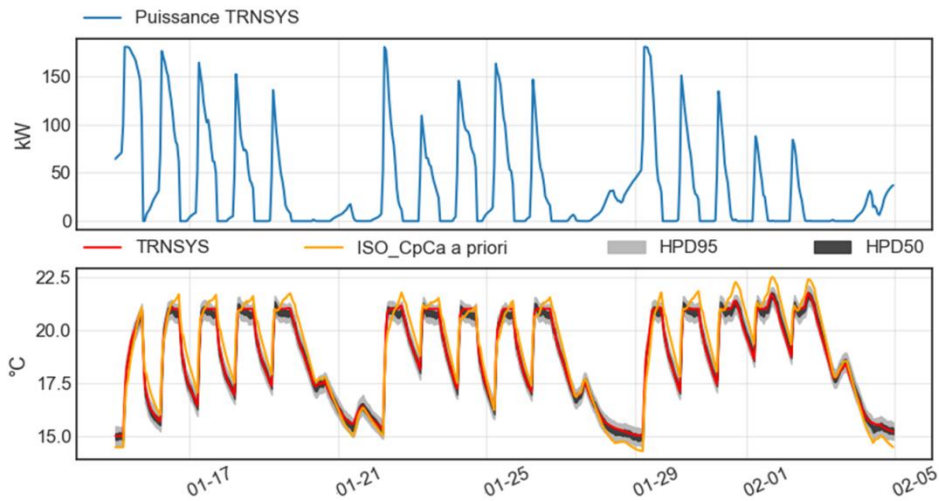


Figure 65 : Prédictions de températures a priori et a posteriori sur la période de calibration, pour le cas d'étude de la fonction de vraisemblance L_{GP}

La matrice de covariance a posteriori est représentée sur la Figure 66. Les lignes diagonales correspondent aux corrélations entre les données espacées de 24 h. Les lignes claires, verticales et horizontales, correspondent aux changements de température de consigne. Les trois semaines de la période calibration peuvent ainsi être distinguées sur cette matrice de covariance.

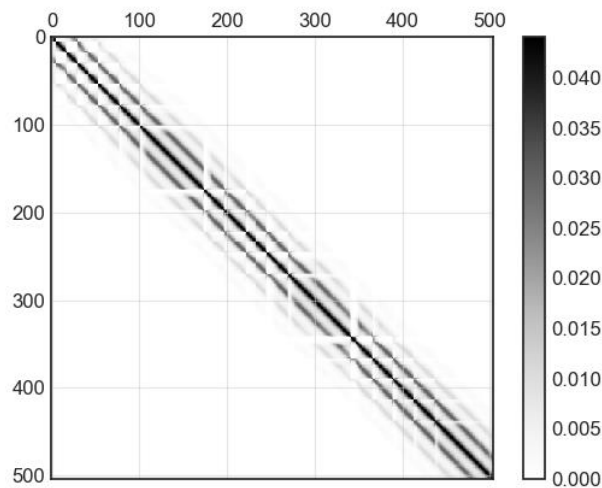


Figure 66 : Matrice de covariance a posteriori Σ_{LGP}

5.2.3 Fonction de vraisemblance L_{Kal}

La fonction de vraisemblance calculée à partir du filtre de Kalman est rappelée en équation (125). En plus des paramètres physiques du modèle RC, sont calibrés les éléments diagonaux de la racine carrée de la matrice de covariance du bruit du processus : σ_{q1} et σ_{q2} (homogène à des écarts-types), et l'écart-type du bruit de

mesure σ_R (126). Les distributions a priori des paramètres calibrés sont présentées dans le Tableau 29.

$$L = P(D_0 | \eta_n(\theta)) \prod_{k=1}^N \frac{1}{\sqrt{2\pi S_k}} e^{-\frac{1}{2} \frac{(\tilde{e}_k(\theta))^2}{S_k}} \quad (125)$$

$$Q^{1/2} = \begin{bmatrix} \sigma_{q1} & 0 \\ 0 & \sigma_{q2} \end{bmatrix} \quad R^{1/2} = \sigma_R \quad (126)$$

- D_k : données de température intérieure moyenne
- $\eta_k(\theta)$: simulations de la température intérieure moyenne Ta

	Distribution a priori
Cp	Normale (0;0,3) (normalisée)
Rg	Normale (1;0,3) (normalisée)
Ap	Normale (1;0,3) (normalisée)
σ_R	Gamma (0,05;2) (°C)
σ_{q1}	Gamma (0,05;2) (°C)
σ_{q2}	Gamma (0,05;2) (°C)

Tableau 29 : Résumé des distributions a priori (cas de la fonction de vraisemblance L_{Kal})

Les résultats de calibration sont présentés dans les Tableau 30 et Tableau 31. Les distributions a posteriori apparaissent moins resserrées que pour le cas de la vraisemblance $L_{i.i.d}$. Les valeurs moyennes des paramètres physiques calibrés diffèrent légèrement de celles obtenues précédemment, notamment pour le paramètre Ap (20 % plus élevé). Les prédictions de température a posteriori apparaissent aussi moins bonnes que pour le modèle calibré avec la fonction de vraisemblance $L_{i.i.d}$ (Figure 67, Figure 68).

	$L_{i.i.d}$		L_{Kal}				
	Moyenne	Ecart type	Moyenne	Ecart type	hpd_5	hpd_95	\hat{R}
Cp	1,317	0,0046	1,32	0,027	1,264	1,370	1,000
Rg	0,984	0,0004	1,00	0,019	0,971	1,043	1,000
Ap	0,678	0,0036	0,90	0,0088	0,886	0,921	1,001
σ_{q1}			0,005	0,0029	0,0003	0,011	1,000
σ_{q2}			0,057	0,0019	0,053	0,061	1,000
σ_R			0,006	0,0035	0,001	0,013	1,001

Tableau 30 : Résumé statistique des distributions a posteriori pour le cas de la fonction de vraisemblance L_{Kal}

	$L_{i.i.d}$		L_{Kal}
	RMSE a priori	RMSE a posteriori	RMSE a posteriori
Janvier (période de calibration)	0,62	0,11	0,40
Février	0,65	0,18	0,44
Mars	0,86	0,33	0,59

Tableau 31: RMSE a priori et a posteriori pour le cas de la fonction de vraisemblance L_{Kal}

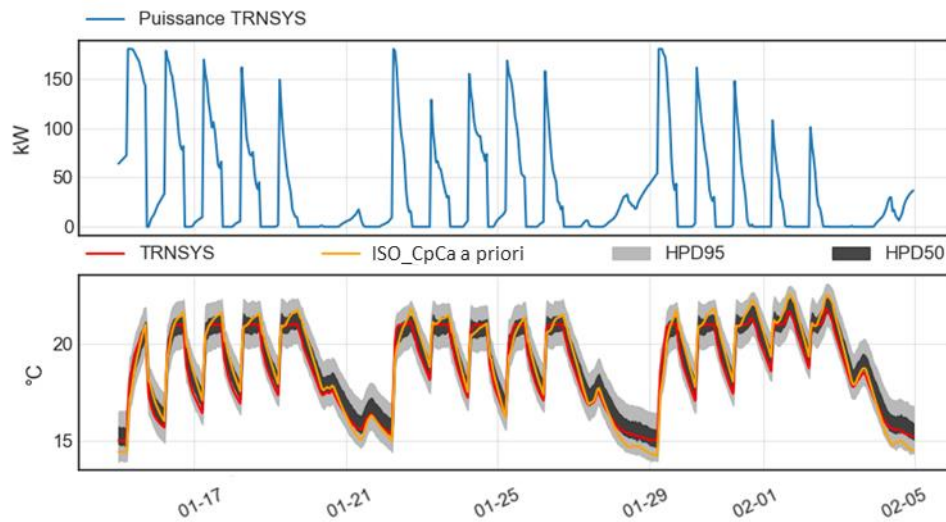


Figure 67 : Prédictions de températures a priori et a posteriori sur la période de calibration pour le cas de la fonction de vraisemblance L_{Kal}

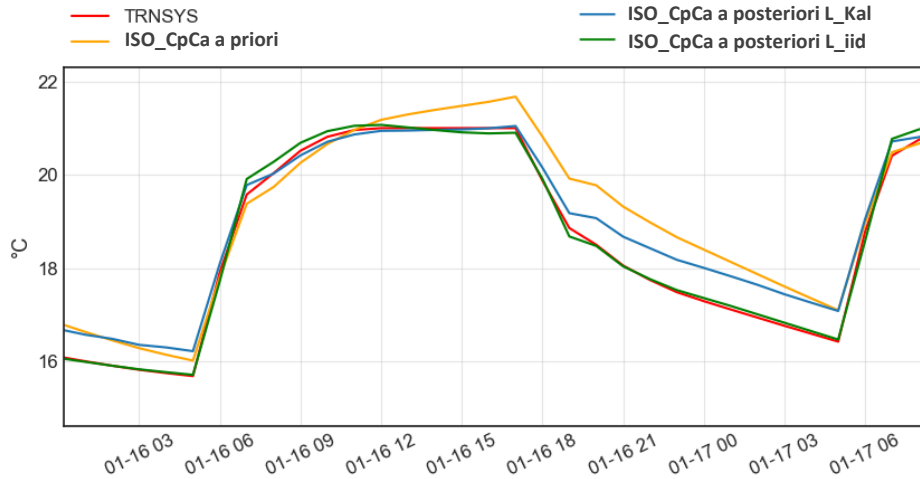


Figure 68 : Prédictions de températures a priori et a posteriori sur la période de calibration (cas de la fonction de vraisemblance L_{Kal})

Afin de mieux comprendre ces résultats, une série de calibrations a été réalisée avec la fonction de vraisemblance L_{Kal} , en fixant le paramètre σ_R à différentes valeurs. Ce paramètre permet de faire varier le degré d'actualisation des données. Lorsque σ_R est faible, l'incertitude sur les données est supposée faible, les simulations sont alors fortement actualisées. A l'inverse, lorsque σ_R est élevé, les simulations sont peu actualisées. Les résultats de cette analyse sont présentés dans le Tableau 32. Des prédictions a posteriori avec actualisation, ou sans actualisation peuvent être calculées ; des RMSE actualisés et non actualisés sont alors présentés.

On vérifie que plus les simulations sont actualisées (σ_R faible), plus les RMSE actualisés sont faibles; et que les RMSE a posteriori sont toujours plus faibles que les RMSE a priori. Il apparaît aussi que les valeurs moyennes a posteriori des paramètres du modèle RC, et les RMSE a posteriori non actualisés, varient avec le degré d'actualisation. Quand le degré d'actualisation baisse, les résultats tendent vers les valeurs obtenues dans le cas de la fonction de vraisemblance $L_{i.i.d}$. Quand le degré d'actualisation augmente, le RMSE a posteriori non actualisé se dégrade.

		L_{Kal}					$L_{i.i.d}$
		σ_R	0,005	0,05	0,1	0,2	
Moyenne a posteriori	Cp	1,34	1,31	1,29	1,31	1,31	1,31
	Rg	1,01	1,00	0,99	0,98	0,98	0,98
	Ap	0,91	0,81	0,72	0,68	0,68	0,68
RMSE non actualisé	a priori	0,62	0,62	0,62	0,62	0,62	0,62
	a posteriori	0,57	0,39	0,22	0,16	0,11	0,11
RMSE actualisé	a priori	0,098	0,099	0,099	0,10	0,14	
	a posteriori	0,044	0,053	0,073	0,10	0,11	

Tableau 32 : Analyse de l'influence du degré d'actualisation sur les résultats de calibration

5.2.4 Conclusion sur le choix de la fonction de vraisemblance

La fonction de vraisemblance $L_{i.i.d}$ a l'avantage d'avoir un seul hyper-paramètre à calibrer, ce qui permet des temps de calibration réduits (environ 5 échantillons/s). L'utilisation de la fonction de vraisemblance L_{GP} devrait permettre, du fait de la prise en compte des corrélations temporelles, d'obtenir des distributions a posteriori plus justes. En revanche, cela implique des temps de calibration beaucoup plus longs (environ 6 s par échantillon) car trois paramètres supplémentaires sont calibrés et l'inverse de la matrice de covariance ($\Sigma_L \in \mathbb{R}^{n \times n}$) doit être calculée à chaque itération. Enfin, l'utilisation de la vraisemblance L_{Kal} , pourra être intéressante pour l'identification de certains paramètres physiques, comme la résistance Rg , en permettant d'éviter d'éventuels biais engendrés par des phénomènes non modélisés, comme expliqué au chapitre 2. Elle sera aussi particulièrement pertinente pour identifier un modèle de commande prédictive. Cependant, les modèles calibrés par cette fonction de vraisemblance peuvent présenter de moins bonnes capacités prédictives.

5.3 Discussion sur le nombre de capteurs de température

Jusqu'ici, la température moyenne des bâtiments étudiés était calculée comme la moyenne des températures de chaque zone, pondérée par les surfaces au sol. Cependant, en pratique, il peut être difficile d'instrumenter toutes les zones d'un bâtiment. Il paraît alors pertinent de se demander si un nombre de capteurs réduit

peut être utilisé. Pour éclairer cette problématique, différents plans d'instrumentation sont étudiés:

- Tref : plan de référence incluant 15 mesures température : une par zone.
- T3mid : 3 mesures température situées dans les zones principales de l'étage intermédiaire : « Bureau 1 », « Bureau 2 » et « Salle de réunion ».
- T4mid : 4 mesures température situées dans les zones de l'étage intermédiaire : « Bureau 1 », « Bureau 2 », « Salle de réunion », et « Circulation ».
- T3roof : 3 mesures température situées dans les zones principales de l'étage supérieur : « Bureau 1 », « Bureau 2 » et « Salle de réunion ».
- T3floor : 3 mesures température situées dans les zones principales de l'étage inférieur : « Bureau 1 », « Bureau 2 » et « Salle de réunion ».
- T3réu: 3 mesures température situées dans les zones «salles de réunion » de chaque étage.

Pour cette étude, en plus du scénario « Base », le scénario « Zones », dans lesquels les zones de circulation ne sont pas chauffées, est considéré. Dans le Tableau 33 les données températures moyennes calculées selon ces différents plans d'instrumentation sont comparées au cas de référence Tref. Pour le cas d'étude « Base », le plan d'instrumentation T3mid semble pertinent, l'estimation de la température moyenne du bâtiment est similaire au cas de référence (0,1 °C d'écart en moyenne). En revanche, pour le cas d'étude « Zones », ce plan d'instrumentation entraîne des écarts importants (0,5 °C). L'ajout d'une mesure dans la zone de circulation non chauffée (T4mid) permet de réduire les écarts avec la température de référence. On remarque que pour le cas d'étude « Base », ce plan d'instrumentation (T4mid) fait augmenter légèrement les écarts par rapport au plan T3mid. Cela vient du fait que le plan T3mid, en ne prenant pas en compte les températures plus froides du dernier étage, surestime la température moyenne ; cette surestimation est amplifiée par la prise en compte de la température de la zone de circulation de l'étage intermédiaire. Les plans d'instrumentation T3roof, T3floor, et T3conf engendrent des biais plus importants sur l'estimation de la température moyenne du bâtiment.

	T3mid	T4mid	T3roof	T3floor	T3conf
Base	0,09	0,27	0,89	0,25	0,84
Zones	0,45	0,23	0,81	0,58	0,98

Tableau 33 : RMSE (°C) entre la température de référence Tref, et les températures calculées selon les différents plans d'instrumentation étudiés

L'effet du plan d'instrumentation sur les résultats de calibration est observé dans les Tableau 34 et Tableau 35, respectivement pour les scénarios «Base » et « Zones » . Ces résultats confirment les observations faites dans le Tableau 33. Les plans d'instrumentation T3mid et T4mid sont les plus pertinents, ils permettent de fortement réduire le nombre de capteurs tout en limitant les biais sur la calibration du modèle.

