



HAL
open science

Développements méthodologiques pour la validation des outils de simulation énergétique urbaine

Lucie Lefort

► **To cite this version:**

Lucie Lefort. Développements méthodologiques pour la validation des outils de simulation énergétique urbaine. Thermics [physics.class-ph]. Université Paris sciences et lettres, 2022. English. NNT : 2022UPSLM099 . tel-04335151

HAL Id: tel-04335151

<https://pastel.hal.science/tel-04335151>

Submitted on 11 Dec 2023

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 Mines Paris

**Methodological developments for the validation of Urban
Building Energy Models**

***Développements méthodologiques pour la validation
des outils de simulation énergétique urbaine***

Soutenue par

Lucie Lefort

Le 14 octobre 2022

École doctorale n°621

ISMME

**Ingénierie des Systèmes,
Matériaux, Mécanique, Éner-
gétique**

Spécialité

**Energétique et Génie des
procédés**

Composition du jury :

Michaël KUMMERT Professeur, Polytechnique Montréal	<i>Rapporteur</i>
Monika WOLOSZYN Professeure, Université Savoie Mont Blanc	<i>Rapporteuse</i>
Christian INARD Professeur, Université de La Rochelle	<i>Président</i>
Mindjid MAIZIA Professeur, Université de Tour	<i>Examineur</i>
Bruno DUPLESSIS Maître assistant, Mines Paris	<i>Examineur</i>
Thomas BERTHOU Ingénieur de recherche, Mines Paris	<i>Examineur</i>
Pascal SCHETELAT Ingénieur de recherche, CSTB	<i>Examineur</i>
Eric PEIRANO Directeur général adjoint en charge de la R&D, Efficacity	<i>Directeur de thèse</i>

Remerciements

Alors voilà, une page de ma vie se tourne aujourd’hui, un chapitre se clôt. Et quand je regarde le chemin parcouru, je me rends compte que ce qui me restera de ces trois ans de thèse, ce ne sont pas les graphiques sur lesquels j’ai passé des heures ou les milliers de lignes de code. Non ce qui me restera, ce sont vos visages à vous tous qui m’avez accompagnée pendant ces trois ans.

Alors merci à mes encadrants : Pascal, Bruno, Peter, Thomas, Romain et Eric. Merci Thomas pour ta positive attitude permanente et ta gentillesse. Merci Eric pour ton accompagnement tout en délicatesse. Merci Peter d’avoir aidé à garder les pieds sur terre à cette thèse et merci pour l’accueil à Sophia ! Merci Romain de m’avoir tirée dans cette aventure, tu avais raison, ça valait le coup ! Et merci de m’avoir fait confiance ! Merci Bruno d’avoir ramené l’odeur des embruns en V014 à travers nos discussions de voileux mais surtout merci d’avoir été ce pilier rassurant sur lequel je pouvais m’appuyer. Enfin merci Pascal pour ta passion communicative, merci pour les sessions vacances-travail à Grenoble et puis merci pour ces points thèses passés autant à se raconter nos vies et à refaire le monde qu’à regarder mes résultats !

Voilà le joyeux bazar de nos réunions à 7 puis 6 va beaucoup me manquer !

Une thèse c’est aussi un laboratoire et une équipe et je pense que je ne pouvais pas mieux tomber qu’avec l’équipe MDE et sa bande de doctorants alors merci pour les goûters, les cassettes impossibles, les soirées au Confessionnal et surtout merci pour les rires.

Merci tout particulier à Antoine et Rémi pour avoir été ma bouée de sauvetage pendant le COVID, à Enora pour avoir été ma grande soeur de thèse et à Valentin pour ton écoute bienveillante et attentive. Mais l’équipe MDE ne serait pas la même sans son chef d’équipe alors merci Pascal pour le soin que tu prends de chacun de nous. Bref MDE, je vous adore et vous allez me manquer mais je me rassure en me disant que les doctorants de l’équipe ne partent jamais complètement.

Dans les équipes qui m’ont accompagnée, il y a aussi celles du CSTB. Big Up tout particulièrement à l’équipe de la BDNB. Merci les gars de m’avoir permis de m’amuser avec un tel joujou, je resterai toujours votre première fan ! Mention spéciale à Nicolas, magicien des serveurs et de l’origami, qui a dû réparer mes bêtises un certain nombre de fois, merci Nico!

J’ai aussi eu la chance de faire partie de l’Efficaciteam et je voudrais remercier en particulier Romain, Arni, Nico, Sophie et Sam pour tous les bons moments que nous avons pu avoir. Merci Sam d’avoir été un n+1 hors norme et d’être un ami sur lequel on peut tout autant compter.

En mer la nuit, chaque bouée, chaque balise, chaque phare a une lumière bien spécifique pour guider les marins et pendant ces trois ans de vie, j’ai eu beaucoup de petites lumières un peu inclassables qui m’ont accompagnées alors merci Lucien, Angèle, Gwendo, la grande Marie, merci Vie en Mer, le MCC, Penboc’h et j’en oublie sûrement !

J'ai aussi la chance d'avoir bien ancré dans ma vie trois balises qui sont des repères fidèles : Elé, Nono, Zaza. Les filles, vous êtes mon filet de sécurité et peu importe ce qui arrive dans ma vie, tant que vous en ferez partie, je sais qu'il y aura des rires, des cocktails et des paillettes, tant que vous en ferez partie, je sais qu'il y aura de la joie.

Enfin un immense merci à mes phares : Papa, Maman, Jean, Loïc et Bertille. Merci de m'avoir aidée à devenir l'adulte que je suis tout en me laissant libre. Merci de m'aimer même lorsque je suis insupportable, même lorsque je me plante, merci de m'aimer telle que je suis tout entière. Je vous aime très fort.

Voilà, j'arrive au bout de cette thèse et quand je regarde en arrière, je me rends compte que vous tous, chacun à votre manière, vous avez pris soin de moi et Dieu sait qu'au cours de ces trois ans, j'en ai eu besoin. Alors du fond du coeur, merci à tous !

De mémoire de marin, on n'a jamais vu une tempête ne pas finir.

Abstract

Following the last IPCC report, the urge to act against climate change has become more pregnant than ever. In order to reduce the impact of human activities on climate, cities stand at a crossroads. With 55% of the world population living in urban areas [1], cities are responsible for a majority of the world energy consumption but represent also an important political power to shift toward an ecological transition. Thus, cities such as London [2] or New-York [3] have set ambitious environmental goals, both aiming to be carbon-neutral by 2050. To achieve such targets, reducing energy consumption in buildings is an important leverage. For instance, in France, buildings accounted for 45% of the national final energy consumption in 2015 [4].

In order to help decision makers choose the best energy saving measures for buildings at the urban scale, the scientific community developed decision-aiding tools among which lie the urban building energy models (UBEM) [5]. An UBEM is able to compute energy consumption for each building of a district or a city and thus enables decision-makers to have a clear understanding of the impacts of different energy saving measures such as buildings refurbishment or flexibility.

However, to ensure that UBEM provide a real aid in the decision-making process, their accuracy should be verified through a thorough validation process. Even though great efforts have been made to elaborate validation procedures for building energy models [6, 7], further work is required to take into account UBEM particularities (models and enrichment processes) into the validation process. UBEM should also be assessed in their general conditions of use, and not only on synthetic districts as proposed by Saelens et al. [8], to verify how they perform in real life decision-making process. Examples of comparison between measurements and UBEM results can be found in the scientific literature [9, 10], but such examples often lack reproducibility, the validation dataset being rarely available. Indicators used to assess UBEM performances can also vary greatly from a study to another, thus preventing inter-model comparisons. A standardization of the UBEM assessment process would then help compare different UBEMs and surely lead to improvements in urban building energy modelling fields.

The main research question tackled in this thesis can be described as follows: **How to verify the adequacy of UBEM to the decision process they are used for?**

This main research question leads to several other questions that will be discussed in this thesis dissertation:

- What are the decision processes UBEM are involved in and which UBEM outputs should be assessed in each use case ?
- How to develop a comprehensive and generic validation framework for UBEM ?

A validation framework composed of an assessment stage and a diagnostic stage has been developed and is presented in this dissertation. An application of this assessment framework is proposed on use cases for which open data can be used to compose a validation dataset. The assessment stage and the diagnostic stage are then described in depth for this set of use cases along with the statistical tools designed to assess and analyse the UBEM performance.

The proposed assessment framework has been designed to be applicable to a large range of UDEM and to a large range of use cases, but future works should verify its applicability to a broader range.

Résumé

Suite au dernier rapport du GIEC, le besoin d’agir contre le changement climatique est devenu plus prégnant que jamais. Afin de réduire l’impact des activités humaines sur le climat, les villes sont des acteurs incontournables. Avec 55% de la population mondiale vivant dans des zones urbaines [1], les villes sont responsables de la majorité des consommations énergétiques mondiales mais représentent également un pouvoir politique important pour s’orienter vers une transition écologique. Ainsi, des villes comme Londres [2] ou New-York [3] ont fixé des objectifs environnementaux ambitieux, visant toutes deux à être neutres en carbone d’ici 2050. Pour atteindre ces objectifs, la réduction de la consommation d’énergie dans les bâtiments est un levier important. Par exemple, en France, les bâtiments représentaient 45% des consommations nationales d’énergie finale en 2015 [4].

Afin d’aider les décideurs à choisir les meilleures mesures d’économie d’énergie pour les bâtiments à l’échelle urbaine, la communauté scientifique a développé des outils d’aide à la décision parmi lesquels figurent les outils de simulation énergétique urbaine (Urban Building Energy Model, UBEM) [5]. Un UBEM est capable de calculer les consommations d’énergie pour chaque bâtiment d’un quartier ou d’une ville et permet ainsi aux décideurs d’avoir une compréhension claire des impacts de différentes mesures d’économie d’énergie telles que la rénovation de certains bâtiments ou la mise en place d’actions d’effacement des consommations.

Cependant, pour s’assurer que les UBEM fournissent une aide réelle au processus de prise de décision, leur exactitude doit être vérifiée par un processus de validation approfondi. Même si de nombreux travaux de recherche ont permis d’élaborer des procédures de validation pour les modèles de simulation énergétique des bâtiments [6, 7], des travaux supplémentaires sont nécessaires pour prendre en compte les particularités des UBEM (modèles et processus d’enrichissement) dans le processus de validation. Les UBEM doivent également être évalués dans leurs conditions générales d’utilisation, et pas seulement sur des quartiers synthétiques, comme le proposent Saelens et al. [8], ceci afin de vérifier leur performance dans le processus de décision réel. Des exemples de comparaison entre les mesures et les résultats d’UBEM peuvent être trouvés dans la littérature scientifique [9, 10], mais ces exemples manquent souvent de reproductibilité, le jeu de données de validation étant rarement disponible. Les indicateurs utilisés pour évaluer les performances des UBEM peuvent également varier fortement d’une étude à l’autre, ce qui empêche les comparaisons inter-modèles. Une standardisation du processus de validation des UBEMs aiderait alors à comparer les différents UBEMs et conduirait sûrement à des améliorations dans les domaines de la modélisation énergétique des bâtiments à l’échelle urbaine.

La principale question de recherche abordée dans cette thèse peut donc être décrite comme suit : **Comment vérifier l’adéquation des UBEM au processus de décision pour lequel ils sont utilisés ?**

Cette question de recherche principale mène à plusieurs autres questions qui seront discutées dans ce manuscrit :

- Quels sont les processus de décision dans lesquels les outils de simulation énergétique ur-

baine sont impliqués et quelle sortie de ces outils doit être évaluée dans chaque cas d'usage ?

- Comment développer un cadre de validation complet et générique pour les outils de simulation énergétique urbaine ?

Un processus de validation composé d'une étape d'évaluation et d'une étape de diagnostic a été développé et est présenté dans cette thèse. Une application de ce processus de validation est proposée sur des cas d'usages pour lesquels des données accessibles librement peuvent être utilisées pour composer un jeu de données de validation. L'étape d'évaluation et l'étape de diagnostic sont ensuite décrites en profondeur pour cet ensemble de cas d'usages de même que les outils statistiques sélectionnés pour évaluer et analyser les performances des UBEM.

Le cadre d'évaluation proposé a été pensé pour être applicable à un large éventail d'UBEM et à un large éventail de cas d'usages, mais des travaux futurs devraient vérifier son applicabilité à un éventail plus large.

Contents

Remerciements	i
Abstract	viii
Résumé	viii
Contents	ix
List of Figures	xi
List of Tables	xvii
Acronyms	xix
1 Urban Building Energy Models as decision-support tools: overview and thesis approach	1
1.1 Urban Building Energy Models: specificities and challenges	2
1.2 Urban Building Energy Models: main features and simulation processes	3
1.3 Urban Building Energy Models: modelling purposes and decision-making processes	9
1.4 Thesis approach and structure of the manuscript	14
2 Urban Building Energy Models and Validation	17
2.1 Validation: a state of the art	18
2.2 A need for a new validation framework	28
3 Application of the validation framework to a first set of use cases	35
3.1 Ad hoc validation of both the physical model and the enrichment process	36
3.2 Choice of the use cases	51
3.3 Creation of a validation dataset	53
4 Evaluation: Assessing the fitness for purpose of UBEM	59
4.1 Quantifying the existing uncertainty	60
4.2 Applying the validation procedure to a simple UBEM	68

4.3	Applying the validation procedure to other UBEM tools	84
5	Diagnosis: Understanding the weaknesses of UBEM	93
5.1	Presentation of the diagnostic stage applied to Simply	95
5.2	Applying the diagnostic stage to other tools	105
5.3	Studying the influence of different levels of detail on the tools performances . . .	111
6	Conclusion and outlook	145
	Bibliography	151
	Bibliography	151
	Résumé long en français	159
	Appendices	173

List of Figures

- 1.1 UBEM perimeter 3
- 1.2 Existing approaches for the parameters’ allocation step during the enrichment process 8

- 2.1 Validation method developed in the PASSYS project [6] 19
- 2.2 Validation framework for UBEM: a proposal 34

- 3.1 Object structure in DIMOSIM [11]. 37
- 3.2 RiCj thermal zone model used in DIMOSIM [11]. 39
- 3.3 Results obtained by DIMOSIM on the BESTEST benchmark [11]. 40
- 3.4 Results obtained by several UBEM on the DESTEST [8]. 40
- 3.5 Example of a Bayesian network. 42
- 3.6 Comparison between the normalised mutual information and the Pearson coefficient. 46
- 3.7 Bayesian network obtained by training on EPCs. 47
- 3.8 Mutual Information loss. 49
- 3.9 Marginals distributions of the training set, the trained network and of a sample drawn from the network. 50
- 3.10 Annual measured and simulated consumptions of gas and electricity obtained for 100 districts. 51
- 3.11 Correlation coefficients between the error and several input parameters. 52
- 3.12 Example of buildings addressing issues. 55
- 3.13 Sankey diagram of the buildings available to create a validation dataset. 56
- 3.14 Dashboard of presentation of the validation dataset. 58

- 4.1 Inverse transform method applied to continuous and discrete random variables (reproduced from [12]) 62
- 4.2 Locating uncertainty along the simulation chain 64
- 4.3 Test building randomly sampled from the validation dataset - Picture taken from Google Maps 72
- 4.4 Distributions of the annual gas and electricity consumptions obtained by Simply on the test building 73

LIST OF FIGURES

4.5	Variation of the annual gas consumption with respect to the heated ratio and to the air-change rate	73
4.6	Distributions of the percentage of error obtained by Simply on the annual gas and electricity consumptions of the test building	74
4.7	Distributions of the annual electricity and gas consumptions obtained at the sample scale	75
4.8	Distributions of the residual on the annual electricity and gas consumptions obtained at the sample scale	76
4.9	Difference between central interval (a) and HPD region (b) for a multimodal pdf - figure taken from Murphy [13]	77
4.10	HPD region computed for a confidence level of 95% on the annual gas and electricity consumptions of the test building	77
4.11	Distribution of the computed consumptions with a level of confidence of 95% in regard with the DLE	78
4.12	μ_{bldg} and the HPD region of each building against its DLE	79
4.13	Evolution of the empirical coverage for different levels of confidence	80
4.14	Evolution of the empirical coverage obtained by Simply for different number of samples	81
4.15	Distributions of the annual gas and electricity consumptions obtained by Simply, Calomel and Dimosim on the test building	86
4.16	Distributions of the percentage of error obtained by Simply, Calomel and Dimosim on the annual gas and electricity consumptions of the test building	87
4.17	Distributions of the residuals obtained on the validation dataset by Simply, Calomel and Dimosim	88
4.18	Distributions of the median consumptions and HPD region obtained by Simply, Calomel and Dimosim with a 95% confidence level	90
4.19	Evolution of the empirical coverage for different confidence levels for Simply, Calomel and Dimosim	91
5.1	Distributions of the percentage of error obtained by Simply on the annual gas and electricity consumptions for the test building	95
5.2	Mutual information computed between the residual, μ_{bldg} , and buildings characteristics	96
5.3	Relationship between uncertain parameters and the residual	97
5.4	Distributions of the percentage of error obtained by Simply on the annual gas and electricity consumptions of the test building	99
5.5	Mutual information between the percentage of error and uncertain parameters	99
5.6	Mutual information between the percentage of error on the electricity and gas consumptions and the uncertain parameters for all buildings of the validation dataset	101
5.7	Share of variance explained by the first five eigenvectors obtained via PCA	102

5.8	Projection of the mutual information table on the two-dimension space obtained via PCA - coloured based on the heating energy carrier	102
5.9	Eigenvectors obtained via PCA	103
5.10	Projection of the mutual information table on the two-dimension space obtained via PCA - coloured based on the DHW energy carrier	104
5.11	Mutual information computed between the residual and building's characteristics	106
5.12	Vectors of mutual information between uncertain parameters and percentage of error obtained on electricity consumptions for the buildings of the validation dataset and for three UBEM	108
5.13	Vectors of mutual information between uncertain parameters and percentage of error obtained on gas consumptions for the buildings of the validation dataset and for three UBEM	109
5.14	Results of principal components analysis applied to vectors of mutual information for the error on electricity consumptions for Simply, Calomel and Dimosim	112
5.15	Results of principal components analysis applied to vectors of mutual information for the error on gas consumptions for Simply, Calomel and Dimosim	113
5.16	Evolution of the mean and standard deviation of the consumptions obtained for each building between the three levels of data availability	115
5.17	Evolution of the gas and electricity consumptions obtained for a building heated with electricity	116
5.18	Evolution of the gas and electricity consumptions obtained for a building heated with gas	116
5.19	Comparison between the validation dataset and the training set used to train the Bayesian network	117
5.20	Distributions of residuals obtained with Simply for the three levels of data availability on electricity	118
5.21	Distributions of residuals obtained with Simply for the three levels of data availability on gas	119
5.22	μ_{bldg} and the HPD region of each building against its DLE for three levels of data availability	122
5.23	Evolution of the empirical coverage depending on the data availability	123
5.24	Mutual information between the residual and building's characteristics for three levels of data availability	124
5.25	Vectors of mutual information between uncertain parameters and percentage of error on electricity consumption obtained for the buildings of the validation dataset and for three levels of data availability	126
5.26	Vectors of mutual information between uncertain parameters and percentage of error on gas consumption obtained for the buildings of the validation dataset and for three levels of data availability	127
5.27	Results of principal components analysis applied to vectors of mutual information for the error on electricity consumption for three levels of data availability	128
5.28	Results of principal components analysis applied to vectors of mutual information for the error on gas consumption for three levels of data availability	129

LIST OF FIGURES

5.29	μ_{bldg} and the HPD region of each building obtained with Calomel against its DLE for three levels of data availability	131
5.30	μ_{bldg} and the HPD region of each building obtained with Dimosim against its DLE for three levels of data availability	132
5.31	Mutual information between the residual obtained with Calomel and buildings characteristics for three levels of data availability	133
5.32	Mutual information between the residual obtained with Dimosim and buildings characteristics for three levels of data availability	134
5.33	Vectors of mutual information between uncertain parameters and percentage of error on electricity consumptions obtained for the buildings of the validation dataset with Calomel and for three levels of data availability	136
5.34	Vectors of mutual information between uncertain parameters and percentage of error on gas consumptions obtained for the buildings of the validation dataset with Calomel and for three levels of data availability	137
5.35	Vectors of mutual information between uncertain parameters and percentage of error on electricity consumptions obtained for the buildings of the validation dataset with Dimosim and for three levels of data availability	138
5.36	Vectors of mutual information between uncertain parameters and percentage of error on gas consumptions obtained for the buildings of the validation dataset with Dimosim and for three levels of data availability	139
5.37	Results of principal components analysis applied to vectors of mutual information obtained for error on electricity consumptions with Calomel for three levels of data availability	140
5.38	Results of principal components analysis applied to vectors of mutual information obtained for error on gas consumptions with Calomel for three levels of data availability	141
5.39	Results of principal components analysis applied to vectors of mutual information obtained for error on electricity consumptions with Dimosim for three levels of data availability	142
5.40	Results of principal components analysis applied to vectors of mutual information obtained for error on gas consumptions with Dimosim for three levels of data availability	143
6.1	Validation framework for UBEM: a proposal	149
1	Principales étapes de simulation pour les outils de simulation énergétique urbaine	160
2	Cas d’usages associés aux UBEM	161
3	Proposition d’un cadre méthodologique pour la validation des outils de simulation énergétique urbaine	164
4	Présentation du jeu de données de validation retenu pour la suite du manuscrit .	167
1	Bayesian network obtained by training on PHEBUS.	175
2	Mutual Information loss obtained for the appliances network.	176
3	Marginals distributions of the training set, the trained network and of a sample drawn from the appliances network.	177

4	Bayesian network obtained by training on the census.	178
5	Mutual Information loss obtained for the census network.	179
6	Marginals distributions of the training set, the trained network and of a sample drawn from the census network.	180
7	Evolution of the mean and standard deviation of the consumptions obtained with Calomel for each building between the three levels of data availability	191
8	Distributions of residuals obtained with Calomel for the three levels of data availability on electricity	192
9	Distributions of residuals obtained with Calomel for the three levels of data availability on gas	193
10	μ_{bldg} and the HPD region of each building against its DLE for three levels of data availability for Calomel	194
11	Evolution of the empirical coverage depending on the data availability for Calomel	195
12	Mutual information between the residual and building's characteristics for three levels of data availability for Calomel	196
13	Vectors of mutual information between uncertain parameters and percentage of error on electricity consumption obtained for the buildings of the validation dataset and for three levels of data availability with Calomel	198
14	Vectors of mutual information between uncertain parameters and percentage of error on gas consumption obtained for the buildings of the validation dataset and for three levels of data availability with Calomel	199
15	Results of principal components analysis applied to vectors of mutual information for the error on electricity consumption for three levels of data availability obtained for Calomel	200
16	Results of principal components analysis applied to vectors of mutual information for the error on gas consumption for three levels of data availability obtained for Calomel	201
17	Evolution of the mean and standard deviation of the consumptions obtained with Dimosim for each building between the three levels of data availability	202
18	Distributions of residuals obtained with Dimosim for the three levels of data availability on electricity	203
19	Distributions of residuals obtained with Dimosim for the three levels of data availability on gas	204
20	μ_{bldg} and the HPD region of each building against its DLE for three levels of data availability for Dimosim	205
21	Evolution of the empirical coverage depending on the data availability for Dimosim	206
22	Mutual information between the residual and building's characteristics for three levels of data availability for Dimosim	207
23	Vectors of mutual information between uncertain parameters and percentage of error on electricity consumption obtained for the buildings of the validation dataset and for three levels of data availability with Dimosim	208

LIST OF FIGURES

24 Vectors of mutual information between uncertain parameters and percentage of error on gas consumption obtained for the buildings of the validation dataset and for three levels of data availability with Dimosim 209

25 Results of principal components analysis applied to vectors of mutual information for the error on electricity consumption for three levels of data availability obtained for Dimosim 210

26 Results of principal components analysis applied to vectors of mutual information for the error on gas consumption for three levels of data availability obtained for Dimosim 211

List of Tables

- 1.1 Common inputs for the physical model of an UBEM tool. 6
- 1.2 Data sources available for UBEM parameterisation. 7
- 1.3 Review of UBEM use cases 13

- 2.1 Review of validation attempts 25

- 3.1 Single feature performances assessment. 48

- 4.1 Heating system’s efficiency 67
- 4.2 DHW system’s efficiency 67
- 4.3 Hypotheses retained for the uncertainty quantification 69
- 4.4 Simply input parameters 71
- 4.5 Main properties of the test building 72
- 4.6 Comparison of the input parameters required by both Simply and Calomel 85
- 4.7 Results obtained with the three tools on the test building 85
- 4.8 Results obtained with the three tools at the sample level 91

- 5.1 Results obtained thanks to the analysis on the error variance 110
- 5.2 Known and uncertain parameters for the three levels of data availability 114
- 5.3 Mutual information between measurements and simulated consumptions for the three levels of data availability 120

- 1 Single feature performances assessment for the appliances network. 176
- 2 Single feature performances assessment for the census network. 178

Acronyms

AE Absolute Error.

APE Absolute Percentage of Error.

BDNB Base de Données Nationale des Bâtiments.

BEM Building Energy Model.

cdf Cumulative Distribution Function.

CVRMSE Coefficient of Variation of the Root Mean Square Error.

DHW Domestic Hot Water.

DLE Données Locales de l'Énergie.

EPC Energy Performance Certificates.

ESD Empirical Standard Deviation.

FDL Fichiers Détail Logement.

HPD Highest Posterior Density.

KPI Key Performance Indicator.

MAE Maximum Absolute Error.

MAPE Maximum Absolute Percentage of Error.

MBE Mean Bias Error.

ME Mean Error.

MI Mutual Information.

MPE Mean Percentage of Error.

NMBE Normalized Mean Bias Error.

NVF Normalized Variation Factor.

PCA Principal Components Analysis.

pdf Probability Density Function.

PE Percentage of Error.

R² Coefficient of determination.

RE Relative Error.

RIG Relative Information Gain.

RMSE Root Mean Square Error.

RNRMSE Range Normalized Root Mean Square Error.

RSS Residual Sum of Squares.

SMAPE Symmetrical Mean Absolute Percentage of Error.

UA Uncertainty Analysis.

UBEM Urban Building Energy Model.

Chapter 1

Urban Building Energy Models as decision-support tools: overview and thesis approach

Contents

1.1	Urban Building Energy Models: specificities and challenges	2
1.2	Urban Building Energy Models: main features and simulation processes	3
1.2.1	Perimeter of a UBEM tool	3
1.2.2	Physical models in UBEM	3
1.2.3	Data available at the building scale	5
1.2.4	Enrichment processes in UBEM	8
1.3	Urban Building Energy Models: modelling purposes and decision-making processes	9
1.4	Thesis approach and structure of the manuscript	14

1.1 Urban Building Energy Models: specificities and challenges

In the sixth IPCC assessment report [14], the decade to come has been identified as critical in the fight against climate change. A sharp decrease in global energy consumption and greenhouse gas emissions is required to prevent more environmental crises.

Cities are responsible for 75% of global primary energy consumption [15] and this share is likely to increase with the raise of the world population living in urban areas [1]. Cities are then key players of the ecological transition due to their high energy consumption, but also due to their high energy-saving and adaptability potentials. For instance, it would be possible in many cities to develop district heating and cooling networks using renewable energy to cover the heating and cooling needs of offices and dwellings. District cooling networks would namely offer a viable solution to the urban heat island effect that many cities experience during heat waves and would help protect the elderly and fragile people from intense heat.

Another reason for cities being a cornerstone of the ecological transition is their political power. Cities are able to implement environmentally friendly policies in a number of different fields such as waste management, biodiversity preservation, water management or energy consumption in buildings. London [2], New-York [3] and Paris [16] have then set ambitious plans to drastically reduce their energy and greenhouse gas emissions by 2050. In order to achieve such goals, reducing energy consumption in buildings seems to be a keystone. In France, for example, tertiary and residential sectors represented 45% of the national final energy consumption in 2015 [4].

To help cities reduce energy consumption in buildings, decision-aiding tools have been developed by the scientific community among which are the **Urban Building Energy Models (UBEM)**. Urban Building Energy Models (UBEM) have been under development for more than fifteen years in order to address the question of energy transition in buildings at large spatial scale [5]. They are **decision-support tools that model the energy behaviour of the buildings of a district** in order to help decision makers arbitrate between different energy scenarios. A large diversity of models is used to compute the energy consumption of buildings in UBEM tools. It ranges from very detailed thermal models such as Energy Plus [17] to data-driven models [18]. The type of decision-processes that UBEM are intended to assist are also very diverse such as assessing the energy savings due to buildings retrofitting or evaluating the indoor comfort during a heat wave.

Independently of the decision-process or the physical model used, **the data collection to parameterize a UBEM is often a challenging issue**. For example, the heated volume of each building is a typical input of UBEM. Although having this piece of information for a single building may seem trivial, it is a much more complex question to address when it comes to collect the information for hundreds of buildings at once. Most of UBEM inputs share this same issue. For instance, it is extremely rare to know precisely the insulation level, the window type or the roof slope of each building in a district. In order to overcome this lack of information, researchers have developed strategies to fill in missing data with technical surveys and sets of hypotheses. This practice is called **data enrichment process**. This data enrichment process, while necessary, may impair UBEM results and jeopardize their usefulness. Indeed, this process brings an error that is not yet well quantified.

Although the validation of building energy models taken individually has already been discussed extensively in several projects such as the PASSYS project [6] or the BESTEST [7], the UBEM have not yet benefitted from a similar validation framework. This means that there is no common framework to ensure that a UBEM provides a real aid to the decision-making process. The aim of this thesis is then to develop a validation framework dedicated to UBEM in order to verify their reliability. This validation methodology needs to take into account the enrichment process and assess UBEM performances on real buildings to verify the fitness for use of these tools. The

main research question addressed in this manuscript is then: **How to verify the adequacy of UBEM to the decision process they are used for?**

To tackle this research question, it is first necessary to define clearly the perimeter of UBEM by specifying their features and their modelling purpose. To this end, a comprehensive overview of UBEM is provided below, as well as a review of the decision-making processes in which UBEM are involved.

1.2 Urban Building Energy Models: main features and simulation processes

1.2.1 Perimeter of a UBEM tool

Urban Building Energy Models can be defined as decision-support tools that are capable to model energy consumption at the building scale for several buildings at once (the number of buildings being simulated can range from ten buildings to the entire national building stock). The simulation process found in UBEM is summarized Figure 1.1. It is composed of three main steps:

1. Collection of available data: the available data describing each building to model is collected.
2. Enrichment process: the missing input data is filled thanks to statistical surveys and sets of hypotheses.
3. Physical modelling: the energy behaviour of each building is simulated thanks to a physical model.

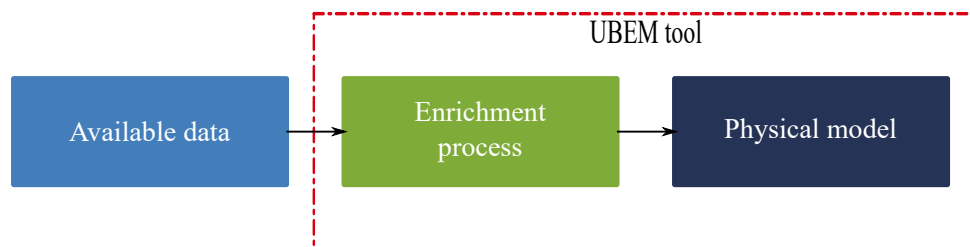


Figure 1.1: UBEM perimeter

Since the data enrichment process is inherent to UBEM and given that it could deeply alter the results obtained by a tool, a UBEM is defined in this manuscript as a set of a given physical model coupled with a specific enrichment process, as illustrated Figure 1.1.

Each step of the simulation process is now discussed more in depth in the following sections.

1.2.2 Physical models in UBEM

To simulate the energy behaviour of each building, the core of any UBEM tool is composed of a physical model. This physical model is often an assemblage of several sub-models modelling different aspects of the building's energy behaviour. These sub-models can be sorted into four categories: occupancy, systems, weather and envelop. The way to model the phenomena related to each of these categories can be very different from one UBEM to another. A review of the main approaches found in each category is then provided hereafter.

1.2.2.1 Occupancy modelling

The modelling of occupants in UBEM aims to take into account diverse phenomena such as blind and window opening, a change in the heating set point or the use of electric appliances. A comprehensive review of occupancy models found in UBEM has been provided by Happle et al. [19]. The authors break down the type of occupancy models into three categories:

- Deterministic and space-based models;
- Stochastic and space-based models;
- Stochastic and person-based models.

Person-based approaches allocate an occupancy scenario to each occupant of the building [20] while space-based approaches allocate an occupancy scenario to a whole area of the building based on its usage. Person-based approaches are interesting to couple UBEM with other tools such as agent-based transport models. Such coupling between a UBEM tool and a transport model was realised by Robinson et al. [21] using the UBEM tool CitySIM. Stochastic approaches are generally favoured over deterministic approaches when the diversity's representation is of great importance. For instance, when it comes to designing a district heating network, a good representation of the diversity in the simulated load profile is a cornerstone and deterministic approaches may not suffice.

1.2.2.2 Energy systems modelling

At the district scale, not only heating, ventilation and air conditioning (HVAC) systems inside buildings need to be modelled, but also district networks such as district heating networks or power grids, as described by Allegrini et al. [22]. Allegrini et al. also highlight the key role of cities for renewable energies deployment. UBEM should then be able to identify renewable energies potentials and pinpoint the match of these resources with the local demand. Such models have been implemented by Fonseca et al. [23] in a UBEM, City Energy Analyst (CEA).

1.2.2.3 Weather and climate modelling

Frayssinet et al. [24] describe the different types of urban climate models that may be used with UBEM. At the scale of the city, urban canopy models take into account a simple geometry of the buildings and calculate the air temperature and wind speed within the city. A more detailed approach is proposed using computational fluid dynamics, but this approach is often considered too expensive in computing resources. The simplest approach but also the most widespread is to use measured weather data from a weather station near the city and to interpolate them. Frayssinet et al. advocate for the use of more detailed models. This seems particularly reasonable if UBEM are to be used to study the urban heat island effect.

1.2.2.4 Thermal modelling

A common starting point for UBEM development has often been to use building energy model (BEM) to simulate more than one building and then add complementary models. However, BEM are detailed models and as such, are computationally expensive. Other approaches have then been applied to lower the computation time of models such as resistance-capacitance models or data-driven models.

Detailed models are a natural choice to build UBEM upon BEM. This strategy was implemented by Reinhart et al. [25] in UMI, by Hong et al. [26] in CityBES or by Polly et al. [27] in UrbanOpt. These three UBEM rely on EnergyPlus [17], a BEM used worldwide and that has a massive users community. Existing BEM have already benefitted from users' feedbacks and, for most of them, a thorough validation process has been carried out. However, BEM require a substantial amount of inputs and although it is possible to parallelize the computation process (each building may be calculated using a different processor), the computation cost remains high as stated by Hong et al. [28].

To tackle these two issues (computation cost and high number of inputs), resistance-capacitance models have been developed for UBEM. This approach was chosen by Garreau et al. [11] in Dimosim, by Berthou et al. [29] in Smart-E, by Fonseca et al. [23] in CEA and by Robinson et al. [21] in CitySim. These four UBEM tools use a resistance-capacitance model parameterized based on wall layers' descriptions. These models present the advantage of lowering computation cost and allow the computation of interactions among buildings like solar shading. However, if the number of layers used to describe the thermal boundaries of buildings is too small, such models can be inaccurate when looking at short time-scale processes.

A third approach quite different can also be found in the scientific literature: using black box models. In this case, the data-driven model generally does not require models from the other categories (weather, system and occupancy) nor an enrichment process. For this reason, **this type of models is not considered in the rest of this manuscript** even though the validation framework should be applicable to this type of UBEM as well. Future work should verify that the proposed validation framework is suitable to such models. An example of this approach is given by Kontokosta and Tull [18]: a predictive model of buildings energy consumption in New-York is implemented based on measured annual energy consumption. Three different algorithms are used and the performances of the algorithms are compared. One of the drawbacks of such approach is the inability of the model to be used outside of its training scope. In Kontokosta and Tull [18], the model is trained on energy consumption of buildings larger than 50,000 square feet. It then struggles to accurately predict energy consumption for smaller buildings. Furthermore, such models require measured energy consumption data at fine spatial and time scales and such data are often difficult to obtain due to privacy.

Putting aside the data-driven models, the common inputs for the physical model in a UBEM tool are detailed in Table 1.1.

1.2.3 Data available at the building scale

Independently of the chosen models, finding all this information to parameterize the physical model is challenging. When parameterizing an UBEM tool, information is required for every building that is to be simulated, such as the heated volume, the walls layer description or the number of occupants. Since this information is rarely fully available, more affordable data are collected and required information is derived from this partial dataset thanks to the enrichment process. The data sources often found in an UBEM case study are presented in the first columns of Table 1.2 along with the information they provide, and the required parameters usually derived from it. They are generally used by UBEM modellers, and provide information at the building scale for almost every building that one aims to simulate.

The information provided at building scale provided by each source is detailed in the second column. It appears quite often that the provided information is not an input directly required by the physical model of the UBEM, but it is often used in the enrichment process to fill in missing parameters. For instance, the tax assessment records provide the construction year of each building. The construction year is not an input required by the physical model of the

Thermal properties	Wall U-value
	Roof U-value
	Floor U-value
	Windows U-value
	Thermal inertia
	Windows-to-wall ratio
Systems	Heating system
	Cooling system
	Domestic Hot Water (DHW) system
	Heating and cooling systems efficiencies
	DHW system efficiency
Occupancy	Ventilation system
	Heating set point
	Cooling set point
	Heated volume
	Cooled volume
	DHW draw off
	Air-change rate
	Number of occupants
Number, type and energy class of appliances	
Outdoor conditions	Outdoor temperature
	Humidity
	Wind speed
	Solar radiation
	Solar shading

Table 1.1: Common inputs for the physical model of an UBEM tool.

UBEM, but thanks to the different thermal regulations that have evolved over time, it enables the modeller to choose a U-value for the walls of the buildings that reflects the construction type of each building.

Some pieces of information in the table can be retrieved from different sources, such as the wall U-value or the windows-to-wall ratio which are provided by the Energy Performance Certificates in some cases and which are derived from the construction year in other cases. This is due to the fact that Energy Performance Certificates are not available for every building (in France for instance only ten million out of 36 million dwellings display an EPC [35]). When this data source is unavailable, these inputs are then retrieved from another source during the enrichment process.

In that respect, the enrichment process is somehow the first modelling step of any UBEM as it makes the transition between the available information and the required parameters.

Data source	Available information	Parameters required by the physical model
Tax assessment record [30]	Construction year and/or refurbishment year	Wall U-value Roof U-value Floor U-value Window U-value Windows-to-wall ratio Heating system DHW system
	Total floor area Building location	Heated volume Building location
Geographic Information System database [31–33]	Building volume Solar shading Adjacent walls Building location	Heated volume Solar shading Share of adiabatic walls Building location
	Wall U-value Roof U-value Floor U-value Window U-value Heating system DHW system Windows-to-wall ratio	Wall U-value Roof U-value Floor U-value Window U-value Heating system DHW system Windows-to-wall ratio
Energy Performance Certificates (EPC) [34]	Outdoor temperature	Outdoor temperature
Closest weather station [30]	Humidity Solar radiation Wind speed	Humidity Solar radiation Wind speed
	Incomplete sources of information	Number of occupants Number, type and performance class of appliances Heating and cooling systems efficiencies DHW system efficiency Cooling system Heating/cooling set point Air-change rate

Table 1.2: Data sources available for UBEM parameterisation.

1.2.4 Enrichment processes in UBEM

The enrichment process consists of filling in the missing input parameters required by the physical model based on available information. Various approaches stand out from the UBEM review.

A first common step is the **creation of groups of buildings**. These groups of buildings are often referred to as **archetypes** in the scientific literature. The main strategies to do so are reviewed by Goy et al. [36] and can be grouped in two main categories:

- Supervised approaches;
- Unsupervised approaches.

The supervised approaches generally create categories based on the features' values of the buildings such as the construction year or the building's usage. Examples of this approach can be found in Rosser et al. [37], in Cerezo-Davila et al. [38] and in Wang et al. [39].

The unsupervised approaches are based on machine-learning techniques that automatically create clusters of buildings with similar properties. Such approach was favoured by Pasichnyi et al. [40], Garrison and New [41], De Jaeger et al. [42] and Ledesma et al. [43].

Once the segmentation of the buildings to simulate has been done, the second step of the enrichment process consists in **attributing the thermal properties, occupancy schedules and all the other missing inputs to each category**.

This parameters allocation can be done using different levels of complexity as depicted Figure 1.2.

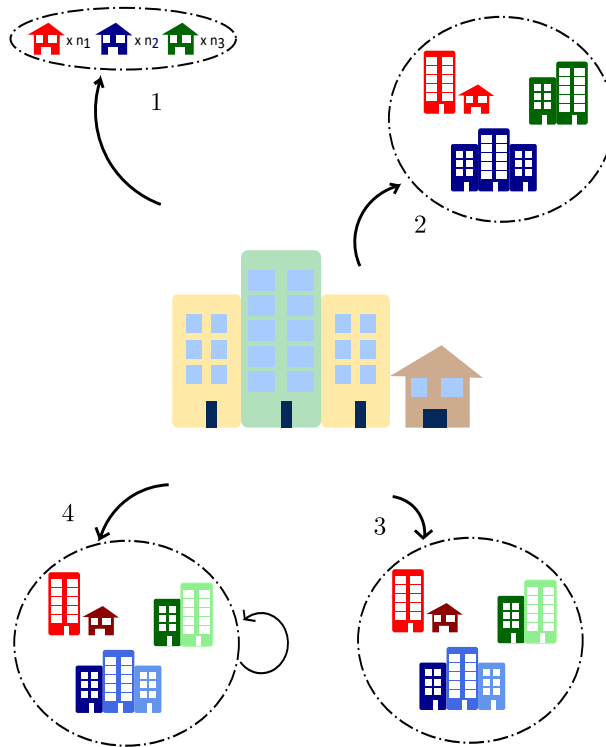


Figure 1.2: Existing approaches for the parameters' allocation step during the enrichment process

The simplest approach is to compute the energy consumption for a given building and then to multiply it by the number of buildings in the category (or by the total floor area of the buildings

in the category). An example of such simple approach is provided by Nagpal and Reinhart [44] but to compare it to a more complex UBEM.

The second approach is to take into account the geometry of each building but to provide each building of the same category with the exact same parameters for thermal properties, schedule occupancy and so on. With this approach, there is almost no diversity in the energy consumption within a same class of buildings. An example of this approach can be found in Cerezo et al. [38].

The third approach consists in considering each unknown parameter as a random variable. Therefore, a probability distribution function is allocated to each class of building for each parameter. The missing parameters for each building of the class are then sampled from these probability distributions. De Jaeger [45] offers a good example of this practice. Such approach can also be found in Schetelat et al. [46].

Lastly, the fourth approach shares a common basis with the previous one, modelling each unknown variable with a probability distribution functions, but takes into account measurements available for each building to update its probability distributions. This approach was well introduced by Cerezo et al. [38] and by Sokol et al. [9], and can be found more recently in Gholami et al. [47].

All of these approaches can be found in UBEM tools and can influence the results obtained with the physical model. For this reason, this step of the simulation should be assessed on the same basis as the physical model to verify UBEM's adequacy to any decision-making process.

1.3 Urban Building Energy Models: modelling purposes and decision-making processes

As it was highlighted above, UBEM are decision support tools. As such, they intervene in decision-making processes. Depending on the decision at stake, different UBEM outputs will be used to inform the decision and these outputs will have a various influence on the final decision. The validation framework dedicated to UBEM should then help to verify that a tool is well-suited to the decision-making process it is expected to support by validating the output at stake in the decision-making process and by verifying that it achieves a level of confidence high enough with regard to its impact on the decision process.

A review of the main UBEM use cases found in the scientific literature is proposed Table 1.3. Each use case corresponds to a decision process, and the output at stake in each case has been identified along with its impact on the final decision. When possible, the accuracy expected for the UBEM output used in the decision process was assessed as well.