		Tref	T3mid	T4mid	T3roof	T3floor	T3conf
Moyenne a posteriori	Cp^*	1,32	1,29	1,32	1,37	1,32	1,50
	Rg^*	0,98	0,98	0,98	1,00	0,98	1,03
	Ap^*	0,68	0,68	0,68	0,78	0,68	0,96
	σ_g	0,11	0,11	0,11	0,15	0,11	0,41
RMSE a posteriori	Janvier (Calib)	0,11	0,12	0,13	0,17	0,13	0,46
	Février	0,18	0,30	0,39	0,26	0,39	1,08
	Mars	0,33	0,52	0,67	0,40	0,67	1,88

Tableau 34 : Résultats de calibration selon différents plans d'instrumentation, pour le scénario « Base »

		Tref	T3mid	T4mid	T3roof	T3floor	T3conf
Moyenne a posteriori	Cp^*	1,32	1,23	1,28	1,29	1,27	1,43
	Rg^*	0,98	1,02	1,00	1,04	0,97	1,06
	Ap^*	0,68	0,46	0,60	0,53	0,30	0,61
	σ_g	0,11	0,17	0,11	0,20	0,35	0,46
RMSE a posteriori	Janvier (Calib)	0,11	0,18	0,12	0,23	0,39	0,52
	Février	0,18	0,50	0,48	0,43	0,72	1,27
	Mars	0,33	0,81	0,77	0,67	0,95	2,15

Tableau 35 : Résultats de calibration selon différents plans d'instrumentation, pour le scénario « Zones »

5.4 Identification des capacités et des limites de la méthode de calibration de modèles RC selon les données disponibles

En plus des paramètres structurels, la méthode de calibration bayésienne nécessite la prise en compte des incertitudes liées aux données d'entrée dynamiques :

- Données météo
- Puissance thermique de chauffage
- Electricité spécifique
- Température de consigne
- Température intérieure
- Scénario d'occupation
- Scénario de ventilation
- Scénario d'utilisation des stores extérieurs

Les disponibilités de ces données d'entrée peuvent fortement varier selon les quartiers et les bâtiments étudiés. Les données météo et les données de puissance de chauffage sont supposées indispensables à la calibration du modèle. Pour les autres données, des estimations a priori peuvent être faites. Pour le cas des bâtiments de bureaux, des modèles a priori sont présentés dans le chapitre 4.

Une étude est proposée afin d'identifier ce que l'on peut attendre de la méthode de calibration des modèles RC, selon les données disponibles, au regard des objectifs de la thèse : étude de scénarios de performance énergétique, évaluation d'actions de performance énergétique mises en place (IPMVP), identification de dérives de consommation et des causes de ces dérives. Dans le Tableau 36, une série de cas d'étude dans lesquels les données d'entrée dynamiques disponibles varient est proposée, on parlera de niveaux de connaissance des bâtiments.

Au niveau 5, seule la puissance thermique est connue. Au niveau 4 sont ajoutées les données de consommations d'électricité spécifique. Ces deux premiers niveaux peuvent être atteints sans instrumentation spécifique. Les données de puissance thermique peuvent être issues de sous-stations de réseau de chaleur, quant aux données de consommation d'électricité spécifique, elles peuvent provenir de compteurs communiquant tels que le Linky. Dans les niveaux suivants, des données liées à la température intérieure sont ajoutées. Ces données sont particulièrement importantes pour la calibration des modèles, mais leur obtention peut s'avérer difficile. Le niveau

3, suppose d'abord la disponibilité de la température de consigne du bâtiment, puis, au niveau 2, la température intérieure est connue. Enfin, le niveau 1 suppose une très bonne connaissance du bâtiment, les profils de ventilation mécanique, d'occupation, et d'utilisation des protections solaires sont connus. Ces scénarios sont étudiés sur le cas d'étude « Base » présenté dans le chapitre 3, correspondant à un bâtiment de bureaux modélisé sur TRNSYS, sur lequel quelques modifications seront apportées, on parlera alors du cas d'étude « TRNSYS ».

	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Données météo	✓	✓	✓	✓	✓
Puissance thermique de chauffage	✓	✓	✓	✓	✓
Electricité spécifique	✓	✓	✓	✓	
Température de consigne			✓		
Température intérieure	✓	✓			
Scénario d'occupation	✓				
Scénario de ventilation	✓				
Scénario de stores extérieurs	✓				

Tableau 36 : Résumé des différents niveaux de connaissance du bâtiment envisagés

Pour les 5 cas d'études présentés ci-dessus, le modèle ISO_CpCa est calibré selon la méthode bayésienne. Lorsque les données de température intérieure sont disponibles, la fonction de vraisemblance $L_{i.i.d}$, présentée ci-dessous est utilisée :

$$L_{i.i.d} = \prod_{k=1}^n \frac{1}{\sigma_g \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{D_k - \eta_k(\theta, x)}{\sigma_g} \right)^2} \quad (127)$$

- d_k : données de température intérieure moyenne
- $\eta(\theta, x_k)$: simulations de la température intérieure moyenne T_a

Quand les données de températures ne sont pas connues, le modèle est calibré sur les données de puissance thermique de chauffage via la fonction de vraisemblance suivante :

$$L_{i.i.d} = \prod_{k=1}^n \frac{1}{\sigma_{gp} \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{D_k - \eta_k(\theta, x)}{\sigma_{gp}} \right)^2} \quad (128)$$

- d_k : données de puissance de chauffage
- $\eta(\theta, x_k)$: simulations de la puissance de chauffage φ_{heat}

5.4.1 Estimation des incertitudes a priori

A nouveau, les paramètres $Rate_m$ et $Rate_l$ sont fixés et les incertitudes liées aux autres paramètres du modèle RC sont pris en compte selon les équations suivantes :

$$\begin{aligned}
 Rg &= Rg_0 \cdot Rg^* & Rg^* &\sim N(1 ; 0,3) \text{ (sans unité)} \\
 Cp &= Cp_0 \cdot Cp^* & Cp^* &\sim N(1 ; 0,3) \text{ (sans unité)} \\
 Ap &= Ap_0 \cdot Ap^* & Ap^* &\sim N(1 ; 0,3) \text{ (sans unité)} \\
 && Rg_0, Cp_0, Ap_0 &: \text{ estimations physiques a priori}
 \end{aligned} \tag{129}$$

Lorsqu'elles ne sont pas connues, les données d'entrée dynamiques sont estimées par les modèles présentés dans le chapitre 4. Ces modèles doivent également être associés à des incertitudes a priori et calibrés. L'objectif est de réussir à approcher les profils réels en calibrant un minimum de paramètres. Les variations de profils liés au comportement des occupants ne pourront notamment pas être parfaitement modélisées. De telles variations sont considérées dans le cas d'étude TRNSYS pour se rapprocher d'un cas réel.

5.4.1.1 Scénario d'utilisation des stores extérieurs (apports solaires)

Quand le scénario d'utilisation des stores extérieurs n'est pas connu, ce dernier est estimé via le modèle présenté dans le chapitre 4. L'incertitude a priori de ce modèle est prise en compte via une variable appelée Sol_in , homogène à un pourcentage, qui décale, sur l'axe des ordonnées, la courbe de corrélation entre l'éclairement et l'indice de protection solaire :

$$IPS = \begin{cases} \frac{0,25 - 0,07}{28000} \cdot E + 0,07 + Sol_in/100, & \text{pour } E < 28000 \text{ lux} \\ \frac{0,45 - 0,25}{100000 - 28000} \cdot E + 0,17 + Sol_in/100, & \text{pour } E > 28000 \text{ lux} \end{cases} \tag{130}$$

$$Sol_in \sim N(0 ; 10) \tag{131}$$

La distribution des scénarios a priori des apports solaires internes, obtenue selon cette formulation, est présentée sur la Figure 69. Les apports solaires internes modélisés dans le cas d'étude TRNSYS sont également affichés cette Figure 69. Dans ce modèle TRNSYS, des variations aléatoires de l'indice de protection solaire, dues aux comportements des occupants, sont considérées. A chaque heure, une variation de

l'indice de protection solaire (IPS) est tirée aléatoirement dans une loi normale $N(0 ; 0,1)$.

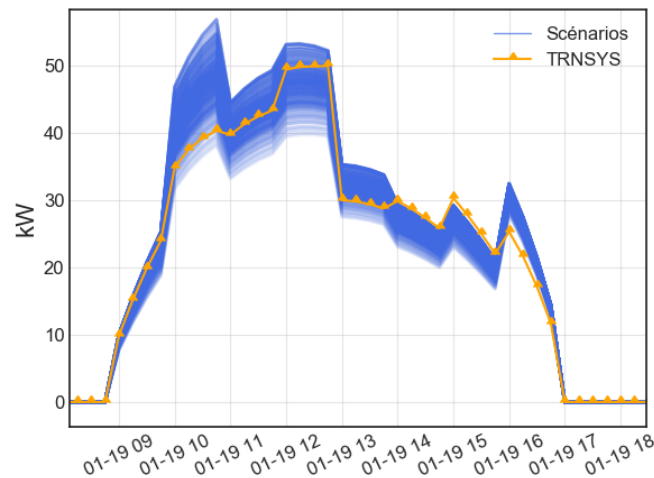


Figure 69 : Distribution a priori des scénarios des apports solaires internes (Abscisse : mois-jour heure)

5.4.1.2 Scénario de ventilation mécanique

L'incertitude associée au modèle de ventilation mécanique est prise en compte via trois variables ($Meca$, $Meca_h1$ et $Meca_h2$) associées à des distributions de probabilité a priori :

$$U_{meca} = \begin{cases} \frac{\dot{V}_{meca} \cdot V_{ch} \cdot \rho \cdot c_{air}}{3600} Meca, & t \in [(h1 + Meca_h1) : (h2 + Meca_h2)] \\ 0, & sinon \end{cases} \quad (132)$$

$$Meca \sim N(1; 0,2) (\text{sans dimension})$$

$$Meca_h1 \sim N(0; 1)(h)$$

$$Meca_h2 \sim N(0; 1)(h)$$

Selon l'équation ci-dessus, la distribution des scénarios a priori du coefficient de transfert U_{meca} , lié au débit global de ventilation du bâtiment, est présentée sur la Figure 70. Dans le cas du modèle TRNSYS, des variations du débit global de ventilation, dues à l'utilisation de la salle de réunion, sont aussi modélisées.

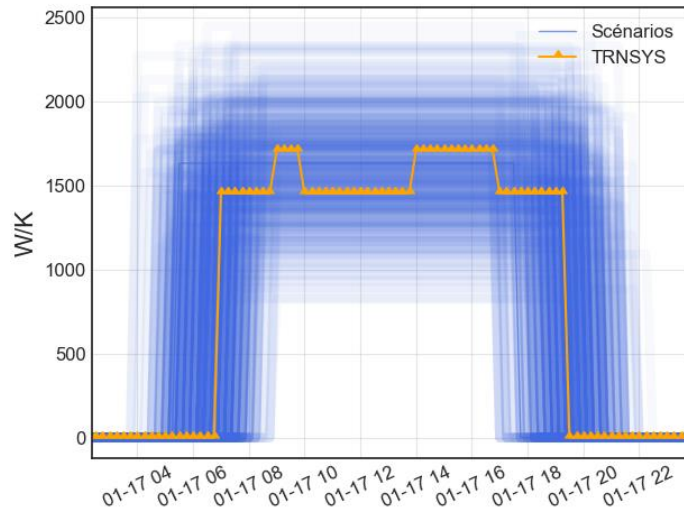


Figure 70 : Distribution a priori des scénarios de ventilation mécanique (*Umeca*) (Abscisse : mois-jour heure)

5.4.1.3 Scénario d'occupation (apports métaboliques)

L'incertitude liée au modèle estimant les apports thermiques engendrés par le métabolisme des occupants est prise en compte via le paramètre *Meta* :

$$\varphi_{meta} = Occ \cdot 80 \cdot \frac{S_{ch}}{r_{pers}} \cdot Meta \quad (133)$$

$$Meta \sim N(1; 0,2) \text{ (sans unité)}$$

La distribution a priori des scénarios d'apports thermiques métaboliques, est présentée sur la Figure 71. Le scénario du modèle TRNSYS inclut, en plus, des pics dus à l'occupation des salles de réunions.

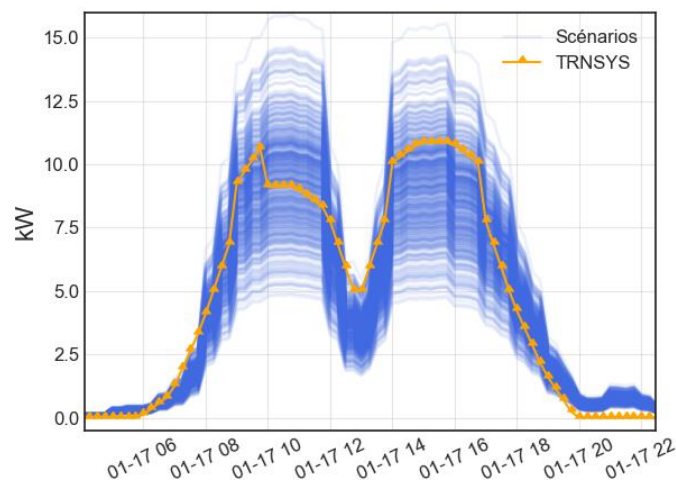


Figure 71 : Distribution a priori des scénarios d'apports métaboliques (Abscisse : mois-jour heure)

5.4.1.4 Scénario de températures de consigne

Lorsque le scénario de température de consigne n'est pas connu, ce dernier est modélisé par une température de consigne Th_0 et une température de réduit Thr_0 . Les heures de changement de consigne sont supposées connues, elles peuvent être observées sur les données de consommation de chauffage. Les incertitudes sur ces scénarios sont prises en compte via les variables Th et Thr , selon l'équation suivante :

$$T_{cons} = \begin{cases} Th_0 \cdot Th, & \text{en période d'occupation} \\ Thr_0 \cdot Thr, & \text{hors occupation} \end{cases} \quad (134)$$

$$Th \sim N(1; 0,1) \text{ (sans unité)}$$

$$Thr \sim N(1; 0,1) \text{ (sans unité)}$$

La distribution a priori des scénarios de consignes de température, est présentée sur la Figure 71.