Use cases		KPI	Spatial resolution	Range of simulation time step	Decision-making indicator	Required accuracy	Impact of the KPI on the final decision	Stakeholders interested in the use case	Tools / Papers
Development of a district heating or cooling network	Identification of potentials for a district heating network	Heat density (yearly heating demand divided by the area of study)	Building scale	10 min - 1 hour	Same as KPI	20 kWh/m ² /year [48]	++	City managers, urban planners, heating network managers	[31, 33]
	Sizing of a district heating network	Fluid velocity, pressure drop, load profiles	District and building scale	10min	Investment cost	High [49]	+++	Urban planners, heating network managers	[19]
Development of renewable energies at the district scale	Identification of potential renewable sources (waste heat, solar, etc.)	Yearly energy potential, share of renewable energies	District and building scale	10 min - 1 hour	Same as KPI		++	Energy utilities, city managers, urban planners	[21, 23, 28, 30, 44]
	Sizing of the power grid to integrate renewable sources	Load profiles, voltage and electrical intensity	District and building scale	10 min		High	+++	Grid operators, energy utilities	

Use cases	KPI	Spatial resolution	Range of simulation time step	Decision-making indicator	Required accuracy	Impact of the KPI on the final decision	Stakeholders interested in the use case	Tools / Papers
Urban energy planning	Yearly energy demand (including cooling, heating, electric appliances and lighting, domestic hot water, industrial processes)	District scale	1 hour	Levelized cost of energy and total annual costs; CO_2 -equivalent emissions and percentage of renewables used	Low	+	Energy utilities, city managers, urban planners	[21, 23, 28]
Future energy use (example: electric vehicle)	Load profiles	Building scale	1 hour	Same as KPI	Medium	++	Grid operators, energy utilities	[28, 31]
Energy regulation and certification programs	Yearly energy demand (primary or final)	District scale	1 hour - 1 day	Set of indicators defined in the certification process	Low	+	City managers	
Performance measurement and verification (IPMVP)	Yearly energy consumption and load profiles	Building scale	1 hour	Same as KPI	High	++	Heating networks managers, city managers	[50]

Use cases	KPI	Spatial resolution	Range of simulation time step	Decision-making indicator	Required accuracy	Impact of the KPI on the final decision	Stakeholders interested in the use case	Tools / Papers
Energy flexibility	Demand-response measures	Load profiles, indoor temperature	Building scale	10 min to 1 hour	Same as KPI		++	Grid operators, heating networks managers, city managers, utilities [51, 52]
	Modification of the setpoint temperature for heating / cooling	Hourly load profiles	District scale	10 min to 1 hour	Same as KPI		++	Heating networks managers, grid operators, energy utilities [30]
Urban heat island effect	Vulnerability to heat waves	Indoor air temperature,	Building scale	10 min to 1 hour	Same as KPI		+++	City managers, cooling networks managers [19, 28]
	Impact of AC in buildings or cool materials on UHI	Outdoor temperature, energy demand	District scale	10 min to 1 hour	Same as KPI	High	+++	Utilities, grid operators [28]
Analysis of building retrofit	Building envelope retrofit	Yearly energy consumption	Building scale	1 hour	Estimation of payback time [53]	Low	+	City managers [19, 21, 26, 31, 33, 44, 53, 54]
		Load profile	District scale	10 min to 1 hour		High	++	Grid operators, utilities, city managers [28]
	Building cooling / heating system retrofit	Yearly final energy consumption	Building scale	10 min to 1 hour	Estimation of payback year [53]	Low	+	Grid operators, utilities, city managers [21, 26, 31, 53]

Use cases	KPI	Spatial resolution	Range of simulation time step	Decision-making indicator	Required accuracy	Impact of the KPI on the final decision	Stakeholders interested in the use case	Tools / Papers
Upgrade of electronic devices and lighting	Yearly final energy consumption	Building scale	10 min to 1 hour	Estimation of payback year [53]	Low	+	Grid operators, utilities, city managers	[21, 26, 31, 44, 53, 54]

Table 1.3: Review of UBEM use cases

A use case is discussed here as an example: the identification of potentials for a district heating network. For this use case, a relevant KPI is the heat density (defined as the buildings yearly heating demand divided by the area of study). The KPI is a UBEM's output which is essential in the decision-making process for the given use case. A spatial resolution for the KPI is also reported here. In this example, the resolution is set at the building scale. This means that for this use case, the level of accuracy of the KPI should be studied at the building scale. In some cases, the KPI may not be the only indicator considered in the decision-making process. For example, the KPI may help to calculate a payback time and the decision-maker will take his decision based on this economic indicator. In this case the payback time will be regarded as the “final indicator”. The accuracy expected on the KPI and the role it plays on the final decision are also assessed in this table in a qualitative way. The column assessing the impact of the KPI on the final decision was filled using the following legend:

- +: The KPI is not steering the decision-making process, other indicators (economic indicators for example) are used;
- ++: The KPI is steering the decision-making process based upon the relative value of the KPI;
- +++: The KPI is steering the decision-making process based upon the exact value of the KPI.

The weight of the KPI on the final decision therefore emphasizes the need for the UBEM to be sufficiently accurate. Finally, the key players that could be the decision-makers in each use case were also reviewed since they are the final beneficiaries of the UBEM tools.

It can be noticed that the KPI presented in Table 1.3 are mostly limited to two usual physical quantities: annual energy consumption and load profiles. However, depending on the use cases, the accuracy expected from the KPI may vary. For example, for the annual energy consumption, in the use case “Identification of potential renewable sources”, it should be accurate at the building scale while in the use case “Urban energy planning”, it needs to be accurate only at the district scale and not necessarily at the building scale. Furthermore, if the impact of the same KPI on decision-making in both use cases is compared, it seems that the KPI in “Urban energy planning” has a lower influence on the decision-making process than it does in the use case “Identification of potential renewable sources”. Working per use case enables then to ensure that the KPI is fitted for the decision-making process of each use case.

1.4 Thesis approach and structure of the manuscript

In order to assess the adequacy of UBEM to each decision-making process identified in this chapter, a validation framework dedicated to UBEM was developed in this thesis and is presented in this manuscript.

As it was previously highlighted, the aim of this thesis is to contribute to **the development of a validation framework enabling to verify the adequacy of UBEM to their use cases.**

To do so, a **state of the art on validation is required** to identify the strengths and the weaknesses of the existing works on the topic. This state of the art is introduced in **Chapter 2** The existing validation procedures dedicated to Building Energy Models such as the BESTEST were namely discussed and the examples of validation applied to UBEM that can be found in the literature were reviewed. Based on this state of the art, a **validation procedure dedicated to UBEM** could be developed and is introduced at the end of the Chapter 2.

To test the proposed validation procedure, it was applied to a first UBEM and to a first set of use cases. The ad hoc validation of a physical model and an enrichment process is then discussed in **Chapter 3** along with the choice of a set of use cases and the creation of a validation dataset.

The way to **assess UBEM** on this first set of use cases was then questioned. It was namely argued that uncertainty should be taken into account in the validation process. The quantification of uncertainty for the given use cases is then discussed in **Chapter 4** and metrics to properly assess UBEM are proposed. To ensure that the proposed framework is well-suited, it is also applied to three different UBEM.

A validation procedure should always offer a **diagnostic stage**. The approach chosen to identify UBEM weaknesses is then discussed in **Chapter 5**.

Chapter 2

Urban Building Energy Models and Validation

Contents

2.1	Validation: a state of the art	18
2.1.1	Validation at the building scale	18
2.1.2	Validation at the urban scale	21
2.2	A need for a new validation framework	28
2.2.1	Research questions addressed in this thesis	29
2.2.2	UBEM validation framework: a proposal	30

Introduction

As discussed in the previous chapter, Urban Building Energy Models (UBEM) are complex decision-aiding tools evolving in an uncertain environment. Such uncertain context could cast a doubt on UBEM reliability. To ensure that UBEM provide a real aid in the decision-making process they are used for, a thorough validation procedure should be applied to each UBEM tool.

Since many UBEM have been built upon Building Energy Models (BEM), a review of validation procedures for BEM is proposed here, followed by a state of the art on UBEM validation. The aim of this literature review at both building and urban scales is to identify the existing validation approaches and to understand the relationship between BEM validation and UBEM validation.

This literature review also helped to refine the scientific questions addressed in this thesis. These research questions are then presented in this chapter, along with a proposal for a validation framework.

2.1 Validation: a state of the art

2.1.1 Validation at the building scale

Among the ten most important challenges in building energy modelling, Hong et al. [55] identified the building energy performance gap as one of the first issues to tackle. With an increasing need to reduce energy consumption in buildings, BEM are thus expected more than ever to provide accurate prediction of buildings energy performances. However, as described by the authors, the building energy behaviour predicted by BEM often falls quite far from reality. A first step towards reducing this performance gap is BEM validation.

Much work has already been carried out in this area over the past thirty years. From these research works have arisen validation standards such as the BESTEST benchmark [7, 56] and the PASSYS method [6]. A common agreement that arose from the research projects on BEM validation is that BEM are complex tools, relying upon hundreds of parameters, making it very difficult to test every possible parameter configuration. Instead, Jensen [6], in the PASSYS method, offered to break down the validation process into two parts:

- Validation of single processes;
- Validation of the whole model.

This approach is presented in depth in Figure 2.1. It is interesting to note that three techniques can be found in both the single process validation and the whole model validation:

- The sensitivity studies;
- The inter-model comparison;
- The empirical validation.

These three techniques play on two levels that are both of equal importance in a validation methodology:

- An **assessment** part: A validation methodology should provide first a framework to test the tool, along with metrics to properly quantify its performances.

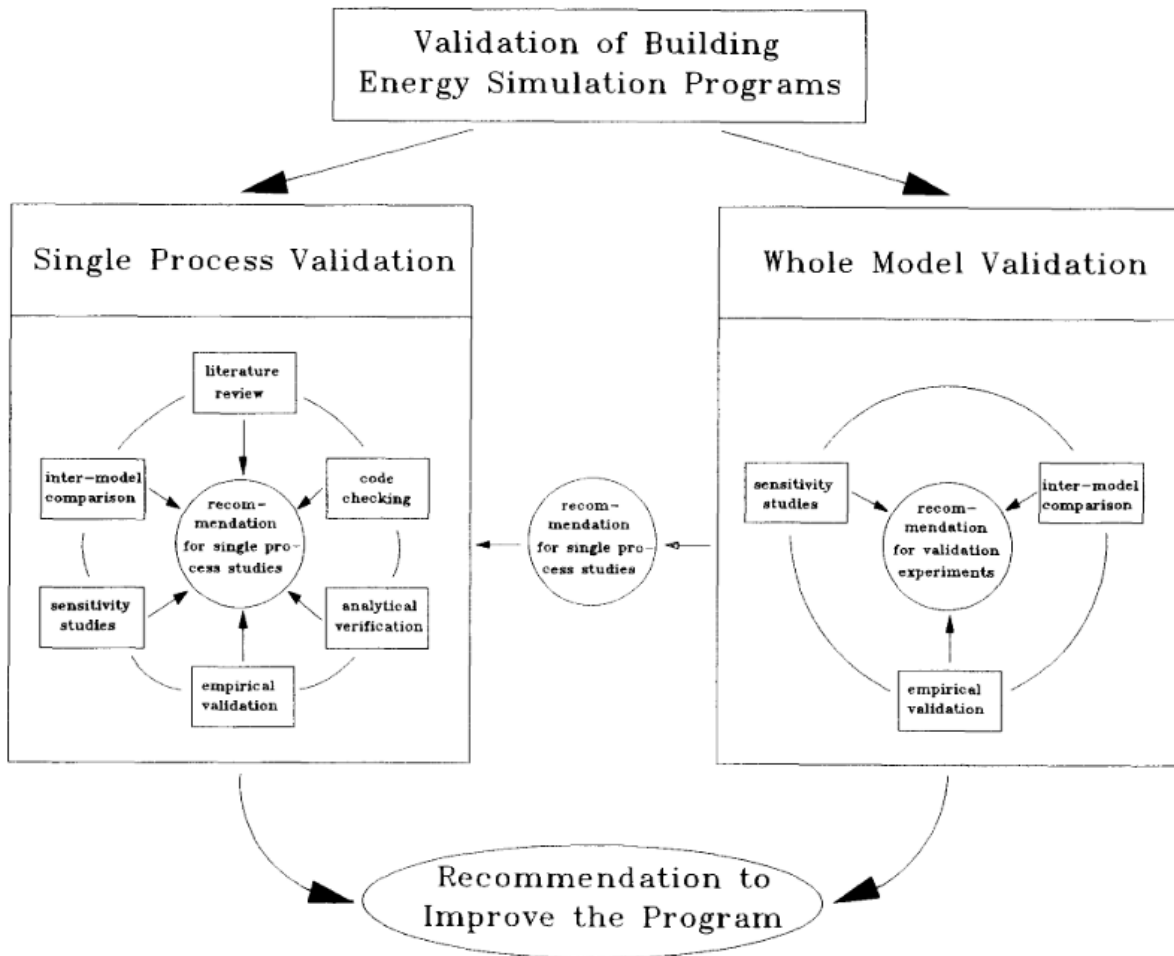


Figure 2.1: Validation method developed in the PASSYS project [6]

- A **diagnosis** part: A validation methodology should then enable modellers to identify where the weaknesses of a tool originate from.

The inter-model comparison and the empirical validation play a role in the assessment step, but, when used jointly, can also provide insights on the diagnosis step. The sensitivity studies help understand inputs influence on outputs variations, thus being an important part of the tools diagnosis.

A similar idea of two-fold validation is proposed in the approach developed by Del Barrio and Guyon [57] and in the BESTEST benchmark [7], the BESTEST being even defined as "Building Energy Simulation Test and Diagnostic Method" by the authors. When looking back at the distinction presented by Jensen [6] between the single processes validation and the whole process validation, the latter could be identified to the assessment part of the validation methodology while the former (single process) could be identified to the diagnosis part, validation of single processes being helpful to identify weaknesses of each tool.

We offer to discuss here first the techniques found in the literature for BEM assessment and then the techniques related to BEM diagnosis.

2.1.1.1 BEM validation: the assessment step

In the assessment step of BEM validation, two main techniques emerge from the scientific literature: the inter-model comparison and the empirical validation.

Inter-model comparison Comparing the results of one BEM to results obtained on the same test case with another BEM is a fairly simple way to start assessing the tool performances. Most of the times, major discrepancies are found among results of different tools. Investigating these discrepancies enables then to identify strong suits and weaknesses of each BEM. Such approach is encouraged in the BESTEST benchmark which provides a set of fully characterized test cases to do so. Other examples of inter-model comparison can be found in Ryan and Sanquist [58] and in Bontemps et al. [59]. A drawback of such technique is that no ground truth is provided, thus making it difficult to assess whether one tool is more valid than another in practice. This drawback can be avoided by selecting for the inter-model comparison at least one BEM already validated against empirical data.

Empirical validation Contrarily to the inter-model comparison, the empirical validation provides a ground truth thanks to measurements on real buildings, although such measurements contain a large share of uncertainty. Ryan and Sanquist [58] identified two types of empirical validation applied to BEM: the **idealised validation** and the **realistic validation**.

The **idealised validation** consists in comparing BEM results to measurements taken on test cells where all parameters are well known. Spitz applied such idealised validation in her thesis [60] where she used the INCA houses to assess EnergyPlus [61]. The INCA houses are experimental houses built and operated by the National Institute for Solar Energy in Le Bourget du Lac, France. The thermophysical parameters of these houses are well known and operating parameters such as operative temperature are well monitored. Uncertainties related to the occupant behaviour are not considered as the houses are uninhabited. Similar work was led by Palomo Del Barrio and Guyon [62] on the ETNA building, an experimental building designed and operated by Electricité de France (EDF) in the Paris region. Such idealised validation enables to reduce uncertainty on most parameters and thus to identify some issues related to the physical model and its numerical implementation. However, as highlighted by Ryan and Sanquist [58], it does not take into account realistic conditions of use of buildings and thus does not assess the ability of BEM to accurately take into account other more uncertain parameters such as occupants behaviour.

Realistic validation on the contrary is a comparison between models and measurements performed on operating buildings, thus accounting for real conditions of use. Examples of this approach can be found in [58]. Such approach is critical to reduce the building performance gap identified by Hong et al. [55].

To do so, taking into account uncertainty in the empirical validation is of the utmost importance as advocated by Palomo Del Barrio and Guyon [57]. Whether working on idealised or realistic validation, uncertainty should always be accounted for. Propagating uncertainty through the model enables the modeller to retrieve a distribution of the output to study instead of a single value. Likewise, the reference value is replaced by an uncertainty band. The validation process would then consist in comparing the uncertainty bands of both measurements and BEM results and their overlapping would assess the tool validity.

When taking a step back and looking at validation of physical models in general, uncertainty is regarded as the cornerstone of most validation processes [63]. With this in mind, a comprehensive review of uncertainty analysis in the building energy modelling field was proposed by Tian et al. [64]. The authors acknowledge a fair spread of uncertainty analysis techniques among the

BEM community namely due to mature techniques in the uncertainty analysis field and to easy to implement techniques such as Monte-Carlo sampling. However, the uncertainty quantification of some complex parameters such as occupancy remains difficult and further work should be led on this aspect.

2.1.1.2 BEM validation: the diagnosis step

The uncertainty quantification enables to implement a strong diagnosis tool: the sensitivity analysis. The sensitivity analysis is defined by Saltelli et al. as "the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" [65]. Depending on the techniques considered, it enables to answer several questions such as:

- Which input has the greatest influence on the output ? And consequently, for which input should we reinforce the quality of data collection ?
- Which input could be set to a fix value ?
- Are there inputs that are highly correlated ?

Examples of sensitivity analysis applied to building energy models are provided by Goffart et al [66], Bontemps [67] and by Spitz [60]. In the case of Bontemps [67] and Spitz [60], sensitivity analysis is used within the framework of empirical validation. It enables thus to test the model and identify the contribution of each uncertain parameters to the output variation.

A similar approach, although not completely characterized as a sensitivity analysis, can be found in the BESTEST benchmark [56]. Several test cases, with an increasing level of complexity, are proposed, thus enabling to assess which part of the model could be erroneous.

A slightly different diagnostic approach is offered by Palomo Del Barrio and Guyon [57]. The authors proposed to proceed first with a principal components analysis (PCA) to study parameters correlation and second, with a parameters' optimisation to compare original parameter values with optimal ones. The parameters optimisation should enable to identify parameters that need to vary the most in order for the BEM results to match the available measurements.

In this part, we highlighted the maturity level reached for BEM validation. As UBEM are more recent tools, their validation has not yet been the focus of as many research projects. This topic is discussed hereafter.

2.1.2 Validation at the urban scale

2.1.2.1 An overview of UBEM validation specificities

When thinking about UBEM validation, one can wonder what the differences are with BEM validation. A first obvious one is the spatial scale. Since UBEM are designed to model several buildings in parallel, there is a question on the number of buildings that should be considered through the validation process. Should the modellers assess their tools on one, ten or a thousand buildings ? A connected question is the aggregation level at which one should analyse a tool results during the validation process: should it be checked building per building or at the aggregated scale of a district or a city ? A second major difference with BEM is the enrichment process inherent to UBEM. This enrichment process brings an additional share of uncertainty and may jeopardize the reliability of UBEM if not properly assessed.

Authors	Validation type	Assessed outputs			UA ¹	Metrics	Location	Open data	Validation data		
		Outputs	Spatial scale						Nb of items	Timestep	Spatial scale
Saelens et al. [8]	Inter-model comparison	Annual heat demand Heating peak power Overheating	Building	No	None	Idealized ditrict tested with a Belgium climate	Yes				
Nageler et al. [68]	Empirical validation	Heating demand and DHW demand	Aggregated by building type and at building scale	No	PE ME RMSE ESD MAE	Gleisdorf	No	69	Year	Building	
Wang et al. [39]	Empirical validation	Annual gas consumption	Postcode scale	Yes	APE CVRMSE	Amsterdam	Yes	84	Year	Postcode	
Dochev et al. [69]	Inter-model comparison	Heated volume Archetypes categorization U-values Window area Heat demand	District and building	No	SMAPE	Berlin	No				
Nagpal et al. [44]	Empirical and Inter-model comparison	Annual energy consumption Areas	Building Usage	No	R ²	Cambridge, USA	No	100	Year	Building	
		Energy savings	Building and Campus								
		Individual energy saving strategies	Building								
					RNRMSE			2	Hour	Building	

¹ Uncertainty Analysis

Authors	Validation type	Assessed outputs			Metrics	Location	Open data	Validation data		
		Outputs	Spatial scale	UA ¹				Nb of items	Timestep	Spatial scale
Nageler et al. [70]	Empirical and Inter-model comparison	Heating load profile	Building and district	No	CVRMSE MBE R ²		No	34	Hour	District
Fonseca et al. [71]	Inter-model comparison	Annual heating energy demand Annual cooling energy demand Annual primary total energy demand	Stock	Yes	RSS	Portugal	No	170000	Year	Distribution at the sample scale
Walter et al. [72]	Empirical and Inter-model comparison	Annual heating consumption Annual cooling demand Peak heating demand Peak cooling demand	Building	No	None	Lausanne, Switzerland	No	1	Year	Building
Sokol et al. [9]	Empirical and Inter-model comparison	Monthly gas and electricity consumption	Building	Yes	PE CVRMSE	Cambridge, USA	No	2263	Month	Building
Garrison et al. [41]	Empirical validation	15-min electricity consumption profiles	Building	No	Kolmogorov-Smirnov test AE RE RMSE threshold CVRMSE NMBE	East Tennessee and Georgia	No	178000	15 min	Building

¹ Uncertainty Analysis

Authors	Validation type	Assessed outputs			UA ¹	Metrics	Location	Open data	Validation data		
		Outputs	Spatial scale						Nb of items	Timestep	Spatial scale
Yang et al. [73]	Empirical validation	Annual gas consumption	Postcode	No	AE MBE CVRMSE	Leiden	Yes	1292	Year	Postcode	
Taniguchi et al. [74]	Empirical validation	Hourly electricity consumption	Sample	No	None	Kansai region, Japan	No	1	Hour	1237 aggregated households	
Widen et al. [10]	Empirical validation	Hourly electricity consumption	Household	No	NVF	Sweden	No	13	Hour	Household	
Fischer et al. [75]	Empirical validation	Hourly electricity consumption	Household	No	MBE Correlations	Germany	No	430	Hour	Household	
Fischer et al. [76]	Empirical and Inter-model comparison	Heating and DHW energy consumption	Household	No	Peak value RMSE R ²	Germany	No	10	Hour	Household	
Fonseca et al. [31]	Empirical and Inter-model comparison	Annual electricity consumption Annual heating consumption Annual cooling consumption	Building	No	PE	Switzerland	No	23	Annual	Building	
Bünning et al. [77]	Empirical and Inter-model comparison	Heating load profile	Building	No	R ²	Switzerland	No	4	15 min	Building	
Nouvel et al. [78]	Empirical and Inter-model comparison	Energy Reference Area Annual Heating Demand	District Building	Yes	PE MAPE APE MPE	Ludwisbrug, Germany	No	28	Year	Building	
Todeschi et al. [79]	Empirical and Inter-model comparison	Annual space heating consumption	Building	Yes	MAE	Fribourg, Switzerland	No	198	Hour	Building	

¹ Uncertainty Analysis

Authors	Validation type	Assessed outputs			Metrics	Location	Open data	Validation data		
		Outputs	Spatial scale	UA ¹				Nb of items	Timestep	Spatial scale

MAPE

Table 2.1: Review of validation attempts

¹ Uncertainty Analysis

Acronym	Full expression
AE	Absolute Error
APE	Absolute Percentage of Error
CVRMSE	Coefficient of Variation of the Root Mean Square Error
ESD	Empirical Standard Deviation
MAE	Maximum Absolute Error
MAPE	Maximum Absolute Percentage of Error
MBE	Mean Bias Error
ME	Mean Error
MPE	Mean Percentage of Error
NMBE	Normalized Mean Bias Error
NVF	Normalized Variation Factor
PE	Percentage of Error
RE	Relative Error
RMSE	Root Mean Square Error
RNRMSE	Range Normalized Root Mean Square Error
RSS	Residual Sum of Squares
R ²	Coefficient of determination
SMAPE	Symmetrical Mean Absolute Percentage of Error

An overview of the main papers related to the topic is proposed in Table 2.1, pointing out the specificities of UBEM validation attempts.

In this table, the focus was set mainly on the assessment part of the validation process and the reviewed assessment approaches are discussed hereafter. However, as stated for BEM validation, a comprehensive validation process should not only assess tools' performances but also offer a diagnosis step to identify the tools' weaknesses. The diagnosis step as found in the literature is then discussed in a following section.