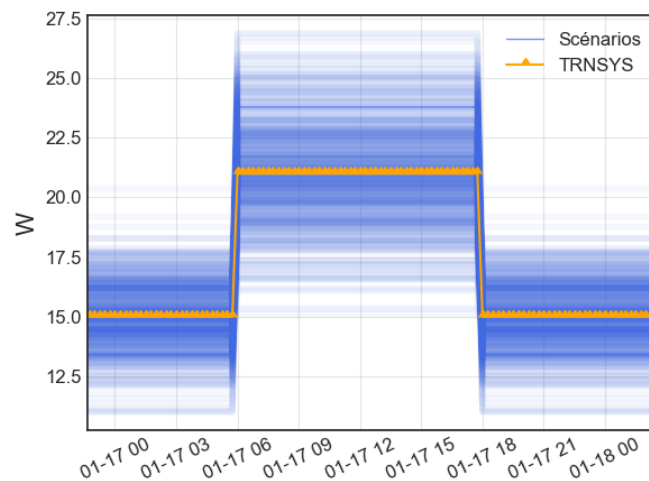


Figure 72 : Distribution a priori des scénarios de consigne de température (Abscisse : mois-jour heure)

5.4.1.5 Scénario de consommation d'électricité spécifique

L'incertitude sur les apports thermiques engendrés par les consommations d'électricité spécifique est considérée via deux variables :

$$\varphi_{elec} = Occ \cdot S_{ch} \cdot r_{elec} \cdot 0,9 \cdot El_{occ} + r_{elec} \cdot S_{ch} \cdot 0,1 \cdot El_{base} \quad (135)$$

$$El_{base} \sim N(1; 0,2)$$

$$El_{occ} \sim N(1; 0,2)$$

La variable EL_{occ} sera fortement corrélée à la variable $Meta$. Pour limiter les corrélations entre ces variables, les incertitudes sur la part de φ_{elec} , dépendante de l'occupation, et sur φ_{meta} , seront considérées via une unique variable $EL_{Me_{occ}}$:

$$\begin{aligned} \varphi_{elec} + \varphi_{meta} &= Occ \cdot \left(S_{ch} r_{elec} 0,9 + 80 * \frac{S_{ch}}{r_{pers}} \right) \cdot EL_{Me_{occ}} \\ &+ r_{elec} \cdot S_{ch} \cdot 0,1 \cdot EL_{base} \\ EL_{base} &\sim N(1; 0,2) \\ EL_{Me_{occ}} &\sim N(1; 0,2) \end{aligned} \quad (136)$$

La distribution a priori des apports thermiques liés à l'électricité spécifique est représentée sur la Figure 71. Pour cette sollicitation, le modèle calibré présente des différences importantes avec le modèle TRNSYS. Cela constitue un cas d'étude intéressant et est représentatif du fait que les consommations d'électricité peuvent présenter de fortes incertitudes dû au comportement des occupants.

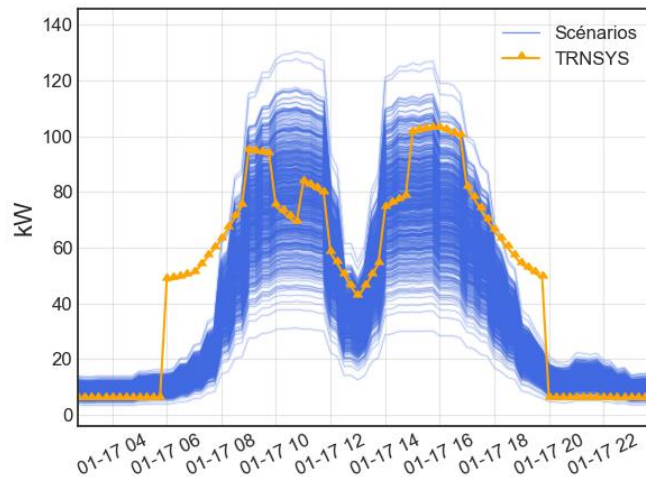


Figure 73 : Distribution a priori des scénarios des apports thermiques liés aux consommations d'électricité spécifique (Abscisse : mois-jour heure)

5.4.1.6 Résumé des hypothèses de modélisation et des distributions a priori

Les distributions a priori utilisées dans cette étude sont résumées dans le Tableau 37 :

	Distribution a priori des paramètres calibrés	Valeurs a priori des paramètres estimés à partir du modèle TRNSYS
Cp*	Normale (1; 0,3) (sans dimension)	$Cp_0/S_{ch} = 183 \text{ (kJ/m}^2 \text{ K)}$
Rg*	Normale (1; 0,3) (sans dimension)	$U_{roof} = 0,65 \text{ W/m}^2 \cdot \text{K}$ $U_{wall} = 0,93 \text{ W/m}^2 \cdot \text{K}$ $U_{win} = 7,5 \text{ W/m}^2 \cdot \text{K}$ $\dot{V}Inf = 0,2 \text{ V}_{ch}/h$
Ap*	Normale (1; 0,3) (sans dimension)	$\frac{Ap_0}{S_{ch}} = 3$
Sol_in	Normale (0; 5) (%)	
Meca	Normale (1; 0,2) (sans dimension)	$\dot{V}meca = 0,6 \text{ V}_{ch}/h$
Meca_h1	Normale (0; 1) (h)	$Meca_h1_0 = 6h30$
Meca_h2	Normale (0; 1) (h)	$Meca_h2_0 = 19h30$
Meta	Normale (1; 0,2) (sans dimension)	$r_{pers} = 25 \text{ pers/m}^2$
El_base El_occ El_Me_occ	Normale (1; 0,2) (sans dimension)	$r_{elec} = 21 \text{ W/m}^2$
Th	Normale (1; 0,1) (sans dimension)	$Th_0 = 21 \text{ }^\circ\text{C}$
Thr	Normale (1; 0,1) (sans dimension)	$Thr_0 = 15 \text{ }^\circ\text{C}$
σ_g	Gamma (0,05;2) ($^\circ\text{C}$)	
σ_{gp}	Gamma (3000;2) (W)	

Tableau 37 : Résumé des distributions a priori

5.4.2 Analyse des modèles a posteriori

Les résultats de l'étude sont résumés dans les Figure 74 et Figure 75. La Figure 74 présente, pour chaque scénario, les distributions a priori et a posteriori des paramètres calibrés, via des intervalles de crédence HDI (Highest Density Intervalles):

- HDI95 a priori : plus petit intervalle contenant 95 % de la densité de probabilité a priori
- HDI95 a posteriori : plus petit intervalle contenant 95 % de la densité de probabilité a posteriori, aussi appelé HPD (Highest Posterior Density) (voir chapitre 2).

La Figure 75 présente, pour chaque scénario, différents indicateurs caractérisant les prédictions a posteriori de température intérieure et de puissance de chauffage :

- RMSE_T : RMSE entre les données de température et les prédictions de température posteriori
- CVRMSE_P : CVRMSE entre les données de puissance et les prédictions de puissance posteriori

- HPD95_T_M : largeur moyenne du plus petit intervalle regroupant 95 % des prédictions de température a posteriori HPD95_T
- HPD95_P_M : largeur moyenne du plus petit intervalle regroupant 95 % des prédictions de puissance a posteriori HPD95_P

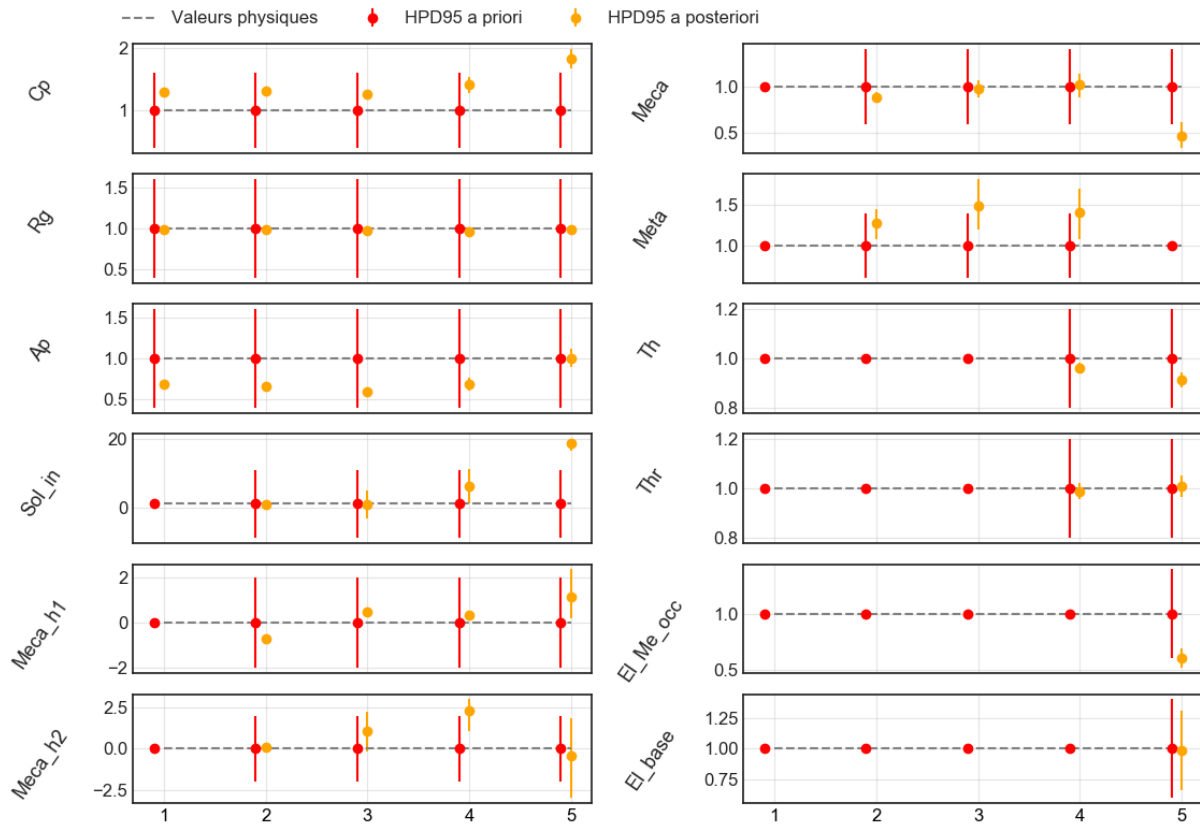


Figure 74 : Distribution a priori et a posteriori des paramètres calibrés pour les 5 niveaux de connaissance considérés

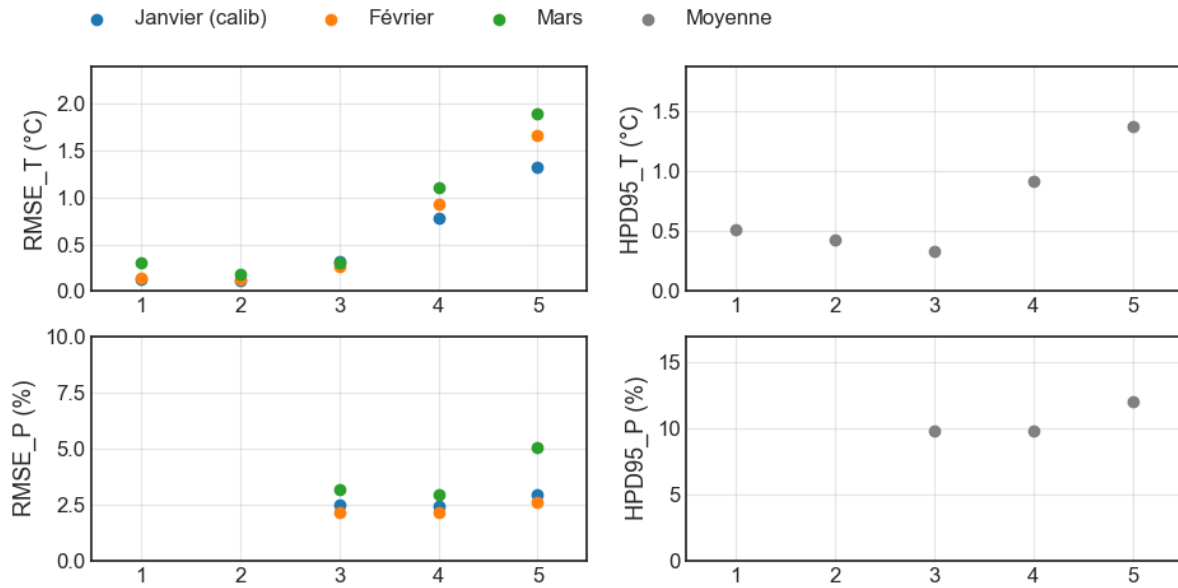


Figure 75 : Indicateurs caractérisant les prédictions a posteriori de température et de puissance, pour les 5 niveaux de connaissance considérés

Au niveau 1, on retrouve des résultats déjà observés précédemment, les distributions a posteriori des paramètres calibrés Ap , Rg et Cp sont très resserrés autour des valeurs moyennes : 0,7 0,98 et 1,3. Les prédictions des températures a posteriori sont très proches des données, le $RMSE_T$ est d'environ 0,1 °C, et la largeur moyenne de l'intervalle de crédence $HPD95_T_M$ est d'environ 0,5 °C.

Au niveau 2, l'ajout de l'incertitude liée au paramètre Sol_in , $Meca_h1$, $Meca_h2$ et $Meca$ change très peu les résultats de calibration. Les distributions a posteriori de ces paramètres sont concentrées autour des valeurs physiques correspondantes.

Au niveau 3, la température intérieure n'est plus disponible ; la calibration est réalisée sur les données de puissance de chauffage, estimées à partir de la température de consigne supposée connue. Cela a principalement pour effet de fortement augmenter l'incertitude a posteriori sur les paramètres : Sol_in , $Meca_h2$ et $Meta$. En plus des prédictions de températures moyennes, les prédictions de puissance de chauffage peuvent être évaluées, le $RMSE_P$ est d'environ 2.5 %, et la largeur de l'intervalle de crédence $HPD95_P_M$ est d'environ 10 %.

Au niveau 4, les deux paramètres Th et Thr , caractérisant le profil de température de consigne, sont calibrés. L'ajout de ces incertitudes ne modifie presque pas les distributions a posteriori déjà calibrées au niveau précédent ; les distributions a posteriori de ces paramètres sont proches des valeurs physiques. Cependant, l'intervalle de crédence des prédictions de température augmente fortement avec la

perte d'information sur la température de consigne, le HPD95_T_M passe atteint les 1 °C.

Au, niveau 5, la calibration du modèle des apports thermiques dus aux consommations d'électricité spécifique, modifie significativement les distributions a posteriori de plusieurs paramètres : *Sol_in*, *Meca_h1*, *Meca*. Aussi, les RMSE_T atteignent des valeurs proches de 1,5 °C pour les trois périodes étudiées. Cela provient du fait le profil du scénario calibré présente des différences importantes avec celui modélisé dans TRNSYS.

Ainsi, dans la majorité des cas, les paramètres calibrés conservent un sens physique, et le modèle présente de bonnes capacités prédictives malgré des incertitudes liées au comportement des occupants. Des limites sont cependant à noter : des erreurs trop importantes sur les profils des données d'entrée dynamique peuvent mener à des valeurs de paramètres calibrés non-crédibles et une dégradation des performances du modèle. Cela est notamment susceptible de se produire pour les données de consommations d'électricité spécifique qui peuvent fortement varier selon le comportement des occupants.

Les modèles a posteriori sont plus représentatifs de l'état initial du bâtiment que les modèles a priori obtenus dans le chapitre précédent, ils permettront des études de scénarios performance énergétique plus précises. Notamment, les études de scénarios de flexibilité sont rendues possibles par la calibration des modèles. Cette étape permet d'identifier les paramètres liés à l'inertie thermique du bâtiment et de réduire les incertitudes sur les données d'entrée dynamique. Dans les paragraphes suivants, la méthode de calibration sera confrontée aux objectifs d'identification des dérives de consommation et des causes de ces dérives.

5.4.3 Identification de dérives de consommation

Un des objectifs de la thèse est l'identification de dérives de consommation et des causes de ces dérives. Pour tester la capacité du modèle à répondre à cet objectif, deux variantes du cas d'étude sont étudiées :

- TRNSYS_drg : Cette variante modélise une dégradation de l'isolation de la toiture par infiltration d'eau, l'épaisseur de l'isolant de la toiture est divisée par 4.

-TRNSYS_dmeca : Cette variante modélise un problème de régulation du système de ventilation, le système reste en permanence en fonctionnement.

La capacité du modèle à identifier ces dérives est étudiée sur les Figure 76 et Figure 77, respectivement pour les niveaux de connaissance du bâtiment : 2 et 5 (Tableau 36). Les modèles sont d'abord calibrés sur le cas d'étude TRNSYS sans dérive de consommation, puis sur une période de suivi, les données issues des modèles TRNSYS_drg et TRNSYS_dmeca sont comparées aux prédictions a posteriori, via les intervalles HPP95.

Dans le cas du niveau 2 de connaissance, les données de puissance thermique de chauffage sont considérées en entrée du modèle. Sur la période de suivi, les températures intérieures mesurées sont comparées aux résultats de simulations. Des écarts entre les températures prédites et mesurées indiquent alors une modification du système modélisé. Des températures mesurées plus basses que les températures simulées peuvent par exemple être la conséquence d'une dégradation de l'isolant ou du système de ventilation. C'est ce qui est observé en Figure 76.

Dans le cas du niveau 5 de connaissance, les températures de consigne sont calibrées, la puissance thermique de chauffage est une sortie du modèle comparée aux données mesurées. Dans la Figure 77, sur la période de suivi, les problèmes d'isolation et de ventilation mécanique entraînent des consommations plus élevées que celles simulées par le modèle.

La méthode semble alors en mesure d'identifier les dérives modélisées. Cependant, la capacité d'identification des dérives dépendra de l'ampleur de ces dernières et de la quantité de données disponibles.

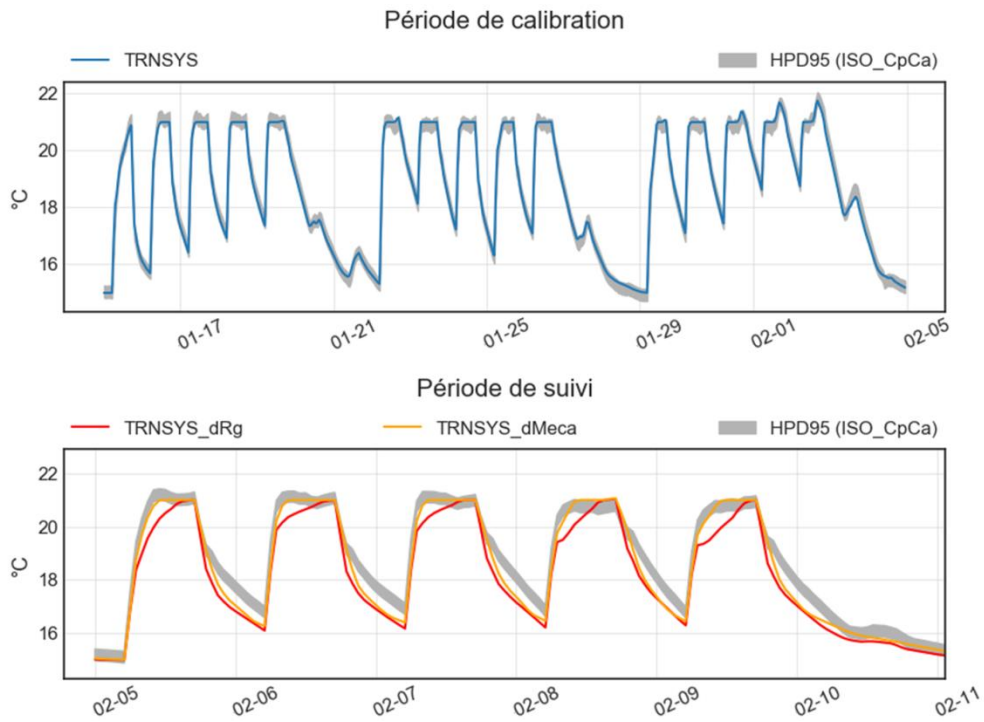


Figure 76 : Identification des dérives de température dans le cas du niveau 5 de connaissance du bâtiment (Abscisse : mois-jour)

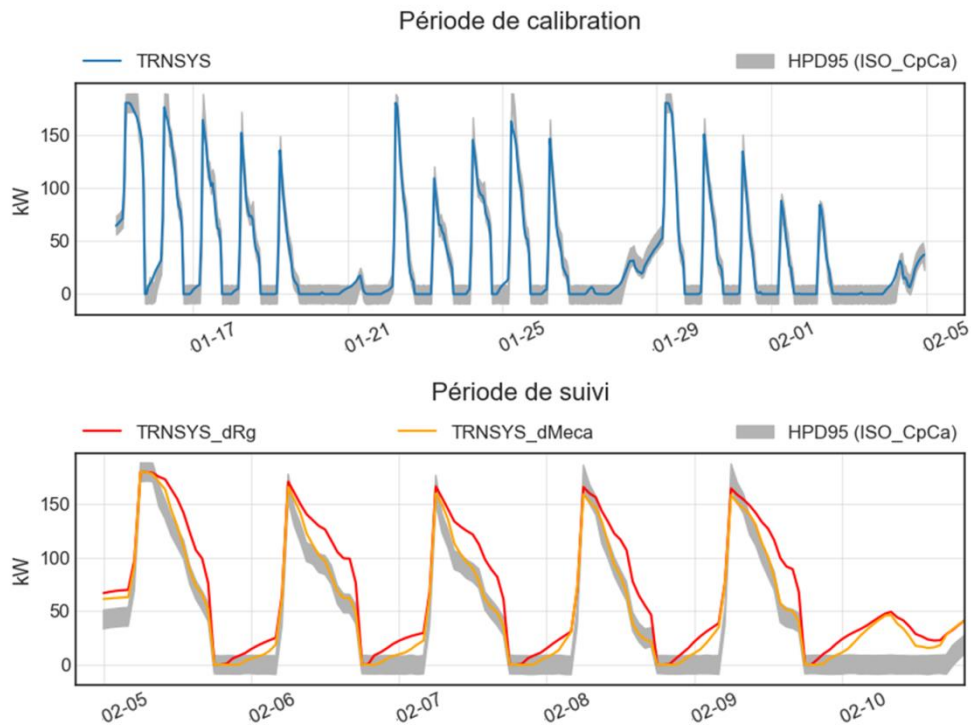


Figure 77 : Identification des dérives de consommation dans le cas du niveau 2 de connaissance du bâtiment 2 (Abscisse : mois-jour)

5.4.4 Identification des causes des dérives de consommation

Une fois que des dérives ont été identifiées, il est intéressant de chercher les causes de ces dérives. Pour cela, le modèle est recalibré sur la période présentant les dérives ; et les paramètres identifiés lors de cette nouvelle calibration sont comparés aux valeurs identifiées avant l'apparition des dérives.

Sur les Figure 78 et Figure 79, sont présentés, pour chaque niveau de connaissance des bâtiments, les résultats de recalibration du modèle sur les données d'exploitation respectivement issue des modèles TRNSYS_drg et TRNSYS_dmeca.

Sur la Figure 78, pour chaque scénario étudié, l'effet de la dégradation de l'isolation de la toiture sur la résistance Rg est clairement identifié (la valeur calibrée de la variable Rg passe d'environ 1 à 0,7). De même, en Figure 79, pour chaque scénario étudié, le paramètre $Meca$ est très faible, cela est symptomatique d'un problème de régulation. Lorsque le profil calibré ne correspond plus au profil réel, l'algorithme de calibration tend à annuler la sollicitation et compense avec d'autres paramètres. Ce phénomène a déjà été observé précédemment pour les gains thermiques liés aux consommations d'électricité spécifique.

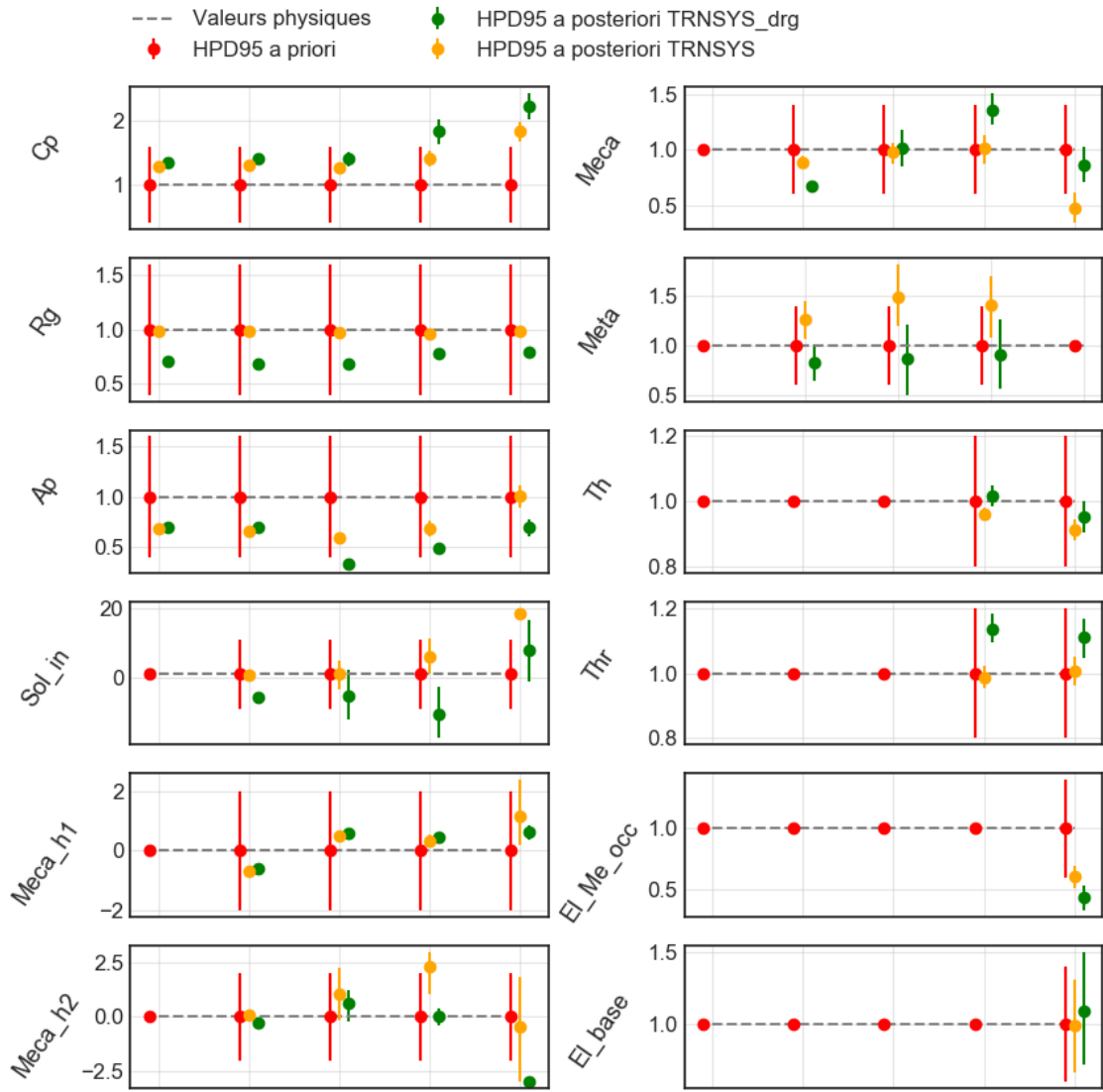


Figure 78 : Résultats de calibration sur les données d'exploitation issues du modèle TRNSYS_drg

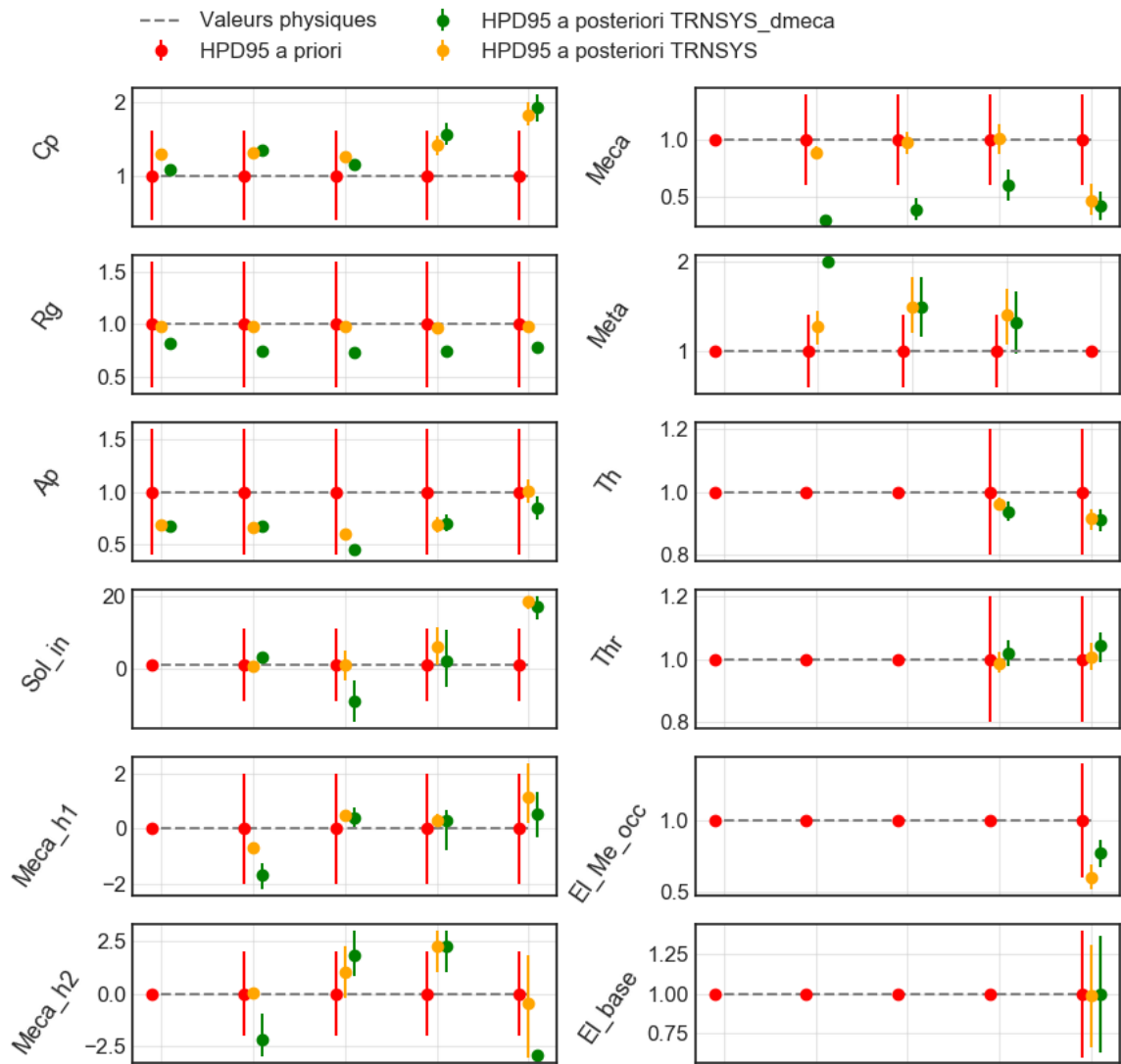


Figure 79 : Résultats de calibration sur les données d'exploitation issues du modèle TRNSYS_dmeca

5.5 Conclusion

La calibration des modèles permet d'identifier les paramètres incertains et d'atteindre de meilleures capacités prédictives. Les modèles calibrés peuvent être utilisés pour l'étude de scénarios de performance énergétique, ainsi que pour l'identification de dérives de consommation et des causes de ces dérives. Les possibilités d'évaluation d'actions de performance énergétique n'ont pas été étudiées spécifiquement, mais cela est similaire à la problématique de l'identification de dérives de consommation. Ces objectifs restent contraints par le niveau d'incertitude sur les données d'entrée. Par exemple, il ne sera pas possible d'identifier de faibles dérives avec d'importantes incertitudes. Aussi, bien que la calibration soit robuste aux incertitudes liées au comportement des occupants, des erreurs trop importantes sur les profils des données d'entrée dynamiques, peuvent mener à des valeurs de paramètres calibrés non-crédibles, et dégrader les capacités prédictives du modèle. Dans le cadre d'un suivi de performance, des paramètres calibrés non-crédibles peuvent être symptomatiques d'une dérive d'un système de régulation. Rappelons enfin que ces résultats concernent un bâtiment de bureaux, les possibilités de calibration seront différentes pour un bâtiment de logements collectifs. Dans ce dernier cas, il sera plus difficile d'obtenir des données d'exploitation, mais les effets de foisonnement sont susceptibles de réduire les incertitudes sur les profils des données d'entrée dynamiques. Ce cas d'étude pourra faire l'objet de futurs travaux.