2.1.2.2 UBEM assessment

As described in Table 2.1, two main approaches are found for UBEM assessment: empirical validation and inter-model comparison. Empirical validation is the most widely spread practise and inter-model comparison is often combined with empirical validation and rarely led on its own. The empirical validation attempts are discussed here first, followed by inter-model comparisons attempts.

Empirical validation Out of the 19 papers reviewed in Table 2.1, 16 papers perform an empirical validation. Although the majority of the reviewed papers can be put into this category, significant differences can be observed among these papers when taking a closer look.

First, when considering **the size of the validation dataset** used in the empirical validation process, it ranges from a single building [72] to 178,000 buildings [41], and the majority of the papers discuss the validity of their models over a validation dataset of less than a hundred buildings. Since UBEM are bottom-up tools, the modellers are often interested in analysing the results of their tools building per building. However, when considering a validation dataset of less than hundred buildings, the statistical representativeness of the conclusions drawn from the validation process could be called into question. There is no guarantee that such a small validation dataset enables to test the UBEM tool on its full range of use, thus leaving blind corners in the validation process. Although it seems rather difficult to create a validation dataset enabling to fully test a UBEM tool, the authors should at least tend towards this aim and identify the limitations of their validation dataset regarding the generalization of their results. More generally, a validation dataset is a sample retrieved from a larger population, and, as for any sample, its representativeness with respect to some key parameters must be verified.

Secondly, **a large variety of metrics** can be found for the tools' assessment, and it seems that no agreement has been reached yet on this topic among researchers. However, several metrics can be found quite frequently such as:

- The Percentage of Error (PE);
- The Root Mean Square Error (RMSE);
- The Coefficient of Variation of the Root Mean Square Error (CVRMSE);
- The Coefficient of Determination (R^2).

An agreement on which metrics to use in the validation process would help compare results among different studies, and thus would enable the comparison of UBEMs performances.

This lack of agreement can also be found when considering inter-model comparison, though to a lesser extent since inter-model comparison frameworks have often been built through international workgroups.

Enabling to compare results among different papers would strengthen the validation process of UBEM tools as would do the **availability of the validation dataset**. As stated in Table 2.1, most of the papers reviewed do not make the validation dataset used in their study public. Thus, Malhotra et al. [80] identified that 95% of the papers published in the UBEM field lack replicability. Such trend is worrying and a thorough validation framework of UBEM should work against such practices by ensuring the complete replicability of the validation procedure.

Another striking point when reviewing validation papers is the **lack of studies taking into account the uncertainty** surrounding UBEM tools in their validation process. Although the modelling of uncertainty in the validation process of physical models should be a standard practice as advocated by Moser and Oliver [63], it is not yet the case for UBEM. However, a few recent works have been observed on the topic of uncertainty modelling such as the thesis of Ina De Jaeger [45] that offered a method to model uncertainty on UBEM inputs data. Pratavia et al. [81] presented a similar approach for uncertainty modelling but studied a dynamic output conversely to Ina De Jaeger who studied mainly annual heating energy. Such trend towards inclusion of uncertainty modelling in the UBEM validation process should be encouraged.

Inter-model comparison There are fewer examples of inter-model comparison in the literature. However, the DESTEST proposed by Saelens et al. [8] deserves to be discussed more extensively here. This validation framework is the result of a group project, IBPSA Project 1, that gathered modellers from different backgrounds to create a virtual district to compare tools' performances on both buildings and energy systems modelling. The virtual district is at the moment composed of 16 identical buildings. These buildings are as simple as possible and simulations were run to compare performances of five different UBEM tools, some working under the Modelica environment and some working under other coding environments. The aim of a very simple test case was to easily identify the origin of discrepancies among tools. However, even this first step appeared to be quite challenging. A second test case to assess the modelling of district heating networks was also developed, and a variety of test cases of increasing complexity should be proposed in future works.

One of the main achievements of this project is to have had researchers from different backgrounds sitting around the table and defining together a validation framework. The proposed DESTEST is thus of great interest to compare performances of several models and to test single processes used in UBEM. However, it does not consider the uncertainty surrounding UBEM in their real conditions of use, and the limitations of this approach are similar to the ones identified by Ryan and Sanquist [58] for the BESTEST.

Other examples of inter-model comparison are found in the literature, namely to assess the validity of individual modules against a baseline. Such approach can be found in Rosser et al. [37] where the authors compare different geometry simplification processes, in Happle et al. [82] where the authors study the influence of different air infiltration models and in Chapman et al. [83] where the authors assess the performances of a stochastic occupancy model against a deterministic approach. A more comprehensive framework to assess the models level of details to be used within UBEM has been developed by Enora Garreau in her thesis and the reader is referred to her manuscript for more information on the topic [84].

2.1.2.3 UBEM diagnosis

Conversely to the assessment step, the diagnosis step of the validation process is barely discussed in the literature. A typical example of this trend is provided in the review conducted by Fennell et

al. [85]. The authors selected the top hundred most cited papers containing the terms: "(energy AND building AND model) AND (uncertainty OR sensitivity OR probabilistic OR stochastic) AND (city OR building stock)". Among these hundred papers, only fourteen actually undertook an uncertainty or sensitivity analysis at the district or city scale.

Furthermore, when a sensitivity analysis is carried out, the authors often limit themselves to a single building as it is the case in Todeschi et al. [79]. Such practise could lead to wrong conclusions since it limits the type of parameters accounted for during the sensitivity analysis. For instance, in the case of Todeschi et al., the most influential parameter identified by the authors was the ventilation rate but when comparing to another sensitivity analysis led by Booth et al. [86], there is no agreement on the result. Booth et al. identified the fraction of heated space as the most influential parameter. Such disagreement can be explained by several facts such as a difference in the studied model or a different way to model the uncertainty, but working on a different perimeter certainly do not help to reach common conclusions.

Quite close from a sensitivity analysis although different, Nouvel et al. [78] studied the influence of several levels of knowledge on the computed annual heating demand. The authors studied thus the impact of the geometry level of detail and the influence of ignoring a heated basement or attic on the predicted heating demand. Such work calls into question the impact of the input data quality and tries to assess its consequences on the UBEM performances.

Along with the quality of input data, the **enrichment process** should be put under the spotlight during the validation process, to challenge the hypotheses made at this step of the simulation and to understand how it influences the results. A good example of validation of the enrichment process alone is offered by Rosser et al. [87]. The authors assessed the performances of a machine-learning algorithm to predict buildings age when unknown. They used common techniques of model selection and model evaluation from the machine learning field to assess the performances of their enrichment process. Among these techniques are the cross-validation for the model selection and the validation on a sample, different from the training sample. The latter is namely a good practice from which the UBEM field would greatly benefit. At the moment, the distinction between the dataset used for the training of the model and the dataset used to assess the model performances is not well acknowledged within the UBEM community. This distinction is of great importance since it enables to verify how well a model behaves when faced with data outside from its training set. Such aspect is particularly critical to enable the generalization of the results obtained through the validation process. A counter-example is provided by Garrison and New [41] where the authors used the same measurements to categorize buildings into archetypes and to assess the performances of the energy simulation of these archetypes.

2.2 A need for a new validation framework

We discussed above the importance of validation for BEM and identified **two main steps in a validation process**: the **assessment of BEM performances** and the **diagnosis** of BEM weaknesses. Several techniques were identified to contribute to the assessment step such as inter-model comparison and empirical validation. The critical role of uncertainty was also discussed along with its appropriation by the BEM community. Regarding the diagnosis step of a BEM validation process, sensitivity analysis, parameters optimisation and the BESTEST test cases were discussed.

When considering UBEM, a lot of research has been carried out on empirical validation and a little less on inter-model comparison. Regarding empirical validation, the generalization of the results is not very robust yet, namely due to the size of the used validation datasets, a lack of common metrics, the non-disclosure of the validation datasets and the lack of uncertainty

modelling. When looking at inter-model comparison, an interesting, though limited for now, work has been proposed within the IBPSA Project 1 workgroup. Similar workgroups should be encouraged on UBEM empirical validation and would help to reach an agreement on metrics and validation datasets constitution.

Regarding the diagnosis of UBEM weaknesses, there is still a long way to go. Thus, uncertainty analysis is still emerging, although good examples have been set by De Jaeger [45] and by Pratavia et al. [81], and sensitivity analysis is almost inexistent among the UBEM community, as stated by Fennell et al. [85].

To summarize, a lot of existing work deals with parts of the UBEM validation process, but no comprehensive framework has been developed so far. Such validation framework should account for the whole simulation process, including the quality of the input data and the enrichment process. Furthermore, requirements should be set to ensure that the validation dataset is large and diverse enough to test the UBEM tools on their full range of use. The opening of such dataset would also improve the replicability of the validation procedure.

In order to overcome the challenges identified above, the goal of this thesis is to put forward a first full validation framework.

The proposed framework was developed to be as comprehensive as possible and to be applicable to the majority of the existing UBEM tools that were reviewed in Chapter 1. The scope of this research work is introduced in the next section and the proposed validation framework is presented afterwards.

2.2.1 Research questions addressed in this thesis

UBEM are first and foremost decision support tools. They were thus developed to help decision makers arbitrate among several choices. Their validation should then be considered from this viewpoint. Likewise, the main research question driving this thesis work is: **How to verify the adequacy of UBEM to the decision process they are used for?**

This main question gives way to two sub-questions:

- What are the different decision processes UBEM are involved in ?
- How to develop a comprehensive and generic validation framework for UBEM ?

These two questions are discussed hereafter.

2.2.1.1 What are the different decision processes UBEM are involved in ?

A review of UBEM use cases has been proposed in Chapter 1. In this review, the different decision processes were identified as well as the UBEM outputs each decision process relies on. A certain level of confidence is expected from the UBEM output in each decision process and was assessed in Chapter 1 as well. This level of confidence should also be accounted for in the validation of UBEM tools. Thus, the validation framework should enable to assess the outputs used in the decision process and to verify whether these outputs are robust enough to achieve the desired level of confidence. Such a requirement forces to test the tools under their conditions of use since a large part of the outputs uncertainty may be caused by the input data and the enrichment process. Ideally, the validation framework should be the same for all the use cases even though a validation procedure should be led separately for each decision process.

2.2.1.2 How to develop a comprehensive and reusable validation framework for UBEM ?

As highlighted above, the validation framework should enable to test the tools under their real condition of use. With this in mind and based on the literature review discussed previously, a list of requirements for a comprehensive validation framework can be drawn:

- Work on the full simulation chain, including the enrichment process;
- Take into account the existing uncertainty at each step of the simulation;
- Create a validation dataset large and diverse enough to provide statistically representative results;
- Make the validation datasets publicly available;
- Define metrics to assess UBEM tools;
- Provide indicators for the diagnosis of UBEM weaknesses.

As a result of the work carried out during this PhD to answer these research questions, a proposal for a validation framework has emerged, verifying the list of requirements developed above. This validation framework is presented below.

2.2.2 UBEM validation framework: a proposal

The proposed validation framework is introduced Figure 2.2. It can be broken down in several steps that are discussed here.

The first step of the validation framework consists in ad hoc validation of the two main components of each UBEM tool:

- **Ad hoc validation of the enrichment process,**
- **Ad hoc validation of the physical model.**

To validate the enrichment process, an approach similar to the one proposed by Rosser et al. [87] should be followed. The authors used techniques from the machine learning field, namely cross-validation, to assess the performances of their enrichment process. This practise is a good way to verify the ability of an enrichment process to provide accurate results beyond its training set.

When considering the validation of the physical model, existing procedures provided by both the BESTEST [56] and the DESTEST [8] should be followed. The BESTEST should enable to fully test the building energy modelling part, while the DESTEST should help assess performances of submodels specific to UBEM such as district heating networks.

The ad hoc validations of the tools used in this manuscript is presented **Chapter 3**.

The second step of the validation framework is **to review the UBEM use cases and to identify the different outputs** at stake in the decision process. These outputs are the Key Performance Indicators (KPI) that will be assessed throughout the validation procedure. For

each use case and each linked decision process, a certain level of accuracy is expected on the KPI. This level of accuracy should be identified and fixed at this step. Ideally, this threshold should be fixed through exchanges with the concerned decision-makers.

Several use cases may have the same KPI although with a different expected accuracy. These use cases should be grouped together for the following steps of the validation procedure until the assessment step. At the assessment step, the expected accuracy will be taken into account, thus differentiating the use cases.

The third step consists in **creating a validation dataset** to assess a given KPI. This validation dataset should verify several key points that are:

- Measurements of the KPI should be available for every building of the validation dataset. This would enable to study the validity of the outputs at different spatial scales ranging from the single building to the building stock.
- The validation dataset should contain enough buildings to be statistically representative of the building stock to which the modellers aim to apply the UBEM. Such representativeness is not always easy to assess but the distribution of some general parameters such as the building usage or the year of construction should be compared between the validation dataset and the targeted building stock. Limitations of the validation dataset should thus be assessed and taken into account when looking at the validation results.
- There should be as much descriptive information available at the building scale as possible. For instance, each selected building should display an Energy Performance Certificate which would describe the thermal properties and the heating and Domestic Hot Water (DHW) systems of the considered building. Having information at the disaggregated scale of the building enables to analyse more precisely the error and potentially to identify groups of buildings for which the UBEM fails to accurately simulate their energetic behaviour.
- The validation dataset should be made public to ensure the replicability of the validation procedure. This is a non-trivial question since the opening of private data is generally challenging if not impossible. Therefore, the validation dataset should be built upon existing public datasets as much as possible. Even though privacy concerns often hinders the opening of datasets, a general trend towards open data can be observed in Europe. For instance, in France, annual electricity and gas consumptions were made available at district scale at first and recently became available at the address scale [88]. Examples of similar datasets in the Netherlands can be found in several articles [39, 73]. It seems that such data are at the moment available only at the district scale, but one can hope that it will follow the same trend as in the case of France. Similarly, annual energy consumption data are already available at the building scale in the US as described by Kontokosta and Tull [18]. Regarding energy consumption data at a finer time scale, there is still a long way to go to make such datasets public, and a thorough consideration should be given to this question and to the related privacy issues. However, it is already possible in many countries to access these data for research purpose only. Alternatively, researchers could use virtual data sets like the one developed by Neale et al. [89]. The main limitation is then the prior validation of the virtual data set.

The construction of such validation dataset is discussed more extensively in **Chapter 3**.

The fourth step is **the quantification of the uncertainty** surrounding the UBEM simulation. This step is critical since a poor consideration of uncertainty could bias the rest of the validation procedure.

As explained by Walker et al. [90], uncertainty is often divided into two categories : epistemic and aleatory uncertainties. The former refers to uncertainty due to a lack of knowledge whereas the latter corresponds to the inherent variability of an object such as the occupants' behaviour in a building. Both types of uncertainty are often combined. For instance, when modelling the buildings of a district, the number of occupants per building is unknown (epistemic uncertainty) but, even if it was known, there would remain an uncertainty related to the occupants' behaviour (aleatory uncertainty). The representation of aleatory uncertainty is generally accepted as a probability distribution [91]. However, as highlighted by Oberkampff et al. [91], the representation of the epistemic uncertainty is still subject to discussions. The authors advocated thus to use the Dempster-Shafer theory to model it. Other possibilities include fuzzy theory, possibility theory and classical probability theory. For the sake of simplicity and to facilitate the appropriation of uncertainty analysis methods throughout the UBEM field, the representation of epistemic uncertainty is done with probability distributions in this PhD work.

Another question arises when modelling epistemic uncertainty as probability distribution: how to model the relationship among the different uncertain parameters? In the case of UBEM, it is quite likely that uncertain parameters are related to each other. For instance, a recent building would surely have a low U-value for its walls but also for its roof and windows. Such relationship should then be accounted for in the uncertainty quantification step. To do so, Ina De Jaeger [45] offered to use copulas in a first attempt to account for relationships among uncertain variables in the UBEM field. However, the copulas do not seem to be the most appropriate tool since it does not work well for discrete variables [92]. Therefore, it could not be used to model some important relationships such as the relationship between the heating system of buildings (discrete variable) and the walls U-value (continuous variable). Another approach is proposed here, based on Bayesian Networks as introduced by Schetelat et al. [46]. This approach will be discussed more in depth in **Chapter 4**. More generally, the definition of probability density functions to model uncertain parameters is challenging and good practices introduced by the Joint Committee for Guides in Metrology should be followed [93].

Once a probability density function has been assigned to each parameter, these probability density functions must be sampled thanks to a sampling algorithm. A review of the different sampling algorithms for independent variables is provided by Saltelli et al. [65]

Since the results of an uncertainty analysis always depend on the way the uncertainty was modelled [94], all UBEM tools tested on a same validation dataset should benefit from the same uncertainty quantification. The code defining the probability density functions as well as the selected sampling methods should then be shared among modellers to encourage inter-models comparison. Such practice would reinforce replicability and would help challenge the uncertainty quantification process.

The fifth step consists in **propagating the uncertainty** through the whole UBEM simulation. It consists in simulating each combination of uncertain parameters obtained at the previous step. This should then enable to obtain the distribution of the KPI selected for the validation process. This step may be computationally expensive but is unavoidable to properly assess the UBEM performances. At this step, a convergence study should be led by varying the number of simulated samples in order to ensure that the rest of the analysis will be done on a consistent distribution of the KPI.

The sixth step is then to **assess the obtained KPI distribution** against the measurements. To do so, **metrics must be defined to properly quantify the error**. The UBEM perfor-

mances should then be assessed at the building scale but also at the scale of the whole validation dataset. Although different metrics could be used depending on the studied KPI, the same set of metrics should always be used to assess the same KPI.

At this point, the value of the metrics should enable to verify whether the assessed UBEM is accurate enough for the decision process that was considered initially. The different use cases and their respective level of accuracy should then be considered separately.

The seventh step finally offers to **gain a better understanding regarding the UBEM weaknesses** and the causes for the observed discrepancies between UBEM results and measurements. At this step, it would be interesting to lead a sensitivity analysis. However, since correlations among variables were accounted for during the uncertainty quantification, classical sensitivity analysis techniques cannot be used here as explained by Saltelli et al. [65]. Other statistical indicators are then introduced to identify relationships among uncertain parameters and the error at both the building scale and the scale of the whole validation dataset. Such indicators are detailed in **Chapter 5**.

Conclusion

This chapter presented the state of the art of BEM and UBEM validation. Although the first one benefits from many years of research on validation, a performance gap remains that current validation procedures struggle to close.

Regarding UBEM validation, various attempts are found in the literature, but there is a clear lack of a unified validation procedure. Furthermore, the existing attempts often lack replicability and generalization of their results due to too small and unavailable validation datasets. For these reasons, the focus of this PhD work was set on developing a comprehensive validation framework, aiming to unite the different approaches found within the UBEM field. The research questions addressed in this manuscript were then discussed and a proposal for a UBEM validation framework was presented here.

To verify to what extent the proposed validation framework is well suited to UBEM validation, it was applied to a first set of use cases. The chosen use cases and the construction of the associated validation dataset is presented in the next chapter.

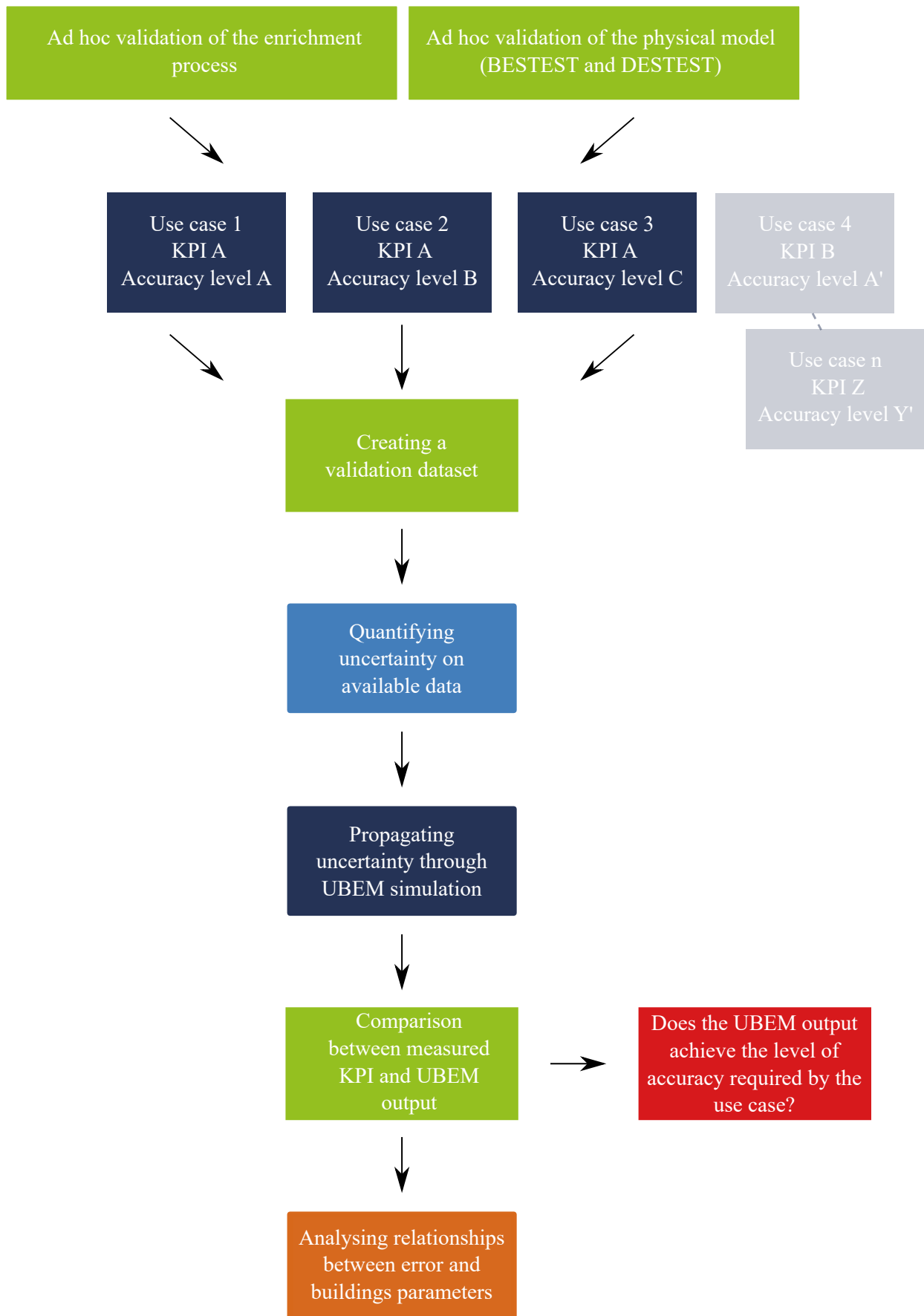


Figure 2.2: Validation framework for UBEM: a proposal

Chapter 3

Application of the validation framework to a first set of use cases

Contents

3.1	Ad hoc validation of both the physical model and the enrichment process	36
3.1.1	DIMOSIM: a bottom-up physical simulation tool	36
3.1.2	An enrichment process: Qiriél	41
3.2	Choice of the use cases	51
3.2.1	Lessons learnt from a first test case on a French department	51
3.2.2	A choice driven by the data availability	52
3.3	Creation of a validation dataset	53
3.3.1	Available databases	53
3.3.2	Presentation of the obtained validation dataset	56

Introduction

In order to legitimate the validation framework presented in the previous chapter, this framework was applied to a first set of use cases and to a first set of UBEM tools.