Conclusion générale et perspectives

Le quartier constitue aujourd'hui une échelle privilégiée des collectivités pour leurs politiques de transition énergétique. Cette maille permet de mener des projets d'envergure, et offre des leviers d'action intéressants pour la baisse des consommations d'énergie et des émissions de CO₂. La rénovation mutualisée des bâtiments peut permettre des économies d'échelle, et des actions de flexibilité thermique et électrique peuvent aider à l'optimisation des réseaux d'électricité et de chaleur. Il semble alors pertinent de pouvoir faire appel à des modèles de quartier pour étudier des scénarios de rénovation et de flexibilité. De plus, partant du constat que les objectifs de performances énergétiques des quartiers ne sont souvent pas atteints, il semble intéressant de pouvoir suivre l'évolution de ces performances et évaluer les actions de performance énergétique mises en place. Dans ce contexte, l'idée centrale de la thèse était de proposer une méthode de construction de modèles de quartier s'appuyant sur des données disponibles à cette échelle, et de répondre aux objectifs suivants : étude de scénarios de performance énergétique (flexibilité, rénovation, sobriété), évaluation d'actions de performance énergétique mises en place (IPMVP), et identification de dérives de consommation et les causes de ces dérives. La méthode proposée consiste en la calibration bayésienne de modèles simplifiés de type RC. Des modèles dits a priori sont d'abord paramétrés à partir de base de données territoriales, de données construction (ex. DOE) et de données d'enrichissement issues de règles statistiques. Ces modèles a priori peuvent ensuite être calibrés par inférence bayésienne avec des données d'exploitation. Ces éléments de contexte et objectifs scientifiques ont été présentés dans le premier chapitre.

Le second chapitre a présenté un état de l'art des méthodes de calibration bayésienne appliquées à des modèles thermiques de bâtiments. Une méthode globale adaptée aux modèles RC est identifiée. Les différentes étapes de cette méthode ont été présentées : estimation des probabilités a priori, réduction du nombre de paramètres calibrés, inférence bayésienne, vérification de la convergence, et calcul des prédictions a posteriori. Certaines étapes ont nécessité des choix d'hypothèses ou d'algorithmes. Notamment l'algorithme d'échantillonnage «Slice sampling» a été

sélectionné pour sa robustesse et sa praticité. Aussi, trois fonctions de vraisemblance ont été proposées. Notamment un processus gaussien a été imaginé afin de tenir compte des corrélations temporelles entre les données.

Dans le troisième chapitre, un modèle RC a été sélectionné pour ses capacités prédictives et son identifiabilité, ce qui correspond dans cette thèse à la possibilité de convergence d'un algorithme de calibration vers des valeurs de paramètres correspondant à des propriétés physiques du bâtiment identifiées. L'identifiabilité permet notamment de définir des probabilités a priori, de renforcer la robustesse des modèles calibrés, d'étudier des scénarios de performance énergétique, et d'identifier des causes de dérives de consommation. L'identifiabilité des modèles RC a pourtant peu été étudiée dans la littérature. Le seul modèle disposant d'une méthode complète d'estimation physique de ses paramètres est celui proposé dans la norme ISO 13790. Un nouveau modèle RC a alors été construit sur la base de cette norme. Des modifications et un re-paramétrage ont permis d'améliorer son identifiabilité et ses capacités prédictives. Malgré cela, seul le paramètre de résistance globale peut aujourd'hui être précisément identifié. L'estimation physique des paramètres liés à l'inertie thermique des bâtiments (A_p , C_p) reste incertaine.

Le quatrième chapitre a proposé une méthode de paramétrage a priori des modèles à l'échelle du quartier. Les bases de données territoriales permettent de fournir de nombreuses informations sur les bâtiments : géométries, usages, années de construction. Cependant, ces bases restent incomplètes, des données supplémentaires et des corrections manuelles sont nécessaires. Les modèles a priori peuvent être utilisés pour étudier des scénarios de rénovation. En revanche, l'étude de scénarios de flexibilité via ces modèles a priori est limitée par les fortes incertitudes sur les paramètres C_p , A_p et les données d'entrées dynamiques. Pour l'étude de scénarios de flexibilité, l'utilisation de bâtiments types, modélisés sur un logiciel STD, sera probablement préférable à l'utilisation de modèles RC. Par ailleurs, des modèles de système tels que des pompes à chaleur ou des panneaux solaires, pourront être adjoints aux modèles RC afin d'étendre la diversité des scénarios de performance énergétique modélisables.

Dans le chapitre 5, des cas d'application basés sur un modèle de bâtiment de bureaux TRNSYS ont montré que la méthode de calibration bayésienne de modèles RC permet d'aboutir à des modèles capables de répondre aux objectifs de la thèse : étude de scénarios de performance énergétique, identification de dérives de

consommation et des causes de ces dérives, évaluation d'actions de performance énergétique mises en place. Une difficulté de la méthode réside dans l'incertitude associée aux données d'entrée dynamiques telles que le débit de ventilation ou l'indice de protection solaire, qui peuvent présenter des variations non prévisibles liées au comportement des occupants. Ces incertitudes sur les profils des données d'entrée sont modélisées par un nombre limité de variables calibrées. Bien que la méthode de calibration ait fait preuve de robustesse vis-vis de ces incertitudes, des différences trop importantes entre les profils réels et calibrés peuvent significativement dégrader les capacités prédictives des modèles.

Finalement, comme conjecturé en introduction, les modèles RC se présentent comme un intermédiaire entre les modèles de signature énergétique et les modèles STD, et présentent certains avantages par rapport à ces deux autres approches. Les modèles de signature constituent une méthode efficace pour suivre les performances mensuelles ou annuelles. Cependant, les modèles RC calibrés sont capables de réaliser un suivi plus fin, et peuvent plus facilement prendre en compte des changements dans les sollicitations telles qu'une modification du profil de température de consigne. Aussi, les modèles de signature ne permettent pas d'identifier les causes des dérives de consommation, ni d'étudier des scénarios de performance énergétique. Les modèles STD pourraient également répondre aux objectifs d'étude de scénarios et de suivi de performance, mais les temps de paramétrage et de calibration sont beaucoup plus longs. La méthode a été créée pour être applicable à la maille du quartier, mais reste intéressante à l'échelle du bâtiment, ou pour des parcs de bâtiments diffus appartenant à un même acteur qui souhaite suivre et améliorer les performances énergétiques de ses bâtiments.

Principales contributions

- Une méthode de calibration bayésienne de modèles RC a été proposée.
- Un modèle RC performant et identifiable a été construit sur la base de la norme ISO 13790
- Une méthode de paramétrage a priori de modèles RC à l'échelle du quartier a été développée. Les possibilités et les limites offertes par les différentes sources d'informations disponibles à cette échelle ont été présentées. Des avancées ont été faites pour l'enrichissement statistique des bâtiments de bureaux.

- Les possibilités et des limites des modèles RC ont été identifiées au regard des objectifs suivants : étude de scénarios de performance énergétique, identification de dérives de consommation et des causes de ces dérives, évaluation d'actions de performance énergétique.

Perspectives

Un effort reste à faire pour construire, pour chaque usage des bâtiments, une base de données d'enrichissement statistique complète, argumentée et validée. L'approche collaborative proposée de la plateforme CEA (Fonseca et al. 2016) semble pour cela prometteuse. L'étape de paramétrage a priori des modèles pourra dans le futur être facilitée par l'évolution des bases de données territoriales et un potentiel développement des données BIM.

En ce qui concerne les modèles RC, des études futures peuvent encore être menées dans le but d'améliorer l'identifiabilité des modèles, et plus spécifiquement les méthodes d'estimation physique des paramètres liés à l'inertie des bâtiments.

La méthode d'inférence bayésienne pourrait gagner en efficacité. Notamment un algorithme d'échantillonnage hamiltonien, tel que le NUTS pourrait être utilisé. Aussi, une solution pourrait être trouvée pour prendre en compte, dans la fonction de vraisemblance, les corrélations entre les données, sans trop augmenter les temps de calcul.

La méthode de calibration a pour le moment été testée sur un modèle TRNSYS de bâtiment de bureaux ; mais des cas d'étude supplémentaires sont nécessaires. Prioritairement un modèle détaillé de bâtiments de logement collectifs pourrait être considéré, puis des cas réels de bâtiments de bureaux et de logements pourront être étudiés.

Il serait intéressant d'étudier les possibilités offertes par les modèles bayésiens hiérarchiques qui, dans le cas de la calibration de modèles de bâtiments à l'échelle d'un quartier, peuvent permettre de prendre en compte l'information supplémentaire selon laquelle les bâtiments calibrés sont liés entre eux par des méthodes constructives communes. Selon cette approche hiérarchique, au lieu d'être fixé, les paramètres des distributions de probabilité a priori sont calibrés, et associés à des « hyper » lois a priori communes à l'ensemble des bâtiments du quartier.

Enfin, des questions perdurent concernant les jeux d'acteurs à l'échelle du quartier. Notamment, lorsque les bâtiments appartiennent à des acteurs multiples. Comment impliquer ces acteurs dans le suivi et l'amélioration des performances énergétiques du quartier en phase d'exploitation ? Peut-on créer une nouvelle entité « facilitateur énergétique », capable de regrouper les acteurs et de garantir une amélioration continue des performances énergétiques du quartier dans le temps ?

Table des figures

Figure 1 : Evolutions des besoins de flexibilité en volume d'énergie déplacée par jour pour répondre aux besoins du système électrique suivant 4 scénarios (RTE 2017)....	17
Figure 2 : Quartier de Clichy-Batignolles	18
Figure 3 : Processus d'évaluation d'actions de performance énergétique sur un quartier	20
Figure 4 : Extrait de la base de données BD TOPO® autour du quartier de Clichy-Batignolles (75017 Paris)	22
Figure 5 : Classification des approches de modélisation des consommations d'énergie territoriales (Swan et Ugursal 2009)	23
Figure 6 : Comparaison de trois approches de modélisation	27
Figure 7 : Méthode globale d'estimation de scénarios et de courbes de référence ajustées sur un quartier	31
Figure 8 : Méthode globale de calibration bayésienne.....	37
Figure 9 : Comparaison de diverses méthodes d'analyse de sensibilité (Kristensen et Petersen 2016)	38
Figure 10 : Illustration de l'effet de l'actualisation sur des données simulées	45
Figure 11 : Exemple d'une distribution bimodale incompatible avec l'utilisation du maximum a posteriori.....	46
Figure 12 : Distribution de la masse de probabilité de lois normales $N(0n, In)$, en fonction de la distance par rapport au mode	46
Figure 13 : Illustration d'une distribution de probabilité a posteriori résumée par une moyenne et un intervalle de crédence à posteriori HPD95.....	47
Figure 14 : Illustration de la méthode de « Slice sampling » par (Neal 2003).....	49
Figure 15 : Exemple d'un échantillonnage de prédictions a posteriori issues de la calibration d'un modèle linéaire sur des données bruitées.....	52
Figure 16 : Méthode globale retenue de calibration bayésienne	54
Figure 17 : Illustration des problèmes sur-apprentissage et sous-apprentissage.....	59
Figure 18 : Modèle R5C1 ISO	70
Figure 19 : Illustration de la notation utilisée pour les représentations de modèles RC	70

Figure 20 : Représentation détaillée du modèle R5C1 ISO	71
Figure 21 : Représentation étoile du modèle R5C1 ISO	71
Figure 22 : Représentation finale du modèle R5C1 ISO	72
Figure 23 : Comparaison entre les températures simulées par les modèles TRNSYS et ISO sur le cas d'étude mono-zone	81
Figure 24 : Représentation du modèle ISO_Ca	81
Figure 25 : Comparaison entre les températures simulées par les modèles TRNSYS et ISO_Ca sur le cas d'étude mono-zone.....	81
Figure 26 : Chaînes d'échantillonnage de la fonction de vraisemblance du modèle ISO_Ca appliqué au cas mono-zone	83
Figure 27 : Graphique des corrélations des valeurs échantillonnées du modèle ISO_Ca appliqué au cas mono-zone	83
Figure 28 : Coefficients totaux de l'analyse de sensibilité de Sobol, sur les dix paramètres du modèle ISO_Ca, avec barres d'incertitude à 95 %.....	84
Figure 29 : Chaînes d'échantillonnage de la fonction de vraisemblance du modèle ISO_Ca appliqué au cas mono-zone, avec un jeu de paramètres réduit	85
Figure 30 : Illustration de la division du modèle ISO_Ca en trois branches	86
Figure 31 : Chaînes d'échantillonnage de la fonction de vraisemblance du modèle ISO_Ca appliqué au cas mono-zone, avec re-paramétrage	89
Figure 32 : Graphique des corrélations des valeurs échantillonnées du modèle ISO_Ca appliqué au cas mono-zone, avec re-paramétrage.....	90
Figure 33 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de calibration dans le cas d'étude mono-zone (les courbes bleue et orange se superposent).....	90
Figure 34 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de validation de février dans le cas d'étude mono-zone (les courbes bleue et orange se superposent).....	91
Figure 35 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de validation de mars dans le cas d'étude mono-zone (les courbes bleue et orange se superposent).....	91
Figure 36 : Analyse de Sobol du modèle ISO_Ca re-paramétré.....	92
Figure 37 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de calibration, dans le cas de la configuration de base : 'Base' (les courbes bleue et orange se superposent)	97

Figure 38 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de validation de février, dans le cas de la configuration de base : ‘Base’	98
Figure 39 : Courbes de températures simulées par les modèle TRNSYS et ISO_Ca sur le jeu de données de validation de mars, dans le cas de la configuration de base : ‘Base’	98
Figure 40 : Modèle ISO_CpCa	99
Figure 41 : Plan du quartier de Clichy-Batignolles	109
Figure 42 : Superposition des données de la BD_TOPO® , avec le quartier de Clichy Batignolles	111
Figure 43 : Comparaison entre la BD_TOPO® et la table SIG construite manuellement.....	112
Figure 44 : Table SIG du quartier de Clichy-Batignolles	112
Figure 45 : Exemple d’identification de surfaces mitoyennes horizontales et verticales	113
Figure 46 : Surfaces mitoyennes horizontales et verticales identifiées sur le quartier de Clichy-Batignolles	114
Figure 47 : Modèle numérique de terrain utilisé en entrée de Daysim (extrait de MNT RGE ALTI® 2018)	115
Figure 48 : Modèle du quartier en entrée du logiciel Daysim, créé par la plateforme CEA (Fonseca et al. 2016) à partir de la table SIG	116
Figure 49 : Représentation de la base de données des fichiers fonciers (Cerema 2017)	121
Figure 50 : Comparaison pour un même local, des surfaces : Carrez, habitable et surface réelle (stoth) (Cerema s. d.).....	121
Figure 51 : Superposition des couches SIG : contours de parcelles issus de la base de données MAJIC, base de données BD_TOPO® et modèle SIG du quartier	122
Figure 52 : Illustration de la méthode liant les surfaces chauffées aux informations contenues dans la table SIG.....	124
Figure 53 : Exemple d’un semainier rempli dans le cadre de l’Enquête emplois du temps de l’INSEE	130
Figure 54 : Profil d’occupation des bureaux (<i>Occ</i>) sur une année.....	132
Figure 55 : Profil d’occupation des bureaux (<i>Occ</i>) sur une semaine	133

Figure 56 : Comparaison, sur un bâtiment de bureaux du quartier de Clichy-Batignolles, de données de consommation mensuelles avec les résultats de simulation du modèle ISO_CpCa a priori.....	136
Figure 57: Comparaison, sur un bâtiment de bureaux du quartier de Clichy-Batignolles, de données de consommation horaires avec les résultats de simulation du modèle ISO_CpCa a priori.....	137
Figure 58 : Comparaison, sur un bâtiment résidentiel du quartier de Clichy-Batignolles, de données de consommation mensuelles avec les résultats de simulation du modèle ISO_CpCa a priori.....	137
Figure 59 : Comparaison, sur un bâtiment résidentiel du quartier de Clichy-Batignolles, de données de consommation horaires avec les résultats de simulation du modèle ISO_CpCa a priori.....	138
Figure 60 : Loi de probabilité normale $N1; 0,3$	144
Figure 61 : Loi de probabilité gamma $G0,05; 2$	144
Figure 62 : Distribution des probabilités a posteriori échantillonnées en considérant la fonction de vraisemblance <i>Li. i. d</i>	145
Figure 63 : Prédictions de températures a priori et a posteriori sur la période de calibration dans le cas de la fonction de vraisemblance <i>Li. i. d</i>	146
Figure 64 : Analyse des résidus, pour le cas d'étude de la fonction de vraisemblance <i>Li. i. d</i>	147
Figure 65 : Prédictions de températures a priori et a posteriori sur la période de calibration, pour le cas d'étude de la fonction de vraisemblance <i>LGP</i>	149
Figure 66 : Matrice de covariance a posteriori ΣLGP	149
Figure 67 : Prédictions de températures a priori et a posteriori sur la période de calibration pour le cas de la fonction de vraisemblance <i>LKal</i>	151
Figure 68 : Prédictions de températures a priori et a posteriori sur la période de calibration (cas de la fonction de vraisemblance <i>LKal</i>).....	152
Figure 69 : Distribution a priori des scénarios des apports solaires internes (Abscisse : mois-jour heure).....	159
Figure 70 : Distribution a priori des scénarios de ventilation mécanique (<i>Umeca</i>) (Abscisse : mois-jour heure).....	160
Figure 71 : Distribution a priori des scénarios d'apports métaboliques (Abscisse : mois-jour heure).....	160

Figure 72 : Distribution a priori des scénarios de consigne de température (Abscisse : mois-jour heure).....	161
Figure 73 : Distribution a priori des scénarios des apports thermiques liés aux consommations d'électricité spécifique (Abscisse : mois-jour heure)	162
Figure 74 : Distribution a priori et a posteriori des paramètres calibrés pour les 5 niveaux de connaissance considérés.....	164
Figure 75 : Indicateurs caractérisant les prédictions a posteriori de température et de puissance, pour les 5 niveaux de connaissance considérés	165
Figure 76 : Identification des dérives de température dans le cas du niveau 5 de connaissance du bâtiment (Abscisse : mois-jour)	168
Figure 77 : Identification des dérives de consommation dans le cas du niveau 2 de connaissance du bâtiment 2 (Abscisse : mois-jour)	168
Figure 78 : Résultats de calibration sur les données d'exploitation issues du modèle TRNSYS_drg.....	170
Figure 79 : Résultats de calibration sur les données d'exploitation issues du modèle TRNSYS_dmeca	171
Figure 80 : Schémas d'une paroi monocouche.....	196
Figure 81 : Représentation en π d'un quadripôle.....	198

Liste des tableaux

Tableau 1 : Inventaire de publications sur les modèles RC ;.....	64
Tableau 2 : Description du cas d'étude : un bâtiment de bureau type.	78
Tableau 3 : Résumé du calcul de C_m selon les méthodes : détaillée, simplifiée et à point, pour le cas de l'étage de bureau mono-zone	78
Tableau 4 : Résultats du calcul de C_m et A_m selon les méthodes : détaillée, simplifiée et à point, pour le cas de l'étage de bureau mono-zone	79
Tableau 5 : Estimation physique des résistances du modèle ISO	79
Tableau 6 : RMSE entre les températures simulées par les modèle TRNSYS et ISO_Ca calibré, sur le cas d'étude mono-zone	92
Tableau 7 : Description des scénarios étudiés	93
Tableau 8 : Résultats de calibration du modèle ISO_Ca sur chacun des 12 scénarios étudiés.....	94
Tableau 9 : Comparaison entre les valeurs des paramètres $Rate_m$ et $Rate_I$ calibrés et non normalisés, avec les estimations physiques de ces derniers.....	96
Tableau 10 : Résultats de calibration du modèle ISO_CpCa sur chacun des 12 scénarios étudiés	100
Tableau 11 : Comparaison entre les valeurs des paramètres $Rate_m$ et $Rate_I$ calibrés et non normalisés, avec les estimations physiques de ces derniers.....	101
Tableau 12 : Récapitulatif des différentes définitions de surfaces de bâtiments	118
Tableau 13 : Typologies des bâtiments de bureaux (Alessandrini et al. 2003)	126
Tableau 14 : Classe d'inertie des bâtiments de bureaux selon leurs typologies (Alessandrini et al. 2003)	126
Tableau 15 : Valeurs par défaut de C_m et A_m selon la classe d'inertie du bâtiment (AFNOR 2013)	126
Tableau 16 : Facteur correctif ϵ (AFNOR 2004).....	128
Tableau 17 : Facteur correctif e (AFNOR 2004)	128
Tableau 18 : Taux de renouvellement de l'air par default (AFNOR 2017a).....	129
Tableau 19 : Attributs concernant le type de la journée enquêtée, issus de la table individu de l'enquête emplois du temps.....	131
Tableau 20 : Répartition mensuelle des jours de congés pris par les actifs.....	132
Tableau 21 : Répartition mensuelle des jours d'arrêt maladie pris par les actifs.	132

Tableau 22 : Répartition selon la taille et le lieu géographique des ratios de surfaces chauffées par personne : <i>rpers</i> (ARENE 2009).....	133
Tableau 23 : Résumé des distributions a priori.....	145
Tableau 24 : Résumé statistique des distributions a posteriori échantillonnées en considérant la fonction de vraisemblance <i>Li. i. d.</i>	146
Tableau 25 : RMSE a priori et a posteriori pour les périodes de calibration et de validation dans le cas de la fonction de vraisemblance <i>Li. i. d.</i>	146
Tableau 26 : Résumé des distributions a priori pour le cas d'étude de la fonction de vraisemblance <i>LGP</i>	147
Tableau 27 : Résumé statistique des distributions a posteriori pour le cas d'étude de la fonction de vraisemblance <i>LGP</i>	148
Tableau 28 : RMSE a priori et a posteriori pour les périodes de calibration et de validation, pour le cas d'étude de la fonction de vraisemblance <i>LGP</i>	148
Tableau 29 : Résumé des distributions a priori (cas de la fonction de vraisemblance <i>LKal</i>).....	150
Tableau 30 : Résumé statistique des distributions a posteriori pour le cas de la fonction de vraisemblance <i>LKal</i>	151
Tableau 31: RMSE a priori et a posteriori pour le cas de la fonction de vraisemblance <i>LKal</i>	151
Tableau 32 : Analyse de l'influence du degré d'actualisation sur les résultats de calibration.....	153
Tableau 33 : RMSE (°C) entre la température de référence Tref, et les températures calculées selon les différents plans d'instrumentation étudiés.....	155
Tableau 34 : Résultats de calibration selon différents plans d'instrumentation, pour le scénario « Base ».....	155
Tableau 35 : Résultats de calibration selon différents plans d'instrumentation, pour le scénario « Zones ».....	155
Tableau 36 : Résumé des différents niveaux de connaissance du bâtiment envisagés.....	157
Tableau 37 : Résumé des distributions a priori.....	163

Bibliographie

- (IEA), The International Energy Agency. 2018. « Key world energy statistics 2018 energy statistics ».
- AFNOR. 2004. « NF EN 12831 : Systèmes de chauffage dans les bâtiments Méthode de calcul des déperditions calorifiques de base ». 1-74.
- AFNOR. 2010. « Nf EN ISO 13786 : Performance thermique des composants de bâtiment Caractéristiques thermiques dynamiques ».
- AFNOR. 2013. « Nf EN ISO 13790 : Performance énergétique des bâtiments — Calcul des besoins d'énergie pour le chauffage et le refroidissement des locaux ».
- AFNOR. 2017a. « NF EN 12831-1 : Performance énergétique des bâtiments — Méthode de calcul de la charge thermique nominale — Partie 1 : Charge de chauffage des locaux, module M3-3 ».
- AFNOR. 2017b. « NF EN ISO 13789 : Performance thermique des bâtiments — Coefficients de transfert de chaleur par transmission et par renouvellement d'air — Méthode de calcul ».
- Alessandrini, Jean-Marie, E. Fleury, S. Filfli, et Dominique Marchio. 2003. *Quelles solutions pour des bâtiments tertiaires climatisés à moins de 100kWh/m²/an ? Rapport intermédiaire 1 Description des bâtiments types.*
- Alessandrini, Jean-Marie, Emmanuel Fleury, Sila Filfli, et Dominique Marchio. 2006. « Impact de la gestion de l'éclairage et des protections solaires sur la consommation d'énergie de bâtiments de bureaux climatisés. » P. 227-36 in *Climamed Congress*. Lyon.
- ARENE. 2009. *Etude sur les consommations d'énergie dans les bureaux en Île-de-France.*
- ASHRAE. 2002. *Guileline 14-2002 Measurement of Energy and Demand Savinf.*
- ASHRAE. 2013. *Ashrae Handbook*,.
- Bacher, Peder, et Henrik Madsen. 2011. « Identifying suitable models for the heat dynamics of buildings ». *Energy and Buildings* 43(7):1511-22. doi: 10.1016/j.enbuild.2011.02.005.
- Baude, Manuel, François-Xavier Dussud, Mathieu Ecoiffier, Jérôme Duvernoy, et Charlotte Vailles. 2018. « Chiffres clés du climat France, Europe et Monde - Edition 2018 ». *Commissariat général au développement durable* 80.
- Berthou, Thomas. 2013. « Développement de modèles de bâtiment pour la prévision de charge de climatisation et l'élaboration de stratégies d'optimisation énergétique et d'effacement ». <http://www.theses.fr>.
- Berthou, Thomas, Bruno Duplessis, Philippe Rivière, Pascal Stabat, Damien Casetta, et Dominique Marchio. 2015. « SMART-E: A tool for energy demand simulation

and optimization at the city scale ».

- Berthou, Thomas, Bruno Duplessis, Pascal Stabat, Philippe Rivière, et Marchio Dominique. 2019. « Urban Energy Models Validation in Data Scarcity Context: Case of the Electricity Consumption in the French Residential Sector Thomas ». in *16th International Conference of the « International Building Performance Simulation Association » (IBPSA)*.
- Berthou, Thomas, Pascal Stabat, Raphael Salvazet, et Dominique Marchio. 2014. « Development and validation of a gray box model to predict thermal behavior of occupied office buildings ». *Energy and Buildings* 74:91-100. doi: 10.1016/J.ENBUILD.2014.01.038.
- Betancourt, Michael. 2018. *A Conceptual Introduction to Hamiltonian Monte Carlo*.
- BMVBS. 2010. « Vereinfachung zur geometrischen und technischen Datenaufnahme im Nichtwohngebäudebestand. »
- Brastein, O. M., D. W. U. Perera, C. Pfeifer, et N. O. Skeie. 2018. « Parameter estimation for grey-box models of building thermal behaviour ». *Energy and Buildings* 169:58-68. doi: 10.1016/J.ENBUILD.2018.03.057.
- Cerema. 2017. *Guide de prise en main des Fichiers fonciers*.
- Cerema. s. d. « dictionnaire des Fichiers fonciers - Fichiers fonciers Tables Principales (FFTP) ». Consulté (<http://doc-datafoncier.cerema.fr/ff/docff/>).
- Chaudhary, Gaurav, Joshua New, Jibonananda Sanyal, Piljae Im, Zheng O'Neill, et Vishal Garg. 2016. « Evaluation of “Autotune” calibration against manual calibration of building energy models ». *Applied Energy* 182:115-34. doi: 10.1016/J.APENERGY.2016.08.073.
- Chong, Adrian, Khee Poh Lam, Matteo Pozzi, et Junjing Yang. 2017. « Bayesian calibration of building energy models with large datasets ». *Energy and Buildings* 154:343-55. doi: 10.1016/J.ENBUILD.2017.08.069.
- Chong, Adrian, et Kathrin Menberg. 2018. « Guidelines for the Bayesian calibration of building energy models ». *Energy and Buildings* 174:527-47. doi: 10.1016/J.ENBUILD.2018.06.028.
- Coakley, Daniel, Paul Raftery, et Marcus Keane. 2014. « A review of methods to match building energy simulation models to measured data ». *Renewable and Sustainable Energy Reviews* 37:123-41. doi: 10.1016/J.RSER.2014.05.007.
- CSTB. 2017. « Règles Th-bat - Fascicule Inertie ».
- Darmayan, Loïc, Pablo Puerto, Gaëtan Cherix, et Massimiliano Capezzali. 2016. « Système de management énergétique territorial ». *AQUA & GAS*.
- EVO. 2010. *Protocole International de Mesure et de Vérification de la Performance énergétique Concepts et options pour l'évaluation des économies d'énergie et d'eau Volume 1*.
- Farzana, Shazia, Meng Liu, Andrew Baldwin, et Md. Uzzal Hossain. 2014. « Multi-model prediction and simulation of residential building energy in urban areas of Chongqing, South West China ». *Energy and Buildings* 81:161-69. doi:

10.1016/J.ENBUILD.2014.06.007.