In this third chapter, the three first steps of the validation framework are then discussed:

- The ad hoc validation of the physical model and of the enrichment process;
- The identification of use cases and their related outputs for which UBEM tools should and could be validated;
- The creation of a validation dataset designed to test the selected outputs.

The ad hoc validation procedure for the physical model is detailed here on DIMOSIM [11], a bottom-up model that was already used in several research projects. Regarding the enrichment process, classical validation procedures from the machine learning field are applied to Qiriel [95], a tool dedicated to the enrichment process.

Next, the set of use cases selected for this first application of the validation framework is introduced. These use cases were selected mainly based on data availability and on the simplicity of the outputs to assess.

Finally, the creation of the validation dataset required to validate UBEM on the chosen use cases is detailed.

3.1 Ad hoc validation of both the physical model and the enrichment process

In this manuscript, the proposed validation framework is applied to three different tools, among which is DIMOSIM [11], combined with a specific enrichment process, Qiriel [95]. As DIMOSIM was the most advanced tool for engineering applications among the three tools evaluated in Chapter 4, the ad hoc validation of the physical model required in the validation process is detailed on this model.

DIMOSIM is presented hereafter along with the results it obtained on both the BESTEST and the DESTEST. Qiriel is then introduced and the ad hoc validation procedure of this enrichment process is discussed.

3.1.1 DIMOSIM: a bottom-up physical simulation tool

3.1.1.1 Presentation of DIMOSIM

DIMOSIM is a modular tool whose aim is to allow simulation of energy systems from the scale of the building to that of an entire city. A comprehensive presentation of DIMOSIM and a detailed description of its composition is provided by Garreau et al. in [11]. A brief review of its main abilities is provided hereunder, but the reader is referred to this paper for a more detailed presentation of the tool.

DIMOSIM consists in a Python package which is programmed in an object-oriented structure. The different key objects found in DIMOSIM and their relationships are presented in Figure 3.1. This structure enables to simulate energy behaviour of buildings with different level of details and to study the results at different scales ranging from a single thermal zone to a whole district.

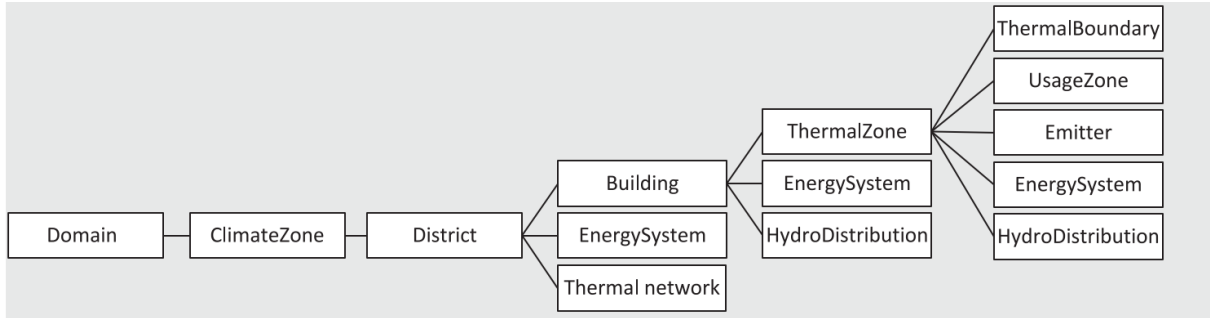


Figure 3.1: Object structure in DIMOSIM [11].

DIMOSIM falls in a rather detailed category of RiCj (electrical-thermal analogy) models and its thermal model is presented in Figure 3.2. This configuration enables DIMOSIM to adjust to the level of information available and to the use case. For instance, it is possible to group together all the walls during the calculation when detailed results are not required, thus reducing the computation time.

DIMOSIM has already been applied to a large variety of projects, as described by Garreau et al. [11], among which are the study of urban heat island effect, the study of district heating networks or the creation of a metamodel for optimization of refurbishment choices. Moreover, PowerDIS, a web-based software is currently developed by Efficacity, a French institute for cities energy transition, to offer a user interface to DIMOSIM which shall help engineering firms and urban policy-makers to use this UBEEM in their day-to-day projects.

3.1.1.2 Ad hoc validation of DIMOSIM

The ad hoc validation of a complex physical model such as DIMOSIM can be manifold.

First, the BESTEST benchmark [7] has been applied to DIMOSIM to verify how the thermal model behaves in regard to BEM. The results were presented by Garreau et al. in [11] and are displayed Figure 3.3. DIMOSIM provides results that are satisfactory enough since it always performs within the range of values obtained by all BESTEST candidates.

Second, DIMOSIM was used within the IBPSA Project One and tested on the proposed DESTEST [8]. The DESTEST test cases were two-fold.

The first test case consists in simulating a cubical building and to compare computed heating power and air temperature among the different tools. These results were presented for a week in March under a Belgian climate and are reproduced here in Figure 3.4. Although, discrepancies among models are noticeable, DIMOSIM behaves within the average and provides acceptable results.

The second test case consists in assessing the ability of the different UBEEM tools to model a district heating network by comparing the thermal losses obtained by each tool. Once again, DIMOSIM provides results that are satisfactory and no major discrepancy is observed with the other tools.

Third, Enora Garreau, in her thesis [84], studied the level of parsimony that should be implemented in UBEEM tools and challenged several sub-models implemented in DIMOSIM thus helping to identify the most parsimonious sub-models to apply for the modelling of solar masks, thermal zones and district heating networks. This increased again the level of confidence one can have in DIMOSIM.

Finally, DIMOSIM has been used on many projects and often compared to measured consump-

tions in buildings and districts. This is still not enough to consider the tool as fully validated, but it helps increase the confidence in the overall robustness of the tool. All in all, the robustness of DIMOSIM and its proper behaviour when considering idealised evaluation frameworks has been well established and all these arguments are considered sufficient for the ad hoc validation of the physical model.

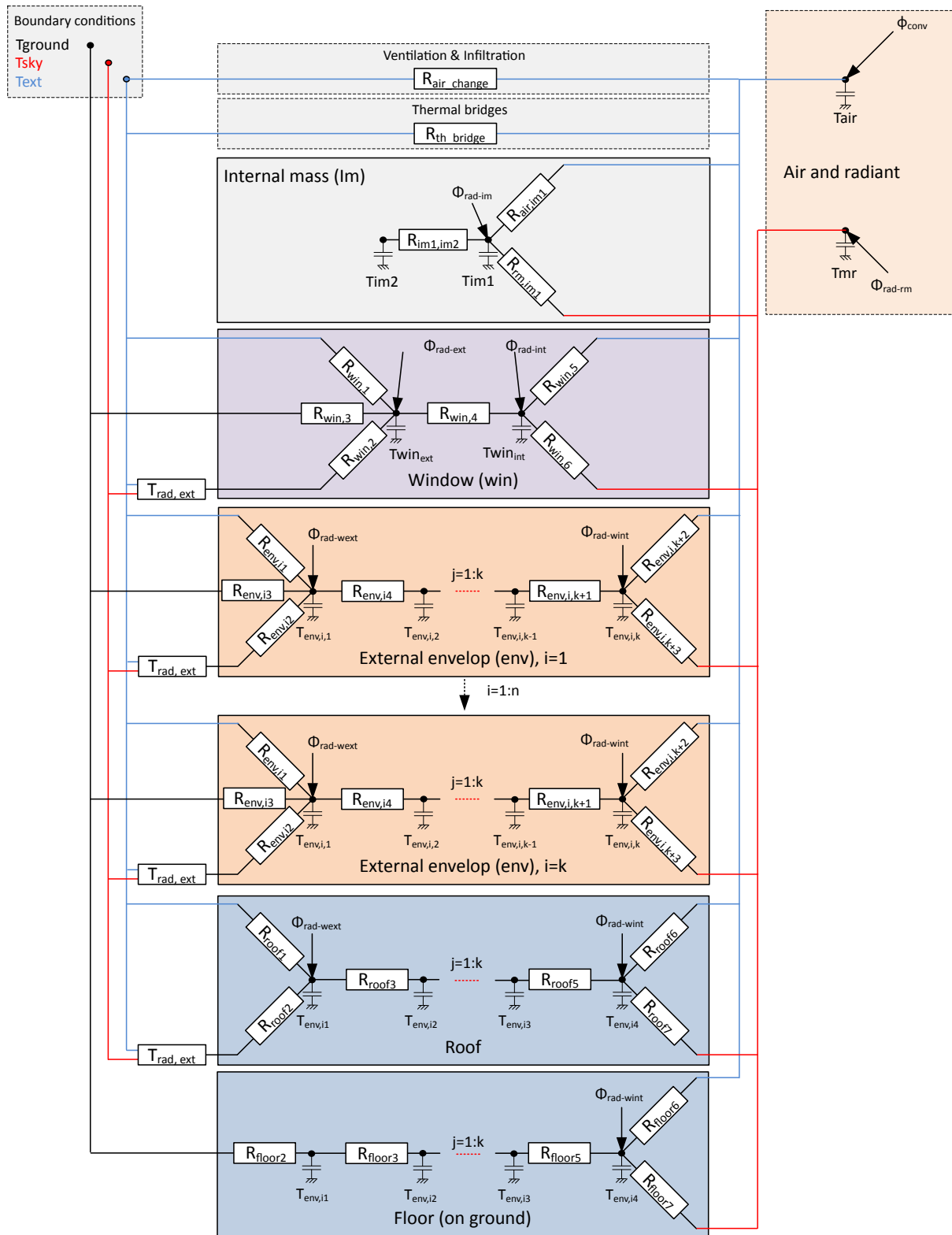


Figure 3.2: RiCj thermal zone model used in DIMOSIM [11].

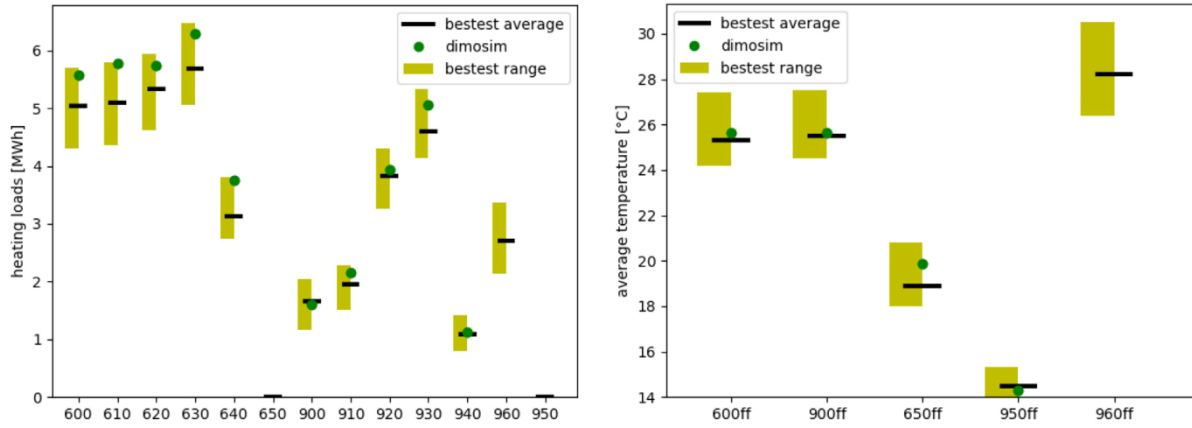


Figure 3.3: Results obtained by DIMOSIM on the BESTEST benchmark [11].

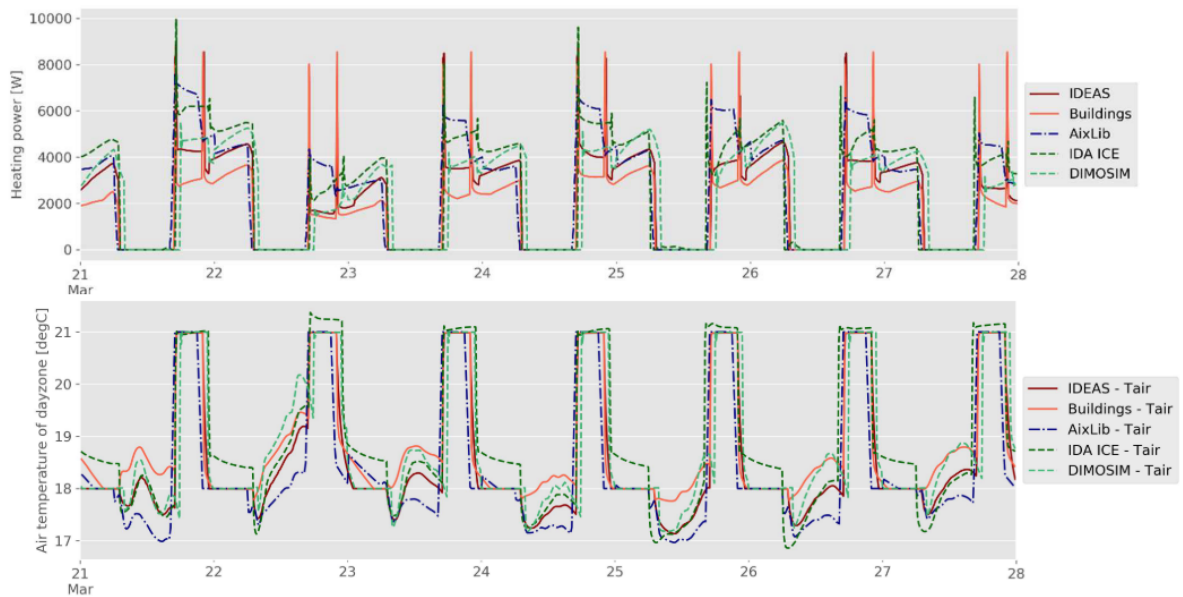


Figure 3.4: Results obtained by several UBE M on the DESTEST [8].

3.1.2 An enrichment process: Qiriél

In order to provide inputs parameters to DIMOSIM, an enrichment process must be used. In this thesis, we used a specific tool to do so: Qiriél [95]. Qiriél is presented more extensively hereafter and a procedure for its ad hoc validation is then discussed.

3.1.2.1 Presentation of Qiriél

During a UBEM simulation, the enrichment process consists in filling in missing variables. This can be reframed as a prediction problem where we try to predict a set of variables $[x_1, x_2, \dots, x_n]$ for each building of our sample. If we consider a single building, we would like to avoid the enrichment process to predict always the same values: as detailed in Chapter 1, such deterministic approach risks to lack the diversity of the building stock, and it could not account for the uncertainty introduced at this step. The aim is then to predict a probable set of variables $[x_1, x_2, \dots, x_n]$. To do so, the enrichment process should model the joint probability $P(x_1, x_2, \dots, x_n)$. Working with the joint probability ensures to properly model the variables relationships and to account for their interdependencies.

A joint probability can always be factorised as follows:

$$P(x_1, x_2, \dots, x_n) = P(x_1)P(x_2|x_1)\dots P(x_n|x_{n-1}, \dots, x_2, x_1) \quad (3.1)$$

With $P(A|B)$ being the conditional probability of A knowing B, as defined in the Bayes formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ with } P(B) \neq 0 \quad (3.2)$$

A Bayesian network [96] is a way to simplify Equation 3.1 by writing:

$$P(x_i|x_{i-1}, \dots, x_2, x_1) = P(x_i|A_i) \quad (3.3)$$

With A_i being a reduced number of parents.

Replacing $[x_{i-1}, \dots, x_2, x_1]$ by a subset A_i implies to make assumptions regarding variables independencies. These independency hypotheses should be carefully examined upon the construction of a Bayesian network.

A joint probability modelled with a Bayesian network can then be factorised as displayed in Equation 3.4.

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i|A_i) \quad (3.4)$$

This factorisation can be graphically represented through a directed acyclic graph. In this graph, each node is a variable and the edges display the connection of each variable x_i to its A_i parents.

An example of a Bayesian network is proposed in Figure 3.5.

It represents the joint probability $P(\text{HeatingEnergy}, \text{ConstructionYear}, \text{WallsUvalue})$ which

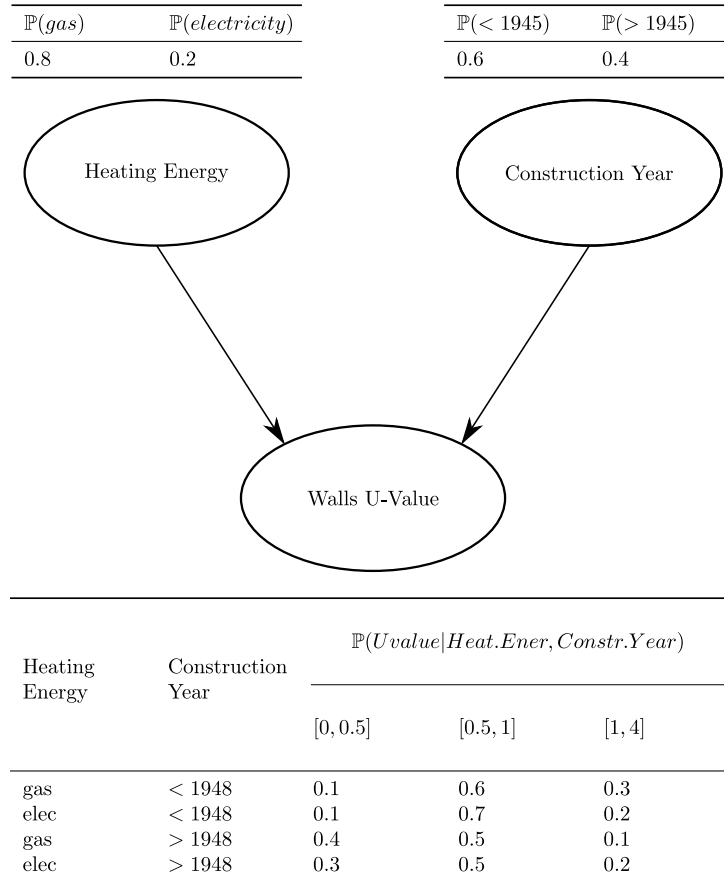


Figure 3.5: Example of a Bayesian network.

is factorised in this network as:

$$\begin{aligned}
 P(HeatingEnergy, ConstructionYear, WallsUvalue) = & \\
 & P(WallsUvalue|ConstructionYear, \\
 & HeatingEnergy) \\
 & \times P(ConstructionYear)P(HeatingEnergy)
 \end{aligned}
 \tag{3.5}$$

We assumed here that the variables *ConstructionYear* and *HeatingEnergy* are independent of each other, otherwise the term $P(ConstructionYear)$ would have been replaced by the term $P(ConstructionYear|HeatingEnergy)$. With the two variables being considered independent of each other, both terms are equal (from the definition provided in the Bayes formula 3.2).

Such modelling of the joint probability enables to easily adjust the probabilities to the available knowledge we have on each building. With the example network, if we know that a building is heated with gas, the walls U value will be sampled from the probability distribution $P(WallsUvalue|ConstructionYear, HeatingEnergy = gas)$ which is different from $P(WallsUvalue|ConstructionYear, HeatingEnergy)$. Taking into account the available information should then reduce the prediction error at the building level.

The training of a Bayesian network consists then in discovering the structure of the network (i.e. making hypotheses on the variables' independency) and in computing the probability tables depicted in Figure 3.5. Once the network has been trained, a sampling algorithm must be used to retrieve set of probables values from the modelled joint probability. To do so, it is common to use a Monte Carlo Markov Chain algorithm such as the Gibbs sampling algorithm.

To summarize, Bayesian networks are a way to model joint probabilities and to offer a graphical

representation of these joint probabilities. It relies on a set of hypotheses regarding variables independencies. It was chosen to be used in Qiriel since it enables to:

- Represent the diversity of the building stock and account for uncertainty;
- Take into account the variables dependencies;
- Adjust the prediction to the available description of the building.

The reader interested in gaining a finer understanding of Bayesian networks is referred to the course note of Ermon and Kuleshov [96] and to the book of Koller and Friedman [97].

In our case, we used different datasets to train different networks. The structure discovery algorithm used to create a Bayesian network from a dataset is not detailed here, but several examples are provided in the references cited above. The training sets we used are:

- **Fichiers Détail Logement (FDL) - INSEE [98]:** The FDL are an extract of the French census located at the IRIS scale. The IRIS are an administrative division of the French territory. Each IRIS is at most the scale of a single city and generally includes around 2000 inhabitants. Each row of the database consists in the description of an anonymised dwelling for which the number of dwellings it represents is provided as well as its IRIS location. For each row, the following parameters are provided:
 - The construction period;
 - The heating energy;
 - The type of dwelling (collective housing or single houses);
 - The number of occupants;
 - The professional status of the occupants;
 - The status of the dwelling’s occupants (Owner or tenant);
 - The number of rooms of the dwelling;
 - The social housing status of the dwelling (whether it is or not part of the social housing stock);
 - The type of heating systems (collective or individual);
 - The surface of the dwelling provided as a range (e.g. between 30 and 40 m²).
- **PHEBUS - MTE [99]:** The PHEBUS database is the result of a survey led in 2013 during which 5000 households were interviewed and listed their electrical appliances along with their energy performances. Contrarily to the two other databases presented here, this dataset is not publicly available.
- **Energy Performance Certificates (EPC) - ADEME [35]:** The realisation of an EPC is mandatory since 2012 when renting or selling a dwelling. The aim of the EPC is to provide a framework to compare buildings thermal quality regardless of its occupancy. All these EPC have been gathered by ADEME and have recently become publicly available. Almost 10 million EPC’s have been gathered so far, providing an interesting picture of the French residential building stock. However, these EPC present several flaws.

First, until July 2021, two different methods were used to rank buildings erected respectively before and after 1948. Buildings built before 1948 were assessed on energy bills only,

while buildings constructed after 1948 were assessed through a regulatory thermal model. For buildings assessed through the regulatory thermal model, their thermal properties had to be reported by the diagnostician. These data provide then a valuable outlook of thermal properties for buildings built after 1948.

Second, the EPC being used as a selling argument, some of them are biased to increase the building's value. Major discrepancies can thus be found within a same EPC or between different EPC made on the same building.

This database is therefore interesting because of its size but is certainly not flawless and should be used cautiously. The following variables are used to train a Bayesian network dedicated to the prediction of buildings thermal properties:

- The period of construction;
- DHW and heating systems types;
- Ventilation system;
- Insulation position (indoor or outdoor);
- Floor U-value;
- Roof U-value;
- Walls U-value;
- Windows U-value;
- Windows-to-wall ratio;
- Building's inertia type.

3.1.2.2 Ad hoc validation of Qiriel

In this part, the focus is now set on the ad hoc validation of Qiriel. The aim is to propose a framework that could be adjusted to other enrichment processes. Following the path opened by Rosser et al. [87], a cross-validation approach was applied here. The cross-validation consists in taking a part of the dataset for the model training and to use the other part to test the model performances. Schetelat et al. [46] proposed statistical indicators in order to assess Bayesian networks performances, but these indicators could be used in a broader extent to assess any enrichment process. These indicators are:

- The log-loss;
- The Relative Information Gain (RIG);
- The loss of Mutual Information.