- Fonseca, Jimeno A., Thuy-An Nguyen, Arno Schlueter, et Francois Marechal. 2016. « City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts ». *Energy and Buildings* 113:202-26. doi: 10.1016/J.ENBUILD.2015.11.055.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, et Donald B. Rubin. 2014. *Bayesian Data Analysis. 3rd edn.* édité par Taylor & Francis Group. Chapman Texts in Statistical Science.
- Gelman, Andrew, et Donald Rubin. 1992. « Inference from Iterative Simulation Using Multiple Sequences ». *Statistical Science* 7(4):457-511.
- Harb, Hassan, Neven Boyanov, Luis Hernandez, Rita Streblov, et Dirk Müller. 2016. « Development and validation of grey-box models for forecasting the thermal response of occupied buildings ». *Energy and Buildings*. doi: 10.1016/j.enbuild.2016.02.021.
- Hastie, Trevor, Robert Tibshirani, et Jerome Friedman. 2008. *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction.*
- Hedegaard, Rasmus Elbæk, Martin Heine Kristensen, Steffen Petersen, Adam Brun, et Steffen Petersen. 2019. « Bottom-up modelling methodology for urban-scale analysis of residential space heating demand response ». *Applied Energy journal* 242:181-204. doi: 10.1016/j.enbuild.2018.07.030.
- Hedegaard, Rasmus Elbæk, et Steffen Petersen. 2017. « Evaluation of Grey-Box Model Parameter Estimates Intended for Thermal Characterization of Buildings ». in *Energy Procedia*.
- Heo, Y., R. Choudhary, et G. A. Augenbroe. 2012. « Calibration of building energy models for retrofit analysis under uncertainty ». *Energy and Buildings* 47:550-60. doi: 10.1016/J.ENBUILD.2011.12.029.
- Higdon, Dave, James Cavendish, John Cafeo, et Robert D. Ryne. 2006. « Combining Field Data and Computer Simulations for Calibration and Prediction ».
- Hoang, Lê Nguyễn. 2018. « La formule du savoir : Une philosophie unifiée du savoir fondée sur le théorème de Bayes ».
- IEE Project. 2016. « European Projects TABULA & EPISCOPE ». Consulté (<https://episcope.eu/welcome/>).
- iNSPiRe Project. 2014. « D2 . 1a - Survey on the energy needs and architectural features of the EU building stock ».
- International Energy Agency. 2020. « IEA – International Energy ». Consulté (<https://www.iea.org/statistics/?country=FRANCE&year=2016&category=Energy supply&indicator=TPESbySource&mode=chart&dataTable=BALANCES>).
- IPCC AR5. 2014. *Fifth Assessment Report of the Intergovernmental Panel on Climate Change.*
- ISO. 2008. *ISO - ISO 13790:2008 - Performance énergétique des bâtiments -- Calcul*

des besoins d'énergie pour le chauffage et le refroidissement des locaux.

- Kang, Yoonsuk, et Moncef Krarti. 2016. « Bayesian-Emulator based parameter identification for calibrating energy models for existing buildings ». *Building Simulation* 9(4):411-28. doi: 10.1007/s12273-016-0291-6.
- Kennedy, Marc C., et Anthony O'Hagan. 2001. « Bayesian calibration of computer models ». *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(3):425-64. doi: 10.1111/1467-9868.00294.
- Kokogiannakis, Georgios, Paul Strachan, et Joe Clarke. 2008. « Comparison of the simplified methods of the ISO 13790 standard and detailed modelling programs in a regulatory context ». *Journal of Building Performance Simulation* 1(4):209-19.
- Kristensen, Martin Heine, Ruchi Choudhary, et Steffen Petersen. 2017. « Bayesian calibration of building energy models: Comparison of predictive accuracy using metered utility data of different temporal resolution ». *Energy Procedia* 122:277-82. doi: 10.1016/J.EGYPRO.2017.07.322.
- Kristensen, Martin Heine, et Steffen Petersen. 2016. « Choosing the appropriate sensitivity analysis method for building energy model-based investigations ». *Energy and Buildings* 130:166-76. doi: 10.1016/J.ENBUILD.2016.08.038.
- Larochelle, Hugo. 2015. « Cours IFT 603, Techniques d'apprentissage (machine learning) ». Consulté 4 mars 2021 (http://www.dmi.usherb.ca/~larocheh/cours/ift603_H2015/contenu.html).
- Lefort, Lucie, Romain Bonabe De Rougé, Pascal Schetelat, Thomas Berthou, Bruno Duplessis, Peter Riederer, Eric Peirano, Lucie Lefort, Romain Bonabe De Rougé, Pascal Schetelat, Thomas Berthou, et Bruno Duplessis. 2021. « Méthodologie de validation d ' outils de simulation énergétique urbaine avec un cas d ' application sur la Seine-et-Marne To cite this version : HAL Id : hal-03202696 ».
- Légifrance. 2017. « Décret n° 2017-486 du 5 avril 2017 relatif au traitement et à la mise à disposition du public des données détaillées de comptage des gestionnaires des réseaux de transport et de distribution d'électricité et de gaz naturel ». Consulté (<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000034379228>).
- Li, Wenliang, Yuyu Zhou, Kristen Cetin, Jiyong Eom, Yu Wang, Gang Chen, et Xuesong Zhang. 2017. « Modeling urban building energy use: A review of modeling approaches and procedures ». *Energy* 141:2445-57. doi: 10.1016/J.ENERGY.2017.11.071.
- Lim, Hyunwoo, et Zhiqiang John Zhai. 2017. « Review on stochastic modeling methods for building stock energy prediction ». *Building Simulation* 10(5):607-24. doi: 10.1007/s12273-017-0383-y.
- Ling, You, Joshua Mullins, et Sankaran Mahadevan. 2014. « Selection of model discrepancy priors in Bayesian calibration ». *Journal of Computational Physics* 276:665-80. doi: 10.1016/j.jcp.2014.08.005.
- MacKay, David J. C. 2003. *Information Theory, Inference, and Learning Algorithms*.

- Madsen, H., et J. Holst. 1995. « Estimation of continuous-time models for the heat dynamics of a building ». *Energy and Buildings* 22(1):67-79. doi: 10.1016/0378-7788(94)00904-X.
- Manfren, Massimiliano, Niccolò Aste, et Reza Moshksar. 2013. « Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation ». *Applied Energy* 103:627-41. doi: 10.1016/J.APENERGY.2012.10.031.
- Martin, Osvaldo. 2018. *Bayesian Analysis with python (Second Edition)*.
- Michalak, Piotr. 2014. « The simple hourly method of EN ISO 13790 standard in Matlab / Simulink : A comparative study for the climatic conditions of Poland ». *Energy* 75:568-78. doi: 10.1016/j.energy.2014.08.019.
- Michalak, Piotr. 2019. « A thermal network model for the dynamic simulation of the energy performance of buildings with the time varying ventilation flow ». *Energy and Buildings* 202. doi: 10.1016/j.enbuild.2019.109337.
- Mikkola, Jani, et Peter D. Lund. 2014. « Models for generating place and time dependent urban energy demand profiles ». *Applied Energy* 130:256-64. doi: 10.1016/J.APENERGY.2014.05.039.
- Neal, Radford M. 2003. « SLICE SAMPLING ». *The Annals of Statistics* 31(3):705-67.
- NégaWatt. 2017a. *La sobriété énergétique La sobriété énergétique : consommer moins pour vivre mieux*.
- NégaWatt. 2017b. « Scénario négaWatt 2017-2050 ».
- ODYSSEE-MURE. 2015. *Energy Efficiency Trends and Policies in the Household and Tertiary Sectors An Analysis Based on the ODYSSEE and MURE Databases*.
- Orgill, J. F., et K. G. T. Hollands. 1977. « Correlation equation for hourly diffuse radiation on a horizontal surface ». *Solar Energy* 19(4):357-59. doi: 10.1016/0038-092X(77)90006-8.
- Raillon, L., et C. Ghiaus. 2018. « An efficient Bayesian experimental calibration of dynamic thermal models ». *Energy* 152:818-33.
doi: 10.1016/J.ENERGY.2018.03.168.
- Reinhart, Christoph. s. d. « Daysim ».
Consulté (<https://daysim.ning.com/page/credits>).
- Reinhart, Christoph F. 2004. « Lightswitch-2002: a model for manual and automated control of electric lighting and blinds ». *Solar Energy* 77(1):15-28.
doi: 10.1016/J.SOLENER.2004.04.003.
- Reinhart, Christoph F., et Marilyn Andersen. 2006. « Development and validation of a Radiance model for a translucent panel ». *Energy and Buildings* 38(7):890-904.
doi: 10.1016/J.ENBUILD.2006.03.006.
- Reinhart, Christoph F., Timur Dogan, J. Alstan Jakubiec, Tarek Rakha, et Andrew Sang. s. d. « UMI -An urban simulation environment for building energy use,

daylighting and walkability ».

- Reinhart, Christoph F., et Sebastian Herkel. 2000. « The simulation of annual daylight illuminance distributions — a state-of-the-art comparison of six RADIANCE-based methods ». *Energy and Buildings* 32(2):167-87.
doi: 10.1016/S0378-7788(00)00042-6.
- Reinhart, Christoph F., et Oliver Walkenhorst. 2001. « Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds ». *Energy and Buildings* 33(7):683-97. doi: 10.1016/S0378-7788(01)00058-5.
- Remmen, Peter, Moritz Lauster, Michael Mans, Marcus Fuchs, Tanja Osterhage, et Dirk Müller. 2018. « TEASER: an open tool for urban energy modelling of building stocks ». *Journal of Building Performance Simulation* 11(1):84-98. doi: 10.1080/19401493.2017.1283539.
- Reynders, G., J. Diriken, et D. Saelens. 2014. « Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals ». *Energy and Buildings*. doi: 10.1016/j.enbuild.2014.07.025.
- Robinson, Darren. 2009. « CITYSIM: COMPREHENSIVE MICRO-SIMULATION OF RESOURCE FLOWS FOR SUSTAINABLE URBAN PLANNING ». in *Eleventh International IBPSA Conference*. Glasgow.
- Rouchier, Simon. 2018. « Solving inverse problems in building physics: An overview of guidelines for a careful and optimal use of data ». *Energy and Buildings* 166(February):178-95. doi: 10.1016/j.enbuild.2018.02.009.
- Roujol, S., E. Fleury, D. Marchio, J. R. Millet, P. Stabat, 2003 « Testing the energy simulation building model of ConsoClim using BESTEST method and experimental data ». *Eighth International IBPSA Conference*. Eindhoven.
- RTE. 2017. « Bilan prévisionnel ».
- Sartori, Igor, Bjørn Jensen Wachenfeldt, et Anne Grete Hestnes. 2009. « Energy demand in the Norwegian building stock: Scenarios on potential reduction ». *Energy Policy* 37(5):1614-27. doi: 10.1016/J.ENPOL.2008.12.031.
- Da Silva, David, Jean-Marie Alessandrini, Jean-Baptiste Videau, et Jean-Robert Millet. 2016. « Evaluation et perspectives du modèle thermique de COMETH , le cœur de calcul de la réglementation thermique des bâtiments neufs ». *IBPSA-France* (May).
- Sobol, I. M. 2001. « Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates ». *Mathematics and Computers in Simulation* 55(1-3):271-80. doi: 10.1016/S0378-4754(00)00270-6.
- Sobol, Ilya Meyerovich. 1993. « Sensitivity estimates for nonlinear mathematical models ». *Math. Model. Comput.* (4):407-414.
- Swan, Lukas G., et V. Ismet Ugursal. 2009. « Modeling of end-use energy consumption in the residential sector: A review of modeling techniques ». *Renewable and Sustainable Energy Reviews* 13(8):1819-35.
doi: 10.1016/J.RSER.2008.09.033.

- Swiss Society of Engineers and Architects. 2006. *SIA Merkblatt 2024: Standard-Nutzungsbedingungen für die Energie- und Gebäudetechnik*.
- The Shift Project. 2017a. *6-Batiment-résidentiel-privé-version-longue*.
- The Shift Project. 2017b. *7-Batiment-public-version-longue*.
- Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, et Paul-Christian Bürkner. 2019. « Rank-normalization, folding, and localization: An improved Rhat for assessing convergence of MCMC ».
- Vivian, Jacopo, Angelo Zarrella, Giuseppe Emmi, et Michele De Carli. 2017. « An evaluation of the suitability of lumped-capacitance models in calculating energy needs and thermal behaviour of buildings ». *Energy and Buildings* 150:447-65. doi: 10.1016/j.enbuild.2017.06.021.
- Wang, Zequn, et Yuxiang Chen. 2019. « Data-driven modeling of building thermal dynamics: Methodology and state of the art ». *Energy and Buildings* 203:109405. doi: 10.1016/j.enbuild.2019.109405.
- Yamaguchi, Yohei, R. Choudhary, A. Booth, Y. Suzuki, et Yoshiyuki Shimoda. 2013. « Urban-scale energy modelling of food supermarket considering uncertainty ». in *The proceedings of BS*. Chambéry, France.
- Yuan, Jun, et Bin Su. 2017. « A Meta Model Based Bayesian Approach for Building Energy Models Calibration ». *Energy Procedia* 143:161-66. doi: 10.1016/J.EGYPRO.2017.12.665.
- Zarrella, Angelo, Enrico Pratavia, Pierdonato Romano, Laura Carnieletto, et Jacopo Vivian. 2020. « Analysis and application of a lumped-capacitance model for urban building energy modelling ». *Sustainable Cities and Society* 63(July):102450. doi: 10.1016/j.scs.2020.102450.

Annexe 1 : Inventaire des bases de données territoriales

Nom:	BD TOPO® Source: Institut National de Géographie (IGN)		
Echelle géographique	Bâtiments		
Entité concernée	Bâtiments	Systèmes	Occupants
	✓		
Attributs	<p>La base de données BD TOPO® contient les contours des bâtiments, géolocalisées selon un système de coordonnées de référence (SCR), et caractérisés par des attributs :</p> <ul style="list-style-type: none"> - hauteurs, - usages (indifférencié, industriel). <p>Depuis 2019: la version 3.0 de la BD TOPO dispose d'attributs complémentaires issus de la base des fichiers fonciers MAJIC :</p> <ul style="list-style-type: none"> - années de construction, - nombres de logements, - nombres d'étages, - usages (résidentiel, commercial et services, industriel). <p>Cependant ces informations restent encore incomplètes, notamment pour les établissements commerciaux.</p>		
Date	Mise à jour plusieurs fois par an		
Licence/ Conditions d'accès	Ouvert depuis 2019		

Nom:	Répertoire des immeubles localisés (RIL)		
Echelle géographique	Adresse (Disponible pour les villes de plus de 10 000 habitants)		
	Bâtiments	Système	Occupants
	✓		
Attributs	<p>Ce répertoire permet de constituer les bases de sondage des enquêtes de recensement des communes de 10000 habitants ou plus.</p> <p>Pour chaque adresse répertoriée les attributs suivants sont renseignés :</p> <ul style="list-style-type: none"> - le nombre de logements, - l'année de construction la plus récente dans l'ensemble des immeubles à cette adresse, - le nombre d'établissements SIRENE <p>Remarque : les données sur les années de construction sont fortement incomplètes pour les établissements professionnels.</p>		
Date	Mise à jour annuelle depuis 2013		
Licence/ Conditions d'accès	Sur devis		

Nom:	Fichiers Logements ordinaires Source: INSEE (Recensement de la population)		
Echelle géographique	IRIS		
Entité concernée	Bâtiments	Systèmes	Occupants
	✓	✓	✓
Contenu/Attributs	<p>Chaque enregistrement correspond à un logement issu du recensement de la population, caractérisé par de nombreux attributs tels que :</p> <ul style="list-style-type: none"> - poids statistique du logement, - date de construction, - nombre de pièces, - superficie du logement (par tranches), - catégorie de logement, - type de logement (maison, appartement). - combustible principal du logement : chauffage urbain, gaz de ville ou de réseau, fioul (mazout), électricité, gaz en bouteilles ou en citerne, autre, - nombre d'habitants par classes d'âge - statut d'occupation détaillé du logement. 		
Date	Mise à jour annuelle depuis 2006		
Licence/ Conditions d'accès	Ouvert		

Nom:	Sitadel2 Source : _ Ministère de la transition écologique et solidaire		
Echelle géographique	Commune		
Entité concernée	Bâtiments	Système	Occupants
	✓	✓	✓
Contenu/Attributs	Données issues des permis de construire "Parmi l'ensemble des données collectées, les principales sont : <ul style="list-style-type: none"> - le nombre et la surface de logements créés, - la surface créée des locaux non résidentiels. Plusieurs types de construction sont distingués : <ul style="list-style-type: none"> - quatre types de logements : logement individuel pur, logement individuel groupé, logement collectif, logement en résidence, - huit types de locaux non résidentiels : hébergement hôtelier, commerces, bureaux, artisanat, bâtiments industriels, entrepôts, exploitation agricole et forestière, services publics ou d'intérêt collectif. La catégorie « service public ou d'intérêt collectif » regroupe les six sous-catégories suivantes : transports, enseignement et recherche, action sociale, ouvrage spécial, santé, culture et loisirs." 		
Date	Mise à jour mensuelle		
Licence/ Conditions d'accès	Ouvert		

Nom:	MAJIC, (Mise À Jour des Informations Cadastrales) Source: Direction Générale des Impôts (DGI)		
Echelle géographique	Adresse		
Entité concernée	Bâtiments	Système	Occupants
	✓		
Contenu/Attributs	Cette base de données contient l'ensemble des informations nécessaires à l'évaluation de la fiscalité. Les locaux d'habitation, commerciaux et industriels sont décrits selon de nombreux attributs tels que : <ul style="list-style-type: none"> - les surfaces habitables - les années de construction - les nombres de pièces et d'étages - les codes NAF pour les locaux professionnels Remarque : pour les locaux commerciaux et industriels: les années de construction ne sont souvent pas disponibles et les surfaces ne sont également pas toujours renseignées.		
Date	Mise à jour annuelle		
Licence/ Conditions d'accès	L'utilisation et la diffusion des fichiers fonciers est réservée aux organismes habilités, ayant effectué une procédure de déclaration à la CNIL		

Nom:	Flores : Fichier localisé des rémunérations et de l'emploi salarié Clap : Connaissance locale de l'appareil productif (avant 2018) Source : INSEE		
Echelle géographique	Commune		
Entité concernée	Bâtiments	Système	Occupants
	✓		✓
Contenu/Attributs	Nombre d'établissements et effectifs salariés par secteur d'activité détaillé et tranche d'effectifs salariés détaillés.		
Date	2013, Mise à jour annuelle depuis 2003		
Licence/ Conditions d'accès	Ouvert		

Nom:	SIRENE® Source: INSEE		
Echelle géographique	Adresse		
	Bâtiments	Système	Occupants
			✓
Contenu/Attributs	Diffusée en accès libre depuis 2017 par l'INSEE, la base de données SIRENE® répertorie l'ensemble des établissements et entreprises actifs du pays, localisés à l'adresse. Environ 10 millions d'établissements sont représentés. De nombreuses données économiques et juridiques sont disponibles, et notamment : <ul style="list-style-type: none"> - les tranches d'effectifs - les codes NAF 		
Date	Mise à jour quotidienne		
Licence/ Conditions d'accès	Ouvert depuis 2017		

Annexe 2 : Méthodes de calcul de Cm selon (AFNOR, 2010)

1. Méthode détaillée du calcul de Cm

La méthode détaillée (AFNOR, 2010) se base sur l'équation de la chaleur appliquée à une paroi, pour laquelle les effets de bord sont négligés. La température dépend donc du temps t et la position horizontale dans la paroi x :

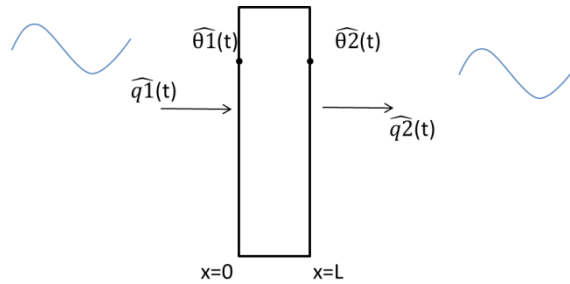


Figure 80 : Schémas d'une paroi monocouche

$$\frac{\partial T(x, t)}{\partial t} = \alpha \frac{\partial^2 T(x, t)}{\partial x^2} \quad (137)$$

$$\alpha = \frac{\lambda}{\rho c} \quad (138)$$

Afin de résoudre cette équation, la température est supposée en régime sinusoïdal selon une période T dite « période de variation thermique d'intérêt ». La période sélectionnée est de 24 h ce qui correspond « aux variations climatiques quotidiennes et aux réductions de température de consigne » :

$$\theta(x, t) = \bar{\theta} + \Delta\theta(x)\cos(\omega t + \psi) \quad (139)$$

$$\Delta\theta(x, t) = \theta(x, t) - \bar{\theta} = \Delta\theta(x)\cos(\omega t + \psi) \quad (140)$$

Il est ainsi possible de définir la valeur complexe associée à ce signal :

$$\widehat{\Delta\theta}(x, t) = \Delta\theta(x)e^{i(\omega t + \psi)} \quad (141)$$

L'équation de la chaleur appliquée à cette écriture complexe devient :

$$\frac{\partial \Delta\theta(x)e^{i(\omega t+\varphi)}}{\partial t} = \alpha \frac{\partial^2 \Delta\theta(x)e^{i(\omega t+\varphi)}}{\partial x^2} \quad (142)$$

$$i\omega \Delta\theta(x)e^{i(\omega t+\psi)} = \alpha e^{i(\omega t+\psi)} \frac{\partial^2 \Delta\theta(x)}{\partial x^2} \quad (143)$$

$$\Delta\theta(x) = \frac{\alpha}{i\omega} \frac{\partial^2 \Delta\theta(x)}{\partial x^2} = \frac{1}{\beta^2} \frac{\partial^2 \Delta\theta(x)}{\partial x^2} \quad \text{avec} \quad \beta = \sqrt{\frac{i\omega}{\alpha}} \quad (144)$$

Les solutions de cette équation différentielle sont de la forme :

$$\Delta\theta(x) = K1 \sinh(\beta x) + K2 \cosh(\beta x) \quad (145)$$

L'expression de la variation de flux est obtenue en appliquant la loi de Fourier :

$$\Delta\overline{\varphi(x)} = -\lambda\beta(K1 \cosh(\beta x) + K2 \sinh(\beta x))\overline{e x} \quad (146)$$

En utilisant ces résultats, les températures et les flux aux conditions limites de la paroi peuvent être liés par une matrice de transfert :

$$\widehat{\theta 1} = \Delta\theta(0) = K2 \quad (147)$$

$$\widehat{\theta 2} = \Delta\theta(L) = K1 \sinh(\beta L) + K2 \cosh(\beta L) \quad (148)$$

$$\overline{q 1} = \Delta\overline{\varphi(0)} = -\lambda\beta K1 \overline{e x} \quad (149)$$

$$\overline{q 2} = \Delta\overline{\varphi(L)} = -\lambda\beta(K1 \cosh(\beta L) + K2 \sinh(\beta L))\overline{e x} \quad (150)$$

$$\begin{pmatrix} \widehat{\theta 2} \\ \overline{q 2} \end{pmatrix} = \begin{pmatrix} T11 = \cosh(\beta L) & T12 = -\frac{1}{\lambda\beta} \sinh(\beta L) \\ T21 = -\lambda\beta \sinh(\beta L) & T22 = \cosh(\beta L) \end{pmatrix} \begin{pmatrix} \widehat{\theta 1} \\ \overline{q 1} \end{pmatrix} \quad (151)$$

Dans le cas d'une paroi multicouche, les matrices de transfert de chacune des parois sont multipliées entre elles pour obtenir une matrice de transfert globale. Des matrices de transfert additionnelles sont également à considérer pour prendre en compte les résistances surfaciques. Cette matrice de transfert peut ensuite être mise sous la forme d'une matrice d'impédance :

$$\begin{pmatrix} \widehat{q 1} \\ -\widehat{q 2} \end{pmatrix} = \begin{pmatrix} Y11 = \frac{-T11}{T12} & Y12 = \frac{1}{T12} \\ Y21 = \frac{1}{T12} & Y22 = -\frac{T22}{T12} \end{pmatrix} \begin{pmatrix} \widehat{\theta 1} \\ \widehat{\theta 2} \end{pmatrix} \quad (152)$$

Dans la norme, la capacité surfacique $k1$ de la paroi 1 est définie à partir des éléments de cette matrice d'impédance :

$$k1 = \frac{Cm}{A} = \frac{1}{\omega} |Y11 + Y12| \quad (153)$$

Le premier terme $Y11$ appelé admittance peut être interprété comme l'effet sur la surface 1 d'une excitation sur cette même face. Le second terme $Y12$, appelé

coefficient de transmission, peut être interprété comme l'effet sur la surface 1 d'une excitation sur la surface 2 :

$$Y_{11} = \left. \frac{\widehat{q1}}{\widehat{\theta1}} \right|_{\widehat{\theta2}=0} \quad (154)$$

$$Y_{12} = \left. \frac{\widehat{q1}}{\widehat{\theta2}} \right|_{\widehat{\theta1}=0} \quad (155)$$

Il est possible d'obtenir une expression analytique de la capacité surfacique :

$$k_1 = \frac{1}{\omega} \left| \frac{\cosh(\beta L)}{\frac{1}{\lambda\beta} \operatorname{sinsh}(\beta L)} - \frac{1}{\frac{1}{\lambda\beta} \operatorname{sinsh}(\beta L)} \right| = \frac{1}{\omega} \left| \frac{\cosh(\beta L) - 1}{\frac{1}{\lambda\beta} \operatorname{sinsh}(\beta L)} \right| \quad (156)$$

Remarque : dans la norme, par convention, le terme Y_{12} est défini de manière négative :

$Y_{12} = -\left. \frac{\widehat{q1}}{\widehat{\theta2}} \right|_{\widehat{\theta1}=0}$, ainsi la capacité surfacique est donnée sous la forme $k_1 = \frac{1}{\omega} |Y_{11} - Y_{12}|$.

On peut aussi interpréter cette expression de la capacité surfacique en considérant l'analogie avec la représentation en π d'un quadripôle (Figure 81):

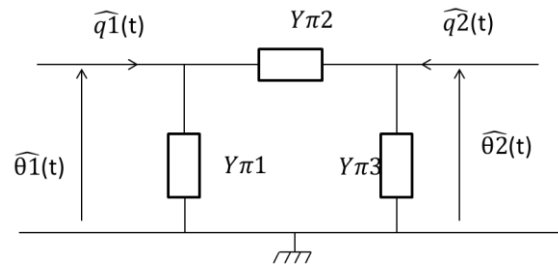


Figure 81 : Représentation en π d'un quadripôle

La matrice d'impédance du quadripôle dit en π peut s'écrire sous la forme :

$$\begin{pmatrix} Y_{11} & -Y_{12} \\ -Y_{21} & Y_{22} \end{pmatrix} = \begin{pmatrix} Y_{\pi1} + Y_{\pi2} & -Y_{\pi2} \\ -Y_{\pi2} & Y_{\pi2} + Y_{\pi3} \end{pmatrix} \quad (157)$$

En faisant l'analogie avec la matrice d'impédance présentée en équation (152), on obtient

$$Y_{\pi2} = Y_{12} = Y_{21} \quad (158)$$

$$Y_{\pi1} = Y_{11} - Y_{12} \quad (159)$$

$$Y_{\pi3} = Y_{22} - Y_{12} \quad (160)$$

$$k1 = \frac{1}{\omega} |Y\pi1| \quad (161)$$

Cette formulation correspond au calcul d'une valeur de capacité à partir d'une impédance purement capacitive (imaginaire).

Finalement, la méthode présentée ci-dessus implique d'importantes hypothèses, et le lien avec la capacité utilisée dans le modèle R5C1 reste peu évident. Les conditions d'application et les incertitudes associées ne sont pas renseignées.

2. Méthode simplifiée du calcul de C_m

La norme (AFNOR 2010), propose aussi une méthode simplifiée pour l'estimation de C_m . Une épaisseur limite (profondeur de pénétration) est définie à partir des caractéristiques de la paroi et la période de variation thermique. Si l'épaisseur des parois est inférieure à deux fois l'épaisseur limite, il est fait l'hypothèse de la couche mince, l'ensemble de la capacité du mur est alors utilisée, la capacité surfacique est calculée selon l'équation :

$$km = \rho Lc \quad (162)$$

Si l'épaisseur est supérieure à deux fois la profondeur de pénétration, la paroi est supposée semi-infinie. Le coefficient de transmission n'a alors plus d'effet sur la capacité, et l'admittance atteint une valeur limite. La capacité surfacique est déterminée selon l'équation suivante :

$$km = \frac{\delta \rho c}{\sqrt{2}} \quad \text{avec} \quad \delta = \sqrt{\frac{\lambda T}{\pi \rho c}} \quad (163)$$

Dans le cas intermédiaire, la méthode dite de l'épaisseur efficace est utilisée. Cette méthode consiste à sélectionner la plus petite valeur de capacité, entre des valeurs calculées selon trois méthodes distinctes non détaillées ici (se référer à la norme (AFNOR 2010) pour les détails).

Enfin les résistances superficielles Rs des parois, incluant les coefficients radiatif, convectif, et les revêtements à effets thermiques, doivent être considérées selon l'équation suivante :

$$km' = \sqrt{\frac{km^2}{1 + \omega^2 km^2 Rs^2}} \quad (164)$$

RÉSUMÉ

Les quartiers constituent aujourd'hui une échelle privilégiée des collectivités pour leur politique de transition énergétique. Cette thèse propose une méthode de construction de modèles thermiques de bâtiments applicable à cette maille. L'objectif est de pouvoir simuler des scénarios de performance énergétique (flexibilité, rénovation, sobriété) pour aider les prises de décisions, ainsi que de réaliser un suivi permettant d'identifier des dérives de consommation et d'évaluer des actions de performances énergétique mises en place. Dans une première phase, des modèles réduits RC, sont paramétrés à partir de bases de données territoriales, de documents de construction et de données statistiques d'enrichissement. Ces modèles, dis a priori, peuvent ensuite être calibrés par inférence bayésienne, à partir des données d'exploitation des bâtiments telles que les puissances de chauffage et les températures intérieures. Un modèle RC est construit sur la base de la norme ISO13790 et sélectionné pour ses capacités prédictives et l'identifiabilité de ses paramètres. Les possibilités et les limites des modèles a priori et a posteriori sont évaluées au regard des objectifs de la thèse, et en fonction des données de calibration disponibles.

MOTS CLÉS

Quartier, Performance énergétique, Modèles RC, Calibration bayésienne, Bases de données.

ABSTRACT

Districts are now a privileged scale for local authorities in their energy transition policy. This thesis proposes a method for constructing thermal models of buildings applicable to this scale. The objective is to simulate energy performance scenarios (flexibility, renovation, sobriety) in order to help decision making, as well as to perform an energy monitoring in order to identify consumption drifts and to evaluate the energy performance measures implemented. In a first phase, RC models are parameterized from territorial databases, construction documents and statistical data. Then, these models can be calibrated by Bayesian inference, with building operating data such as heating power and indoor temperatures. A RC model has been built on the basis of the ISO13790 standard and selected for its predictive capabilities and the identifiability of its parameters. The possibilities and limitations of the prior and posterior models are evaluated with respect to the objectives of the thesis, and according to the available calibration data.

KEYWORDS

District, Energy performance, RC models, Bayesian calibration, Databases.