The log-loss The log-loss is a way to assess the performances of a probabilistic classifier, such as the Bayesian networks, when predicting single features. Its principle is to compare the probability of a predicted output to belong to a given category to the probability of the actual output to belong to the same category.

Its formula is provided Equation 3.6 with H being the cross-entropy function, $P(y_{obs})$ the probability of the actual output and $P(\hat{y})$ the probability of the predicted output.

$$\begin{aligned}
 H(P(y_{obs}), P(\hat{y})) &= \frac{1}{n} \sum_{i=1}^n -P(y_{obs}) \log P(\hat{y}_i) \\
 &= \frac{1}{n} \sum_{i=1}^n -\log P(\hat{y}_i = y_{obs,i})
 \end{aligned}
 \tag{3.6}$$

The proposed simplification in Equation 3.6 is explained by the probability of the actual output which can be written as:

$$P(y_{obs} = C_k) = \begin{cases} 1 & \text{if } y_{obs} = C_k \\ 0 & \text{if } y_{obs} \neq C_k \end{cases} \quad (3.7)$$

When the probability to predict the right category is high, the log-loss is close to zero. If the Bayesian network was a perfect classifier, the log-loss would be equal to zero.

The Relative Information Gain The RIG is a way to compare the performances of the classifier to random chance. If the relative information gain is positive, the considered classifier retrieves more information from the provided inputs than random chance. This type of metrics is particularly interesting in our case since we would like predictions to be fitted to the information we already know on each building. The Relative Information Gain is defined as per Equation 3.8.

$$RIG = \frac{H(P(y)) - H(P(y_{obs}), P(\hat{y}))}{H(P(y))} \quad (3.8)$$

The loss of mutual information The mutual information (MI) is a mathematical tool to assess the relationship between two variables. It is quite similar to a correlation coefficient with the main difference that the correlation coefficient looks for a linear relationship between two variables whereas the mutual information looks for any kind of relationship. This main difference is illustrated Figure 3.6¹.

The mutual information of two variables X and Y is a measure of the difference between the joint distribution P(X,Y) and the factored distribution P(X)P(Y), these two distributions being equal if and only if X and Y are independent of each other. To compare two distributions p and q, we use the Kullback-Leibler divergence presented Equation 3.9.

$$\begin{aligned} D_{KL}(p||q) &= \sum_{j=1}^m p_j \log \frac{p_j}{q_j} \\ &= H(p, q) - H(p) \end{aligned} \quad (3.9)$$

The mutual information of two variables X and Y is then defined as :

$$MI(x_i, y_j) = D_{KL}(P(x_i, y_j)||P(y_i) \times P(y_j)) \quad (3.10)$$

The mutual information is then used to assess how well the Bayesian network represents the relationships among the variables. To do so, we compare the mutual information computed on the training dataset to the mutual information computed on the predicted data. A loss of mutual information indicates that the Bayesian network does not properly reproduce the correlation structure of the variables.

¹Since the mutual information is defined for categorical variables only, the proposed values of the normalised mutual information were computed after a discretisation of both variables. The value of the normalised mutual information could vary depending on the number of categories retained in the discretisation process.

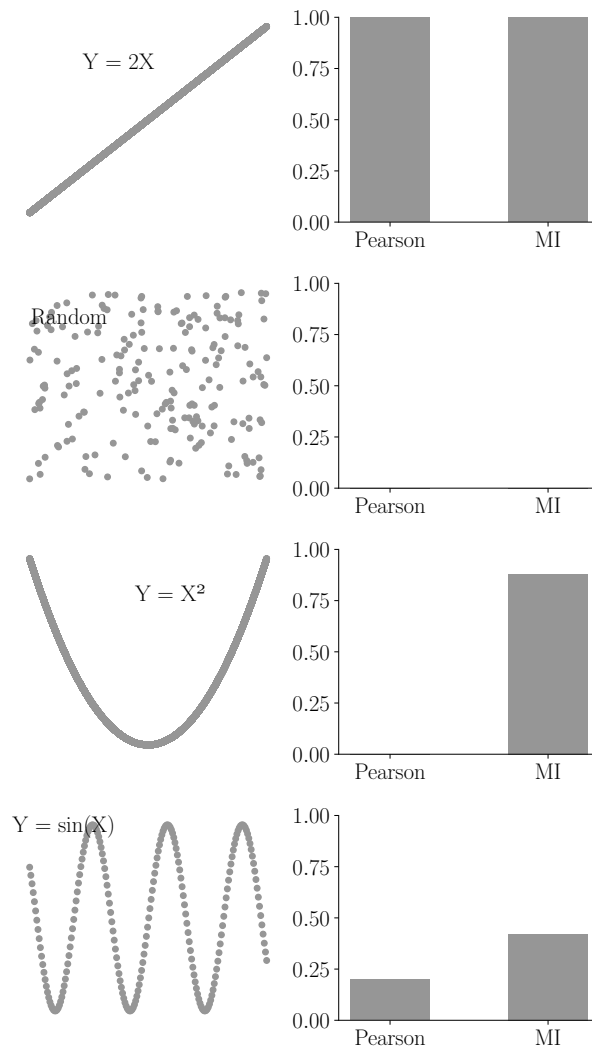


Figure 3.6: Comparison between the normalised mutual information and the Pearson coefficient.

Application to Qiriel As described previously, Qiriel is made of three Bayesian networks:

- One trained on the Fichiers Détail Logement (FDL) to predict the number of occupants per dwelling;
- One trained on Phebus to predict the number and the types of electrical appliances per dwelling;
- One trained on the Energy Performance Certificates (EPC) to predict the buildings thermal properties.

Each individual network should be assessed through cross-validation with the three indicators described above. This process is illustrated here on the network trained on EPC, but results obtained with the two other networks are provided in Appendix.

A dataset of 100,000 EPCs was sampled from the full database. To create this sample, we first selected EPCs from collective housing only since the rest of the work presented in this manuscript is applied solely to collective housing. As it will be described later, the main reason for this choice is the lack of energy consumption measurements for single houses due to privacy

concerns. We then reduced the database to EPCs for which there were no missing values since the graph structure discovery algorithm does not handle well missing values. Beyond these two first criteria, we finally sampled EPCs to respect the building stock distributions regarding three parameters:

- The distribution of the construction years;
- The distribution of the heating energy carriers;
- The geographical distribution of buildings.

Among these 100,000 EPCs, 80,000 were sampled to train the Bayesian network and the other 20,000 were used subsequently to test the network. This process was repeated five times in total, with different training and test sets sampled each time. It enables to study the variation of the log-loss and the RIG when using different (though similar) training sets. This practice is known as 5-fold cross-validation.

The graph structure obtained after the training is depicted in Figure 3.7.

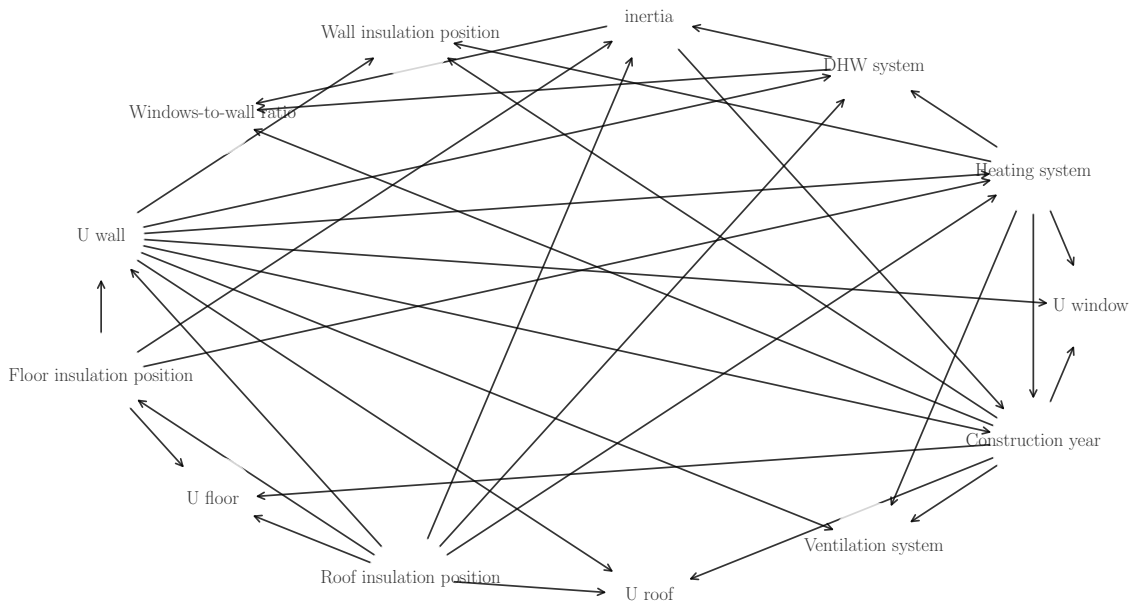


Figure 3.7: Bayesian network obtained by training on EPCs.

When testing the network, we kept the construction year as evidence, and we predicted all the other features.

The log-loss and relative information gain obtained for each feature are presented in Table 3.1 and the loss of mutual information is depicted Figure 3.8.

When considering the RIG obtained for each feature, the overall result is satisfactory. Two features (Roof insulation and Inertia) have a mean value standing slightly below zero which means that for these features, the network does not retrieve well information from the evidences that can be provided. Luckily, the other features display a RIG above zero, thus illustrating the interest of working with such algorithm instead of a simpler approach.

A validation framework is often expected to provide threshold value for this kind of indicators as it is the case with the BESTEST for instance. This kind of threshold should be fixed through

	Log-loss	RIG
Heating system	1.99 (± 0.16)	0.02 (± 0.09)
DHW system	1.65 (± 0.03)	-0.01 (± 0.02)
Ventilation system	1.20 (± 0.04)	0.24 (± 0.03)
Inertia	1.56 (± 0.03)	0.01 (± 0.03)
Wall insulation position	0.95 (± 0.08)	0.23 (± 0.08)
Floor insulation position	1.19 (± 0.05)	0.04 (± 0.05)
Roof insulation position	1.26 (± 0.03)	-0.01 (± 0.03)
U wall	1.80 (± 0.06)	0.22 (± 0.03)
U floor	0.76 (± 0.02)	0.07 (± 0.04)
U roof	0.82 (± 0.03)	0.02 (± 0.06)
U window	1.85 (± 0.07)	0.04 (± 0.04)
Windows-to-wall ratio	2.34 (± 0.02)	0.00 (± 0.00)

Table 3.1: Single feature performances assessment.

a workgroup dedicated to the topic. For the moment, a temporary threshold is arbitrarily set with the network being considered compliant if the RIG is above zero for the majority of the features.

The performances of the network at the features scale can also be illustrated thanks to the marginal distributions of these features as shown Figure 3.9.

On this figure, the marginal distributions of the features in the training set are depicted in blue, the distributions as encoded in the network are depicted in orange and the distributions obtained by sampling the network are in green. There is an overall agreement between the three marginal distributions, although a discrepancy can be observed between the distribution of the building's inertia obtained via sampling and the two other distributions.

This discrepancy is a default linked to the sampler algorithm used to retrieve probable values from the network since it is not visible in the network distribution (i.e. the tables of probability encoded during the training of the network) which is depicted in orange. In this case, we used the Gibbs sampling algorithm [13], which is a well-known Monte Carlo Markov Chain algorithm. However, the error noticed on the building's inertia should not jeopardize the rest of the simulations, and the performances at the feature scale are then deemed acceptable.

We can now turn our attention to the loss of mutual information presented Figure 3.8. In this figure, a positive value between two parameters indicates that the two parameters are modelled with a weaker relationship in the Bayesian network than what they actually have in the training set. On the contrary, a negative value corresponds to a relationship between two parameters that has been exacerbated in the Bayesian network.

In the case displayed Figure 3.8, the mutual information loss seems pretty low for most features. The highest loss is observed between the wall U-value and the construction year. This means that the relationship between these two variables is not entirely accounted for by the Bayesian network, although a link is visible in the network structure (Figure 3.7). As the walls U-value is also linked to many other parameters, the specific influence of the construction year may be depreciated by the other variables. Since the mutual information loss is not very high even in this case, we consider the network as suitable for the rest of the study. Similarly to what was done for the RIG, an arbitrary threshold is proposed for the mutual information loss at 0.1 in absolute value.

As a conclusion, this cross-validation procedure has thus ensured that results provided by Qiri

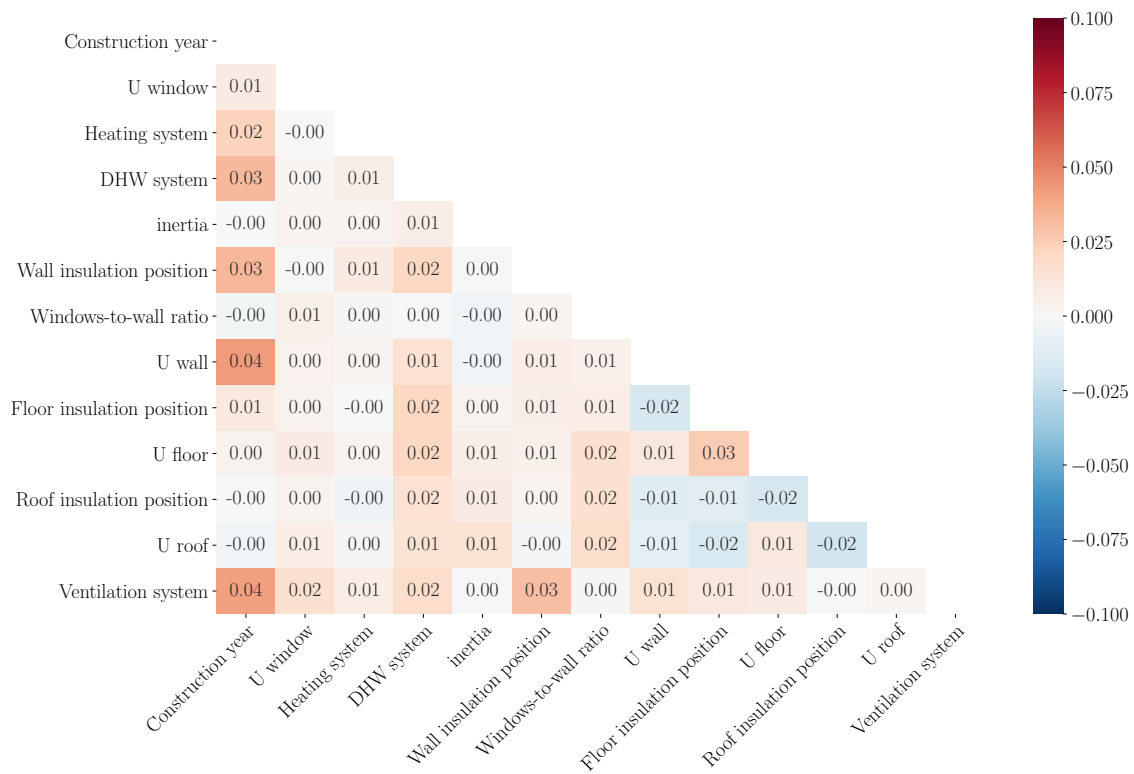


Figure 3.8: Mutual Information loss.

are consistent with the available data, ensuring that there is no major issues with the used enrichment process.



Figure 3.9: Marginals distributions of the training set, the trained network and of a sample drawn from the network.

3.2 Choice of the use cases

3.2.1 Lessons learnt from a first test case on a French department

During the study of a first test case, the annual electricity and gas consumptions were computed for residential buildings of a hundred districts with two UBEM tools, DIMOSIM and Smart-E [29]. Results have been presented in a paper [100] presented at uSIM 2020, and which is provided in Appendix. The computed consumptions were compared to actual consumptions measured at the district scale².

The consumptions simulated by DIMOSIM and Smart-E are depicted against the measured consumptions in Figure 3.10.

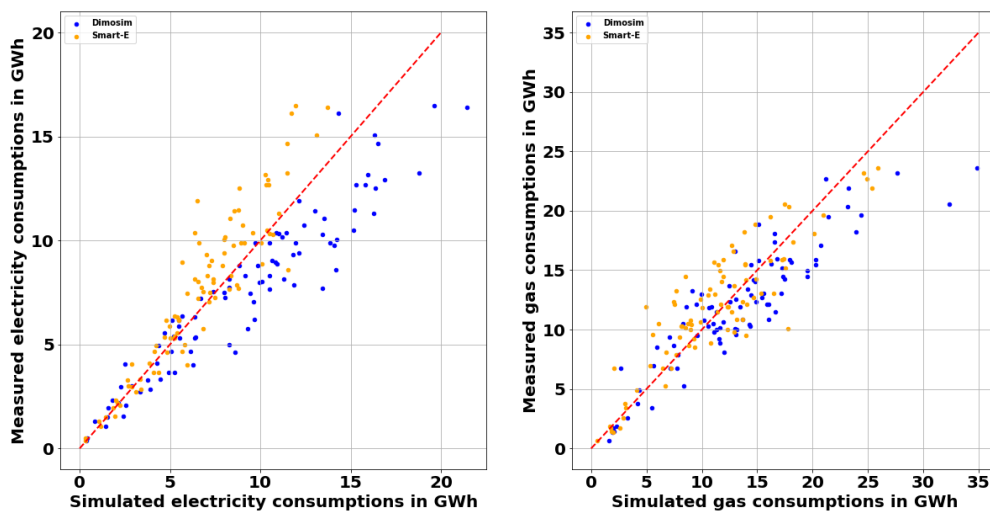


Figure 3.10: Annual measured and simulated consumptions of gas and electricity obtained for 100 districts.

It appears that on this test case, DIMOSIM overestimated the consumption while Smart-E tended to underestimate it. We tried to explain the error with global factors characterizing the districts such as the percentage of buildings heated with gas or the total residential floor area. However, when studying correlations between these factors and the error as depicted Figure 3.11, it was difficult to discriminate between the global factors to identify the most influencing ones and when some global factors stood out it was difficult to explain why.

The only information available at the building scale was then the building geometry and construction year. The rest of the parameters were inferred through the enrichment process.

From this experience, we learned two lessons.

First, we drew the conclusion that having information at the building scale in the validation dataset is of the utmost importance in order to be able to explain the observed error. This point is crucial even when the UBEM tool may not be accurate at the building scale. Having information at the building scale enables to recompose aggregated information at any spatial

²When this PhD project started, there were not as much data publicly available in France as there are now. Thus, annual electricity and gas consumptions provided by the DLE were available only at district scale when it is now available for buildings containing ten dwellings or more. In the same trend, the energy performance certificates (EPC) database was not available by then.

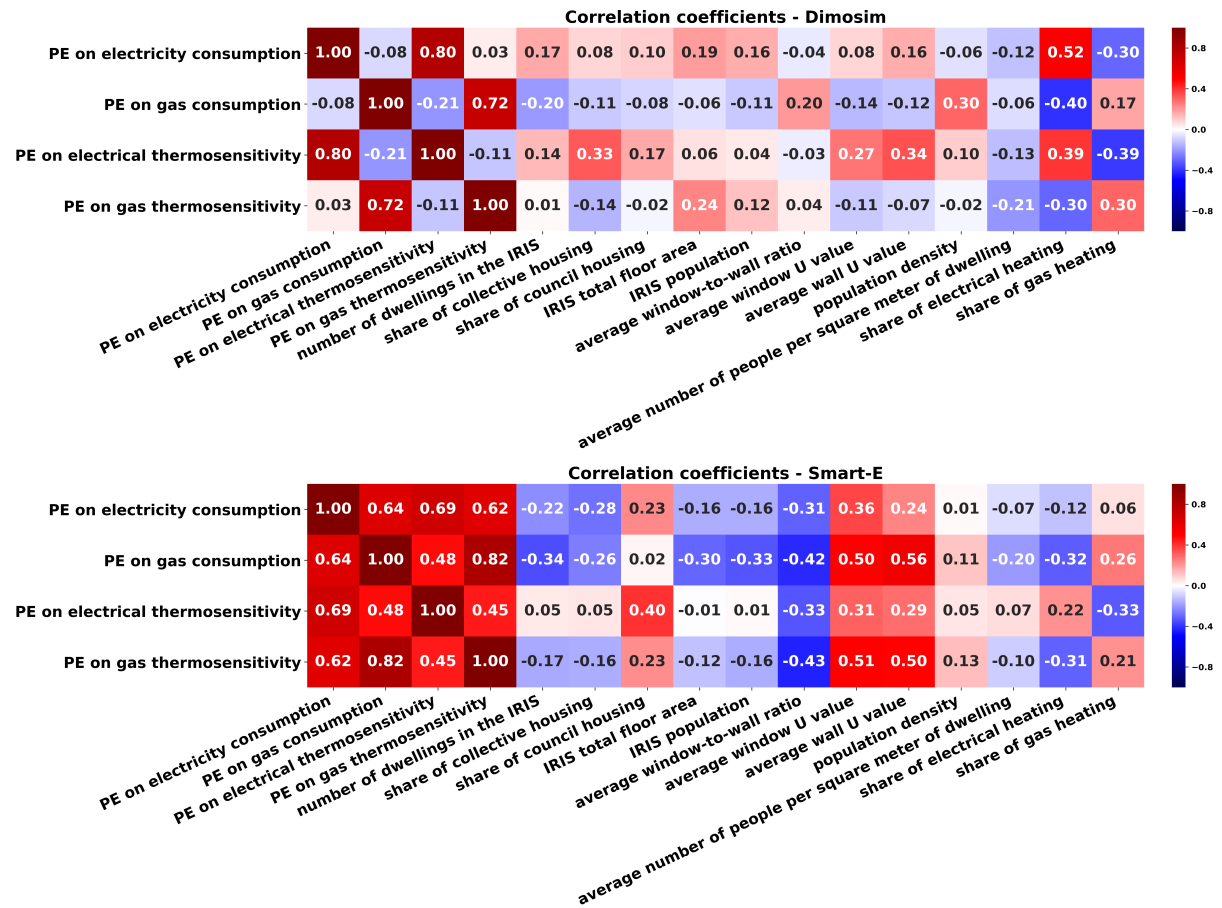


Figure 3.11: Correlation coefficients between the error and several input parameters.

scale.

Secondly, this first attempt taught us that uncertainty needed to be quantified. There are so many sources of uncertainty in a UBE simulation that if it is not properly accounted for, it will be all the more difficult to explain the observed error and conclusions could simply be pointless.

3.2.2 A choice driven by the data availability

The use cases selected to apply the proposed validation framework were chosen based, first and foremost, on data availability. For instance, the data available to describe the buildings thermal properties are almost inexistent for tertiary buildings whereas there is the EPC database to describe residential buildings. Generally speaking, there is a data shortage when it comes to understand energy consumption in tertiary buildings. For this reason, we decided to work only on residential buildings.

As discussed in Chapter 2, a validation dataset should be large enough to test the tools on most of their range of use and should provide with measurements taken on operating buildings. Furthermore, to ensure the replicability of the validation procedure, publicly available data should be used. In France, only the Données Locales de l'Énergie (DLE) matched all these criteria. The DLE provide the annual gas and electricity consumptions for French residential buildings containing at least ten dwellings. Such datasets are also available for industrial and tertiary buildings. The measurements are provided at the address scale, which means that a building is identified thanks to its address. Working on annual consumption to start the validation also seems wise as it is the simplest output that a UBE simulation tool can provide, and thus enables to treat

the basic use case of UBEM. Once this output is validated, the validation procedure could tackle use cases requiring outputs of increasing complexity.

The use cases treated in this manuscript were thus selected based on their associated KPI: the annual gas and electricity consumptions at the building scale. The corresponding use cases were identified:

- Identifying buildings to first retrofit in a district;
- Evaluating the energy saving of a group of buildings being retrofitted;
- Assessing the potential for the extension of a district heating network.

3.3 Creation of a validation dataset

A validation dataset was created based on available databases to assess UBEM performances on the chosen KPI and use cases. The databases that were used to create this validation dataset are presented hereafter. These databases were joined together in a national buildings database which is presented afterwards. Finally, an outlook on the buildings selected in the validation dataset is provided.

3.3.1 Available databases

3.3.1.1 BD TOPO®

The BD TOPO® [101] is a 2.5D database providing the location, the geometry and the height of every building in France. Relative to the CityGML framework [102], this database provides a LoD 1 of the French building stock. Its accuracy regarding the geometry and the height of each building varies depending on how the information was acquired (LiDAR data, orthophotography) but is generally comprised between 1m and 2.5m. These uncertainties due to a lack of accuracy add up to other uncertainties related to the definition of a building. For instance, the building footprint given by the BD TOPO® often comprises the balcony and the dependences. However, this database enables to retrieve the buildings' layout of a district or a city and thus to identify the adjacencies between buildings as well as the solar masks.

This database is updated via aerial photographs gradually covering the whole French territory, so that each land plot is updated every three to four years.

3.3.1.2 Fichiers Fonciers

The Fichiers Fonciers (FF) [103] are French tax records related to buildings. They are associated to an administrative division of the French territory, the administrative plots, each tax file being located in one administrative plot. It contains the following information regarding the buildings that are located within the perimeter of the plot:

- Total floor area,
- Construction year,
- Number of dwellings,
- Number of dwellings that are part of social housing,

- Number of dwellings that are empty,
- Walls materials,
- Roof material.

These data are not publicly disclosed but can be obtained for research purposes. It is the only data source that provides the construction year of every building in France. Among all the databases described here, the Fichiers Fonciers are often deemed to be the most reliable and are generally used as ground truth.

3.3.1.3 Données Locales de l'Énergie

The Données Locales de l'Énergie (DLE) are annual electricity and gas consumptions that are provided for residential buildings with more than ten dwellings and for tertiary and industrial buildings consuming more than 50MWh per year. To locate these buildings, a postal address is given. The DLE are the measurements against which we aim to validate the UBEM annual consumption. The public disclosure of these data has been enforced by law, which obliges every distribution system operator such as Enedis or GRDF to collect these data and to publish them. Since this obligation is quite recent at the building scale (it was previously mandatory to disclose these data at the IRIS scale), some distribution system operators struggle to provide the data. For instance, when considering gas consumption for residential buildings, only a very small number is provided compared to what is expected. A better national coverage of these data should progressively be achieved.

3.3.1.4 The Energy Performance Certificates

This database was already introduced in Section 3.1.2.1, but it is worth to mention here that each EPC is located thanks to the building address. Furthermore, during the audit, only surfaces in contact with the outdoor environment are reviewed. If an EPC is made in a dwelling which is located at an intermediate floor of a collective housing building, it will not contain any information on floor and roof thermal properties. To retrieve properties for a whole building, several EPC must often be grouped together.

For the creation of the validation dataset, we selected buildings for which it was possible to have information on all thermal properties. Having access to this level of detail should help the modeller with the error analysis during the validation process.

3.3.1.5 Joining these databases: a non-trivial problem

In order to create a validation dataset, we would like to join together all these databases in order to have all the information associated to the building geometry. However, there is no common identifier between these databases, thus making the joint difficult.

The BD TOPO® can be joined to the Fichiers Fonciers via a spatial joint between the buildings footprints and the administrative plots. However, when there are several buildings footprints within a single administrative plot, the association of the information to the right building geometry becomes very challenging. When it comes to join the Données Locales de l'Énergie and the EPCs, the difficulty level increases. In both databases, buildings are identified via their postal address. However, the postal address is not a straightforward way to identify a building. A postal address must first be translated to geographic coordinates through a step called "geocoding". Once geographic coordinates have been obtained, it may be located far from

the corresponding building. Examples of confusions that can occur is presented in Figure 3.12. In this picture, the numbers above the blue points are the real building address number that were manually checked thanks to Google StreetView whereas the numbers above the red points are what we get with the address geocoding. In this case, we can see that it seems to be working pretty well for buildings 5, 7 and 9 that are in the top left corner of the picture. However, if we look at the buildings 24, 26, 28 and 30 in the center and in the bottom right corner, it seems almost impossible to automatically associate the red points with the blue ones without any additional information.



Figure 3.12: Example of buildings addressing issues.

This issue of buildings addressing is quite well known within the building stock modelling community [104]. A way to solve this issue would be to create a unique and national building identifier as it already exists in other countries (a great example has been set by Switzerland [105]). Work is ongoing in France to follow the same path and a first trial should be proposed before the end of 2022³.

3.3.1.6 Base de Données Nationale des Bâtiments (BDNB): a step toward joining databases

An important work has been led by the CSTB within the GoRénove project⁴ to join the databases mentioned above. One of the goals of this project was to create a national database characterizing the existing residential buildings with as much information as possible at the building scale. This database was then to be used to give a first estimate of the energy savings and the financial gains that one could get with retrofitting a given building.

The team project put a lot of effort on improving the join of the previously mentioned databases, both spatially and at the address level. They developed a methodology for this join which is publicly available on a Gitlab project [106]. Most of the joined databases are open dataset so the part of the BDNB relying on these data is open too. However, information related to the Fichiers Fonciers is considered sensitive and should not be made public, but may be available for research purposes.

This PhD project closely followed the development of the BDNB, and it was then decided to sample the validation dataset used in this manuscript from the BDNB.

³<https://eig.etalab.gouv.fr/defis/batid/>

⁴<https://programmeprofeel.fr/projets/go-renove/>

3.3.2 Presentation of the obtained validation dataset

To create the validation dataset, we selected buildings from the BDNB for which a geometry was provided. We also restrained our dataset to residential buildings with no mixed usages. Working on purely residential buildings should reduce the number of assumptions made during the modelling process and thus help to identify where the error can come from at the diagnosis step.

Then, we selected buildings for which the geocoding step (i.e. the addressing of the building) was reliable. This constraint was set in order to avoid uncertainties due to information allocation on the wrong building.

After that, we retained buildings for which annual electricity consumption was available.

Finally, we kept buildings that had an EPC with all the thermal properties filled: this left only 1200 buildings for the validation dataset.

The number of remaining buildings in the BDNB at each step is presented on a Sankey diagram Figure 3.13.

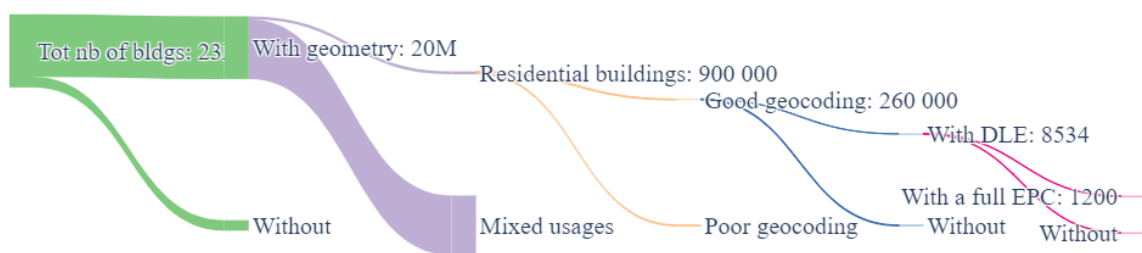


Figure 3.13: Sankey diagram of the buildings available to create a validation dataset.

The availability of the annual electricity consumption seems to be a major bottleneck in the proposed buildings' selection. This is not surprising since the availability of the DLE at the address level is recent and many buildings are still missing in the database. It is hoped that more data will be available in the future, which would remove this bottleneck.

Among the buildings that matched our criteria, we removed buildings with a height set to zero, and we removed buildings for which there was an unexplainable difference between the geometry provided by the BD TOPO® and the surface provided by the Fichiers Fonciers. It left us with 1200 buildings that were selected as the validation dataset.

These buildings are evenly spread across the French territory and the distribution of some parameters such as the construction year and the heating energy were checked and presented in Figure 4.

The annual electricity consumption distribution presents two peaks, the first one around $25kWh/m^2$ corresponds to the consumption of appliances for dwellings that are not heated with electricity whereas the second peak would be more likely dwellings heated with electricity. The same pattern is noticeable on the gas distribution, although it is less clear. In this case, it corresponds to dwellings using gas for cooking only versus dwellings using gas as heating energy.

When considering the distribution of the construction years obtained in the validation dataset

against the distribution provided by the Fichiers Fonciers, it can be noticed that there is a lower representation of recent buildings in our validation dataset. The influence of this difference should be accounted for when trying to scale the validation results to the building stock.

When considering the distribution of heating energies, it seems that a fair panel of each heating energy is represented in the validation dataset.

Regarding the geometry characteristics of the selected buildings, the validation dataset is characterised by an average total floor area of $1075m^2$ and an average height of $15m$. The median number of dwellings per building is 15 which corresponds to an average collective building of four to five storeys.

Conclusion

To conclude, we discussed in this chapter the ad hoc validation of a model, DIMOSIM, along with the ad hoc validation of a given enrichment process, Qiriel.

In the case of DIMOSIM, we used existing frameworks such as the BESTEST and the DESTEST to assess the model performances, but we also acknowledge the fact that the model has already been used in many studies and several theses as an endorsement of the model overall validity.

Regarding Qiriel, we presented this enrichment process built upon Bayesian networks, and we proposed a framework commonly used in the Machine Learning field and several metrics to assess its performances. This framework should enable to test other enrichment processes and to compare their strengths and weaknesses.

We also discussed the available databases and chose a set of use cases based on data availability. The retained outputs that will be assessed in the next chapter are then the annual electricity and gas consumptions at the building level.

Then, we created a validation dataset to assess these two outputs based on publicly available databases. The obtained validation dataset is made of 1200 collective housing buildings for which information regarding their thermal properties and their systems was available. Nonetheless, due to data availability, the obtained validation dataset shows a small though noticeable bias compared to the national building stock, and such bias should be kept in mind when studying the results of the validation procedure.

This validation dataset is used in the next chapter to assess the performances of several UBEM tools.

Geographical distribution across France of the buildings selected in the validation dataset

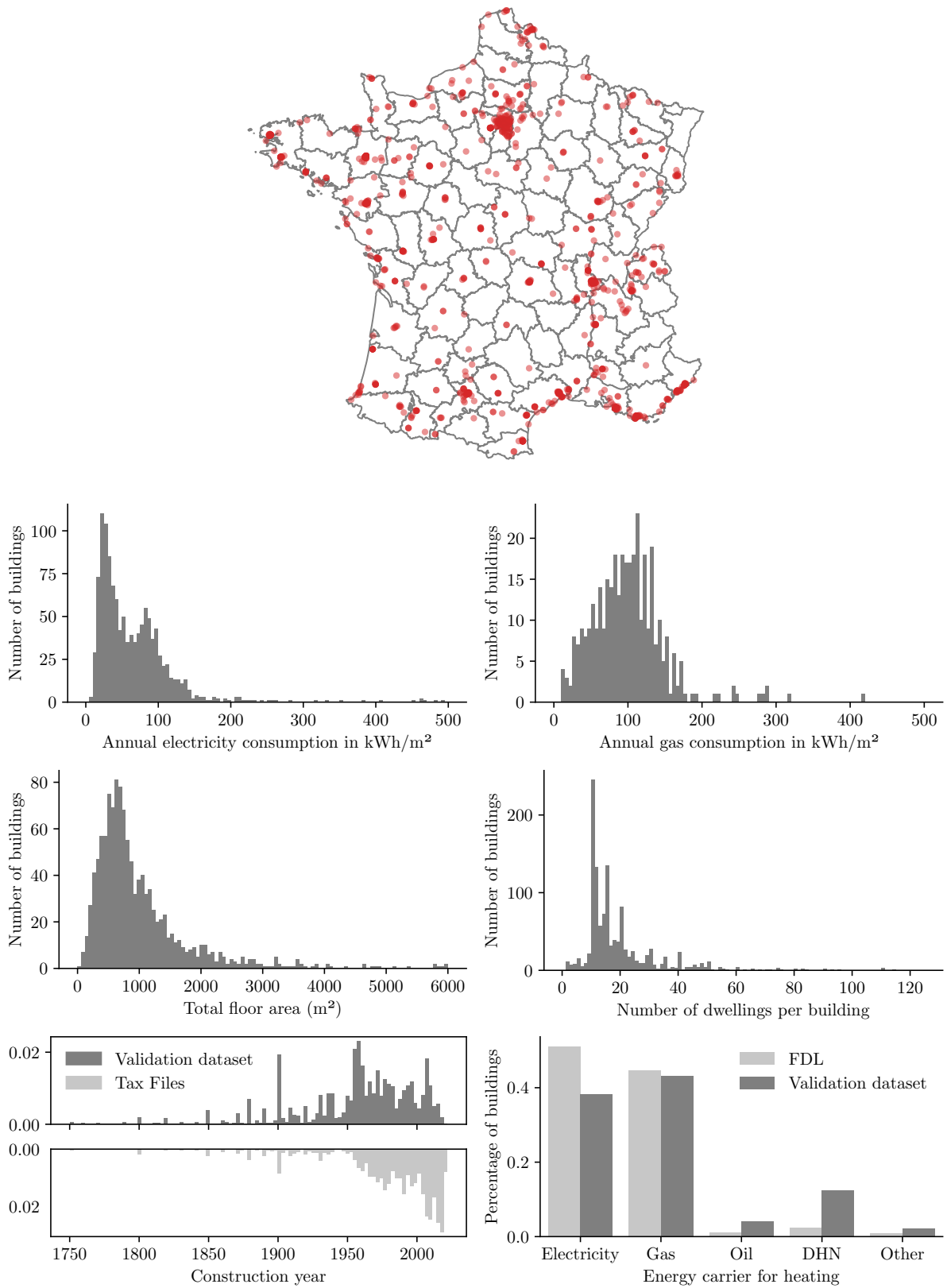


Figure 3.14: Dashboard of presentation of the validation dataset.

Chapter 4

Evaluation: Assessing the fitness for purpose of UBEM

Contents

4.1	Quantifying the existing uncertainty	60
4.1.1	Introduction to the uncertainty quantification	60
4.1.2	Uncertainty quantification applied to UBEM	62
4.2	Applying the validation procedure to a simple UBEM	68
4.2.1	Simply, a simplistic physical model to introduce the proposed validation process	68
4.2.2	Applying the assessment step to Simply	71
4.3	Applying the validation procedure to other UBEM tools	84
4.3.1	Brief description of Calomel	84
4.3.2	Introducing the results on the test building	84
4.3.3	Introducing the results on the entire validation dataset	88

Introduction

In this chapter, the evaluation stage of the proposed validation framework is discussed more in depth.

As advocated in Chapter 2, the assessment of a tool should never be done without accounting for the uncertainty surrounding the tool. The quantification of the uncertainty related to the chosen use cases and to the validation dataset is then described extensively in a first section.

The propagation of the uncertainty is applied to a very simple UBEM, Simply, in the second section of this chapter. By doing so, the focus is set solely on the description of the assessment procedure and on the associated metrics.

Once the metrics have been introduced with Simply, the assessment procedure is applied to two other tools, Calomel and Dimosim in the last section.

4.1 Quantifying the existing uncertainty

As recalled above, the uncertainty surrounding UBEM must be accounted for during the validation process. The quantification of the uncertainty is the first and most critical step of an uncertainty analysis. At this step, the modeller must decide how to represent uncertainty and the hypotheses made here will have consequences on the observed results.

In this section, a brief introduction to uncertainty quantification is provided before applying it to UBEM and to the considered validation test case.

4.1.1 Introduction to the uncertainty quantification

As explained in Chapter 2, the uncertainty may be modelled using different mathematical theories, but the classical approach of probability theory has been retained for the work presented in this thesis. There exists also two types of uncertainty quantification, forward and inverse [64], but we focus solely here on forward uncertainty quantification since the inverse uncertainty quantification would not serve the process presented here. In the rest of this manuscript, uncertainty quantification refers strictly to forward uncertainty quantification.

Within this framework, the uncertainty quantification may be split into two main steps that are:

1. Allocating a probability density function (pdf) to each uncertain variable;
2. Sampling the probability density function.

To discuss the key concepts of these two steps, let us consider an example model, $Y = f(X)$ with $X = (x_1, x_2, x_3) \in \mathbb{R}^3$, on which one wishes to perform an uncertainty analysis. The uncertainty analysis consists in this case in defining a probability density function, p_X for X and then in propagating it through the model f in order to obtain the probability density function of Y , p_Y .

Allocating a probability density function The first step of the uncertainty quantification is then the definition of p_X .

To do so, the information available on X should be accounted for, but so should be the information available on Y . For instance, if Y is defined only on \mathbb{R}^+ , the definition of p_X should be made in order to avoid having $Y = f(X) < 0$.

Furthermore, the probability density function of X should be defined in such a way as to maximize the uncertainty on X to ensure that the uncertainty analysis leaves no blind spots.

A common practise in the uncertainty quantification field to yield a probability density function complying with the constraints mentioned above is to use the Maximum Entropy Principle [94]. This principle relies on the Shannon entropy, defined Equation 4.1, which was first introduced in the Information Theory field [107].

$$H(X) = - \sum_{i=1}^n -P(x_i) \log(P(x_i)) \text{ with } X \text{ a discrete random variable} \quad \text{Or} \quad (4.1)$$

$$H(X) = - \int_{\mathbb{R}^n} p_X(x) \log(p_X(x)) dx \text{ with } X \text{ a continuous random variable}$$

This mathematical function, when applied to p_X , provides a measure of the uncertainty of X . Thus, looking for the maximum of the Shannon entropy for p_X is equivalent to maximizing the uncertainty of X and, the constraints identified on X help restraining the definition space of p_X that is explored during the maximisation process.

When considering the definition of p_X with X being a random vector like in the proposed example, the dependencies among the variables should be assessed too. For instance, back to the proposed example, if x_1 and x_2 are dependent but x_3 is independent of them, the probability density function can be simplified as $p_X(X) = p_{x_1, x_2}(x_1, x_2) \cdot p_{x_3}(x_3)$. In this case, the probability density functions of (x_1, x_2) and of x_3 can be investigated separately.

When considering single independent variables, a review of the probability density functions obtained by the Maximum Entropy Principle for classical constraints is provided in the first supplement of the Guide in Metrology [93]. A usefull example is the case when the modeller only knows in which interval the variable lies. In this case, the probability density function should always be a uniform law defined on the given interval.

When considering dependent variables, the approach proposed in this manuscript is to use Bayesian networks to simplify joint probability distributions, as described in Chapter 3.

Sampling the probability density functions Once the probability density functions (for continuous random variables) or the probability distributions (for discrete random variables) have been defined, procedures to draw samples from them must be defined.

Most programming languages provide the user with a pseudo-random generator whose function is to generate purely random number between 0 and 1. Such a pseudo-random generator then provides samples from a uniform law defined on the interval $[0, 1]$. From these random numbers, random variables following almost any kind of probability law can be generated. This approach is presented as inverse-transform method by Rubinstein and Kroese [12] and use the cumulative distribution function associated to any random variable X . A cumulative distribution function (cdf) F_X of a random variable X is defined as:

$$F_X(x) = P(X \leq x) \quad (4.2)$$

For continuous random variables, the following relationship between the cumulative distribution function and the probability density function stands, as long as the cumulative distribution function is differentiable:

$$p_X(x) = \frac{dF_X(x)}{dx} \quad (4.3)$$

The cumulative distribution function of any random variable is non-decreasing and lies within the interval $[0, 1]$. These two properties enable to draw a random number, u , from a uniform law defined on $[0, 1]$ and to derive from there a sample x from the random variable X :

$$x = F_X^{-1}(u) \quad (4.4)$$

This method is illustrated Figure 4.1.

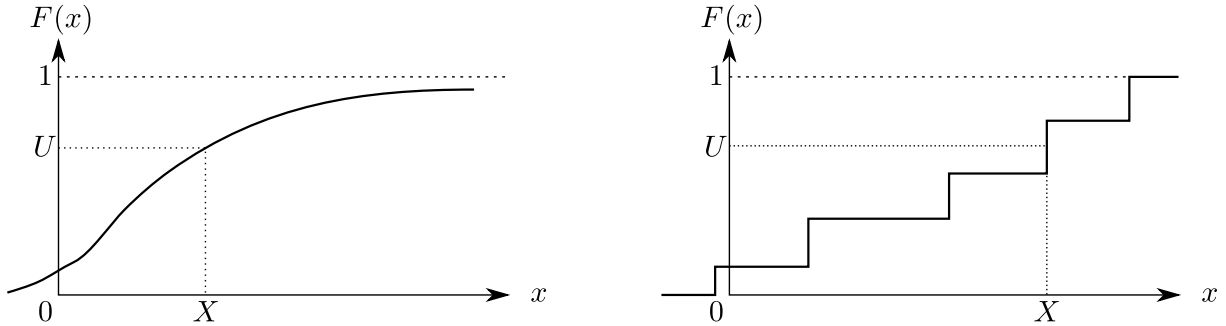


Figure 4.1: Inverse transform method applied to continuous and discrete random variables (reproduced from [12])

Randomly sampling from the interval $[0, 1]$ then enables to generate random sample from any distribution. Since the aim of the generated samples is to be used for uncertainty propagation, the modellers are often interested in exploring the inputs' probability distribution in the most efficient way in order to reduce the number of simulations. To do so, many sampling methods have been developed to efficiently explore the interval $[0, 1]$. A review of these methods is provided by Saltelli et al. [65] and includes:

- One-at-a-time sampling;
- Fractional factorial sampling;
- Latin Hypercube sampling;
- Multi-variate stratified sampling;
- Quasi-Random sampling.

In the rest of this manuscript, we use the Latin Hypercube sampling approach to draw realizations of independent variables. This choice was due to the large penetration of this sampling technic in the literature [38, 108, 109] and to its efficiency in exploring a probability distribution.

Regarding the realizations of dependent variables, it is common to use Markov Chain Monte Carlo algorithms to draw samples of joint distributions. The relationship between the predicted values is naturally taken into account by such processes. In the case of Bayesian networks, the Gibbs sampler, which belongs to the family of Markov Chain Monte Carlo algorithms, was chosen to retrieve realizations from the trained networks based on available information. A presentation of this algorithm is provided by Murphy in [13].

4.1.2 Uncertainty quantification applied to UBEM

In most physical models, uncertainty quantification consists in considering uncertainty for input parameters. In the case of UBEM, the uncertainty is located at two different steps of the simulation chain, as depicted Figure 4.2: a share of uncertainty originates from the available

data and another share of uncertainty comes from the enrichment process. The uncertainty on both available and enriched data must then be accounted for, although the two may be treated slightly differently. In the following of this section, the uncertainty related to the validation dataset introduced Chapter 3 is detailed and the uncertainty quantification for UBEM common inputs parameters is discussed.

4.1.2.1 Uncertainty related to the validation dataset

First, let us discuss here the uncertainty pertaining to the validation dataset. The validation dataset used here was made thanks to a join between five main sources. Uncertainty linked to each of these sources is discussed hereafter.

Fichiers Fonciers This database is deemed to be the most reliable one. Thus, there is no uncertainty considered on the information it provides. For instance, the construction year is deemed to be always true. This hypothesis could be questioned if more detailed data were available but since, to the author's knowledge, this is not the case, the Fichiers Fonciers are considered a reference when it comes to describe the French residential building stock. However, some data were too partial to be used. For instance, the walls material is often provided but with no indication regarding its thickness nor its insulation level, therefore a U-value cannot be retrieved from there, and this information was not used.

Weather data The weather data used for this validation procedure were retrieved from the closest weather station to each building. The weather at the precise location of the building may be rather different, namely due to urban heat island effect. Since the modelling of such phenomena is a complex research question on itself, the uncertainty on these weather data is not considered here. However, offering a way to easily take into account such uncertainty would be a great addition to the proposed validation procedure.

BD TOPO® The BD TOPO® provides the height and footprint of every building of the validation dataset. However, this geometry is precise only up to a certain point. The level of accuracy is provided in the BD TOPO® documentation [101] for both the height and the building footprint. It is quite likely that both uncertainties are linked to each other, but having no way to quantify this relationship, these data were considered independent of each other. The uncertainty on each parameter is then modelled through a uniform law on the interval [value – accuracy, value + accuracy] where the value is the building's value provided by the BD TOPO®.

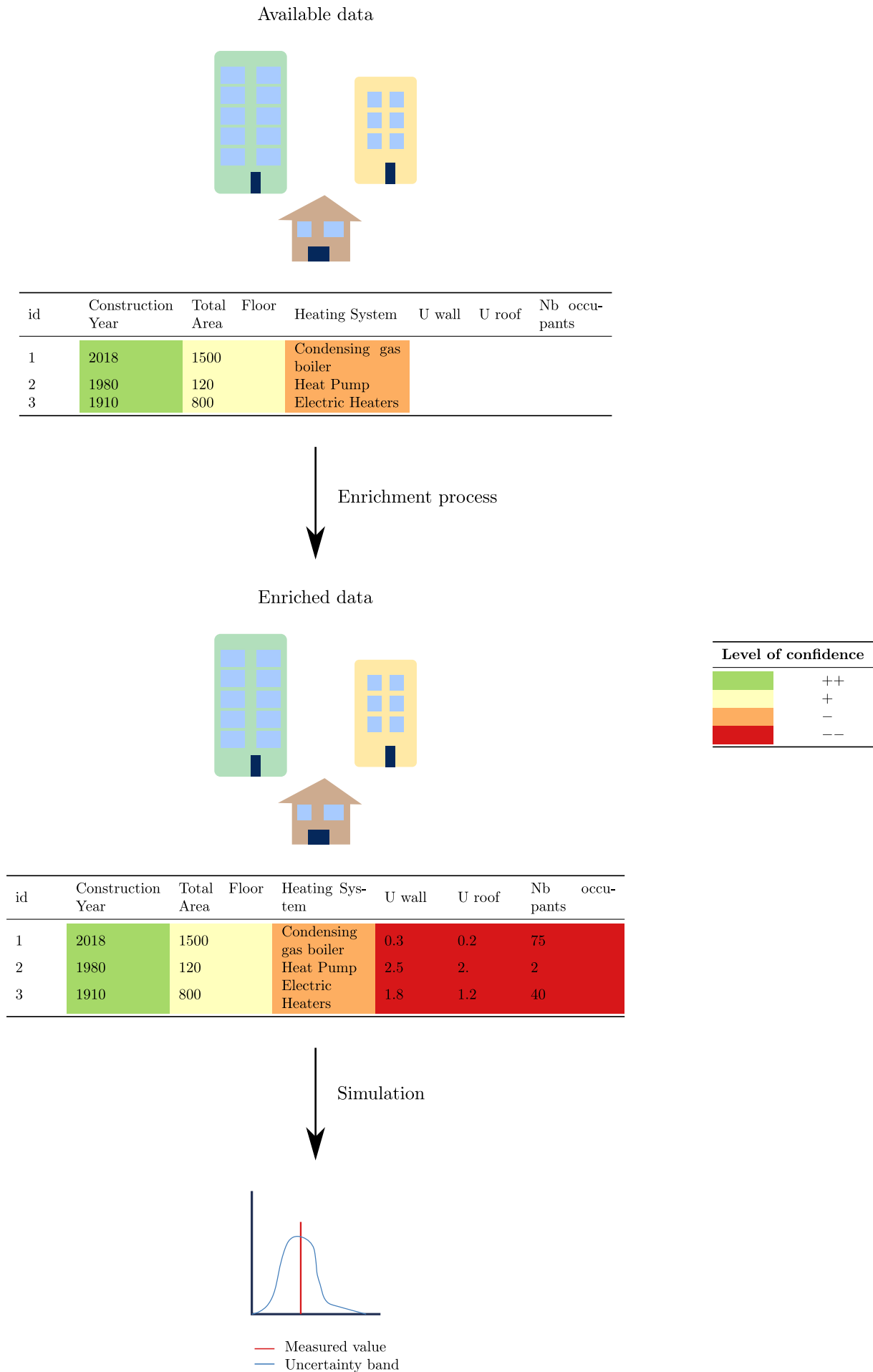


Figure 4.2: Locating uncertainty along the simulation chain

Energy Performance Certificates The Energy Performance Certificates (EPC) database consists in the gathering of 9 million EPC made since 2012. The quality of the provided information is questionable and varies depending on the considered variables. For instance, the heating and the DHW systems are often considered as almost reliable since these systems can be seen by the diagnostician who makes the EPC. However, the information may still be inaccurate due to a refurbishment and a change of the heating system after the diagnostician’s visit. Other parameters such as the U-value of the walls are often pure guesses or estimated based on an abacus since the diagnostician generally does not have access to detailed drawings of the building. The assumption made in this manuscript is that EPC may be very wrong when considered individually but still provide a good picture of the building stock thermal properties. For this reason, EPC were used to train a Bayesian network which is used in the enrichment process and to quantify uncertainty. To assess the influence of these data of questionable quality, simulations with different level of details retrieved from the EPC were made:

- Level 0: No information was retrieved from the EPC at the building scale;
- Level 1: The heating, DHW and ventilation systems are retrieved from the EPC of each building;
- Level 2: All information available in EPC are taken into account for each building.

In this chapter, we introduce results obtained for the simulation run with a Level 1 of data availability. The results obtained with Level 0 and Level 2 are discussed in Chapter 5.

Données Locales de l’Énergie The Données Locales de l’Énergie (DLE) are the measurements against which UBEM outputs are to be validated. These measurements are collected by electricity and gas distribution system operators thanks to smart meters, but may sometimes be rebuilt with their own models. Furthermore, the information is located at the address scale, allowing for error when it comes to allocate these consumptions to a building. However, as there was no better data available and no way to properly model the uncertainty on this database, the measurements were used as is, without further modelling of their uncertainty. The main hypothesis made here is that there is no major bias in this database.

4.1.2.2 Uncertainty quantification applied to common inputs parameters for UBEM tools

Once the uncertainty related to the databases used in the validation dataset has been assessed, the uncertainty on inputs parameters that are commonly required by UBEM tools can be quantified. This uncertainty quantification is split into two steps, following the existing steps of the simulation chain. First, uncertainty on inputs retrieved directly from data available on each building is quantified. Second, uncertainty on inputs obtained through the enrichment process is discussed.

Inputs retrieved directly from the available data The inputs retrieved directly from the available data are:

- **The roof area:** the roof area is retrieved from the building footprint which is provided by the BD TOPO®. Uncertainty on this value is computed in several steps. First, the BD TOPO® provides a value in meter for the accuracy of the footprint. Second, a value α is sampled from a uniform law defined on the interval $[-accuracy, +accuracy]$. Third, this

value α enables to obtain an error coefficient $\epsilon = \frac{\alpha}{\text{Longest edge of the footprint}}$ which scales up or down the footprint, and consequently the roof area.

- **The walls area:** The uncertainty associated to the walls area comes from two sources: the uncertainty on the footprint and the uncertainty on the height. The uncertainty on the footprint is considered as discussed above. The uncertainty on the height of the building is modelled through a uniform law defined on the interval $[\text{height} - \text{accuracy}, \text{height} + \text{accuracy}]$. Since the walls area is computed by multiplying the footprint perimeter by the building's height, the uncertainty is computed likewise.
- **The heating system:** In this chapter, we consider this piece of information as certain. The heating system of each building is provided by its EPC, but two common cases can lead to a wrong heating system:
 - The building underwent a refurbishment and the heating system has changed since the diagnostician's visit;
 - Different heating systems exist within the same building. This is especially the case for old buildings where parts of the dwellings are heated with individual gas boilers while the rest of the dwellings have electric heaters.

This type of uncertainty is not considered in this chapter where the simulations were run with the Level 1 of data availability. In the next chapter, uncertainty on heating system is taken into account through the Bayesian network trained on EPC.

- **The DHW system:** Same assumptions as for the heating system.
- **The outdoor air temperature:** This piece of information is retrieved from the closest weather station and, as discussed above, no uncertainty is considered yet on this parameter.
- **The solar gains:** As for the outdoor air temperature, no uncertainty is considered at the moment on the solar gains.

Inputs obtained through the enrichment process When considering the uncertainty quantification of inputs obtained through the enrichment process, two categories can be distinguished: the inputs that are deemed independent of the others and those that are not. Input parameters are considered independent of each other if the mutual information computed on existing data indicates no relationship or if there is not enough data available to quantify the relationship with the other parameters. The uncertainty quantification associated to the independent inputs is detailed first, followed by the uncertainty quantification of dependent inputs.

Uncertainty associated to independent inputs

- **The building heated volume:** To model the uncertainties of the heated volume, the uncertainty on the building geometry presented above is retrieved. Then, the uncertainty related to the occupation status of the building is quantified before applying a heated ratio on the occupied surfaces. The occupation status is provided by the Fichiers Fonciers: for each dwelling of the building, it says whether it is a vacation home, a primary residence or a vacancy. When a dwelling is vacant, it is considered not heated. When a dwelling is a primary residence, it is assumed to be fully occupied. In the case of a vacation home, the occupancy rate is computed thanks to the French Time of Use survey [110]. In this survey led in 2010, people filled in a diary with a time step of 10 minutes. When surveyed people had a vacation home, they were asked to provide with a rough estimation of the annual number of days they spent there. Once the annual occupation status of the building

is computed, a heated ratio is drawn from a uniform law defined on the interval $[0.4, 1.]$. This interval is arbitrarily chosen and this modelling of uncertainty could be challenged since the heated ratio is likely to be linked to the total floor area of the building. However, there was no available data to properly confirm this hypothesis.

Once the occupation rate and the heated ratio have been computed, these two ratios multiply and are applied to the whole geometry of the building as a scale factor. The building geometry is shrunk accordingly to the total heated ratio, and the physical model is then applied to this shrunken geometry.

- **The total floor area:** The uncertainty on the total floor area is obtained by combining the uncertainty on the building footprint area with the uncertainty on the number of floors since the total floor area is computed by multiplying these two values. The number of floors is obtained by dividing the height drawn when computing the uncertainty on the walls area by a floor height. This floor height is sampled from a uniform law defined on the interval $[2.5, 3.5]$.
- **The heating and DHW systems' efficiency:** The systems efficiencies are sampled from uniform law whose definition interval relied on the heating and DHW systems provided by the EPC. These intervals are presented in Table 4.1 and Table 4.2.

Heating system	System efficiency
Electric heater	$[0.99, 1.0]$
Standard gas boiler	$[0.75, 0.9]$
Air-air heat pump	$[2., 3.]$
Condensing gas boiler	$[0.85, 0.95]$
Air-water heat pump	$[2., 3.]$
Standard oil boiler	$[0.75, 0.9]$
Condensing oil boiler	$[0.85, 0.95]$

Table 4.1: Heating system's efficiency

DHW system	System efficiency
Electric water heater	$[0.7, 0.9]$
Gas boiler	$[0.75, 0.9]$
Oil boiler	$[0.75, 0.9]$
Thermodynamic water heater	$[2., 3.]$

Table 4.2: DHW system's efficiency

- **The air-change rate:** The air-change rate accounts here for the air changes due to both infiltration and ventilation. Its uncertainty is modelled with a lognormal law as advised by Nazaroff [111].
- **The heating set point temperature:** The heating set point is drawn from a database of measured indoor temperature. Since they are measured temperatures, the average temperature on the first thousand values of the annual profile was taken in order to assess the set point without being influenced by the inertia of the building in which the measure was taken. This database of measured indoor temperature was built during the PECOIC project [112]. One of the drawbacks of this database is that none of the profiles presents a lowering of the set point at night. Uncertainty on this practise should be accounted for in future works.

- **The DHW needs:** The PECOIC project [112] gathered DHW draw-off profiles from previous monitoring campaigns and provided draw-off profiles depending on the number of occupants of the dwelling. However, in order to use these profiles, the number of occupants of each dwelling needs to be assessed. To do so, Qiriel is used to sample a number of occupants from a Bayesian network. This Bayesian network is trained thanks to the Fichiers Détails Logements (FDL) [98] of the French census and takes into account the known dwelling's surface. Once the number of occupants has been drawn, a profile corresponding to the household size is sampled from the PECOIC database.
- **The appliances electricity consumption:** The uncertainty on the electricity consumption of appliances is modelled thanks to a monitoring campaign, Panel ElecDom [113], that followed the electricity consumption of appliances in a hundred French dwellings representative of France. Annual electricity consumption for appliances were provided in the final report, and a profile is drawn from this database for each simulation.
- **The internal gains:** In order to model the uncertainty on the internal gains, a ratio of conversion is applied to the electricity consumption of appliances. This ratio is drawn from a uniform law defined on the interval $[0.7, 1.2]$, the value above 1 enabling to account for the metabolic gains that are due to the occupants. This interval is set arbitrarily due to a lack of scientific publications to provide a better interval, but it could be challenged in future works.

Uncertainty associated to dependent inputs The inputs that are deemed dependent are:

- The walls U-value;
- The roof U-value;
- The windows U-value;
- The windows-to-wall ratio.

Uncertainty on these inputs is modelled through the Bayesian network trained on EPC that was introduced in Figure 3.7 in Chapter 3. Predictions of these inputs for a given building are sampled from the Bayesian network parameterized with the evidences gathered for this building. Thus, the samples are taken from a probability distribution conditioned to the information available on that building.

The hypotheses taken at this step of uncertainty quantification are summarized in Table 4.3. In this table, $\mathcal{U}([\cdot, \cdot])$ symbolizes a uniform law.

4.2 Applying the validation procedure to a simple UBLEM

4.2.1 Simply, a simplistic physical model to introduce the proposed validation process

In this chapter as well as in Chapter 5, we use a simplistic physical model to introduce the main concepts proposed in this validation framework. The aim is thus to set the focus on the proposed procedure and metrics before applying them to more complex UBLEM tools such as Dimosim.

This simplistic physical model is called Simply in the rest of this manuscript.

The model used to compute the heating needs in Simply is introduced Equation 4.5.

Inputs retrieved directly from the available data		
Variable	Source	Uncertainty modelling
Roof area	BD TOPO®	$Footprint \times \mathcal{U}([-accuracy; +accuracy])$
Walls area	BD TOPO®	$Footprint \times \mathcal{U}([-accuracy; +accuracy]) \times \mathcal{U}([height - accuracy; height + accuracy])$
Heating system	EPC	No uncertainty
DHW system	EPC	No uncertainty
Outdoor air temperature	Weather files	No uncertainty
Solar gains	Weather files	No uncertainty
Inputs retrieved from the enrichment process		
<i>Independent inputs</i>		
Variable	Uncertainty modelling	
Building heated volume	$\mathcal{U}([0.4, 1]) \times ratio_{occupation}$	
Total floor area	$\frac{height}{\mathcal{U}([2.5, 3.5])} \times Area_{Floor}$	
Heating system efficiency	Uniform law	
DHW system efficiency	Uniform law	
Air-change rate	Lognormal law	
Heating set point	Drawn from measured indoor temperature (PECOIC)	
DHW needs	Drawn from measured DHW consumptions (PECOIC)	
Appliances electricity consumption	Drawn from measured appliances consumptions (Panel ElecDom)	
Internal gains	Appliances electricity consumption $\times \mathcal{U}([0.7, 1.2])$	
<i>Dependent inputs</i>		
Walls U-value	Bayesian network trained on EPC	
Floor U-value		
Roof U-value		
Windows U-value		
Windows-to-walls ratio		

Table 4.3: Hypotheses retained for the uncertainty quantification

$$\varphi_{heat} = (T_{setpoint} - T_{outside}) \cdot U A_{global} - 0.3 \cdot \varphi_{solar} \cdot A_{window} - \varphi_{gains} \cdot A_{TotalFloor} \quad (4.5)$$

With the following parameters:

- φ_{heat} : The hourly heating needs;
- $T_{setpoint}$: The heating set point temperature;
- $T_{outside}$: The outside air temperature;
- $U A_{global} = U_{wall} \cdot A_{wall} + U_{roof} \cdot A_{roof} + U_{window} \cdot A_{window} + AirChangeRate \cdot V_{bldg} \cdot \rho_{air} \cdot (C_{p_{air}} \cdot 3600)$ where $AirChangeRate$ is in volume per hour, ρ_{air} is the air density in $kg.m^{-3}$ and $C_{p_{air}}$ is the specific heat of air in $J.kg^{-1}.K^{-1}$;
- φ_{solar} : The solar gains; in this case, it is simply the sum of direct and diffuse radiations to which was added the arbitrary coefficient 0.3 that accounts for the solar masks;
- A_{window} : The windows' area of the building;
- φ_{gains} : The internal gains; in this case, an arbitrary (although realistic) value in $W.m^{-2}$;
- $A_{TotalFloor}$: The total floor area of the building.

The heating needs are computed at an hourly time step for each building of the sample using Equation 4.5. Then, the total annual electricity and gas consumptions of each building are obtained by assigning an efficiency rate to each building's heating system. Other consumptions considered are those related to the domestic hot water (DHW) and to the electrical appliances. In both cases, an arbitrary value is provided as an input to the model in $W.m^{-2}.h^{-1}$.

The complete model of Simply is displayed Equation 4.6 for the electricity consumptions and Equation 4.7 for the gas consumptions.

$$\varphi_{tot_{elec}} = \delta_{heat_{elec}} \cdot \varphi_{heat} \cdot r_{heatingSystem} + \delta_{DHW_{elec}} \cdot \varphi_{DHW} \cdot r_{DHWsystem} + \varphi_{app} \quad (4.6)$$

$$\varphi_{tot_{gas}} = \delta_{heat_{gas}} \cdot \varphi_{heat} \cdot r_{heatingSystem} + \delta_{DHW_{gas}} \cdot \varphi_{DHW} \cdot r_{DHWsystem} \quad (4.7)$$

With the following parameters:

- $\varphi_{tot_{elec}}$: Total annual electricity consumption;
- $\varphi_{tot_{gas}}$: Total annual gas consumption;

$$\delta_{heat_{elec}} = \begin{cases} 1 & \text{if the building is heated with electricity} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_{heat_{gas}} = \begin{cases} 1 & \text{if the building is heated with gas} \\ 0 & \text{otherwise} \end{cases}$$

- φ_{heat} : Hourly heating needs;
- $r_{heatingSystem}$: Heating system efficiency;

$$\delta_{DHW_{elec}} = \begin{cases} 1 & \text{if the DHW is heated with electricity} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_{DHW_{gas}} = \begin{cases} 1 & \text{if the DHW is heated with gas} \\ 0 & \text{otherwise} \end{cases}$$

- φ_{DHW} : DHW needs;
- $r_{DHW_{system}}$: DHW system efficiency;
- φ_{app} : Electricity consumption of appliances.

The inputs required by Simply to compute the outputs assessed in the validation procedure are summarized Table 4.4.

	Walls U-value
	Roof U-value
	Windows U-value
	Air-change rate
Thermal properties and systems	Set point temperature
	Heating system
	Heating system efficiency
	DHW system
	DHW system efficiency
Building geometry	Roof area
	Windows-to-wall ratio
	Building heated volume
	Total floor area
	Walls area
Weather data	Outdoor air temperature
	Solar gains
Others	DHW needs
	Appliances electricity consumption
	Internal gains

Table 4.4: Simply input parameters

4.2.2 Applying the assessment step to Simply

Now that Simply has been introduced, it is used in this section to detail the assessment step of the validation framework.

Such assessment is made at two spatial scales:

- The building scale: How does the tool perform when considering the results building per building ?
- The sample scale: How does the tool behave when considering the entire validation dataset ?

We first present the metrics used to assess the performances at the building scale by applying them to a building retrieved from the validation dataset. Then, we define the metrics used to assess the performance of the tool at the sample scale.

4.2.2.1 Validation at the building scale

Presentation of the selected building The building selected to introduce the metrics at the building scale is located in the city of Rennes and is depicted Figure 4.3.

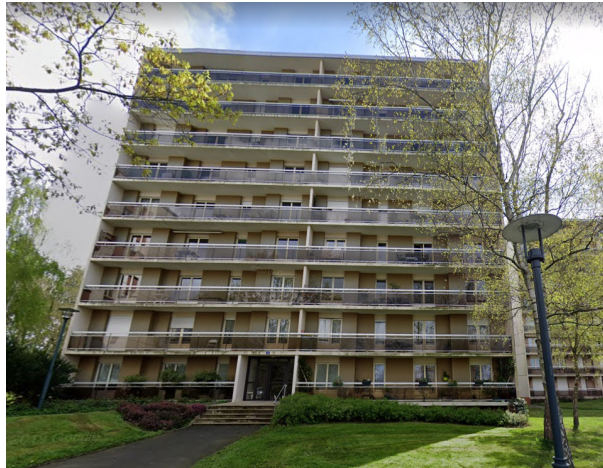


Figure 4.3: Test building randomly sampled from the validation dataset - Picture taken from Google Maps

The main properties of this test building are provided in Table 4.5.

Postal address	Construction year	Total floor area (m ²)	Number of dwellings	Heating & DHW system
4 Square Armand de la Rouërie 35700 Rennes	1982	2901	36	Standard gas boiler

Table 4.5: Main properties of the test building

Distributions of gas and electricity consumptions obtained for the test building The uncertainty quantification described Section 4.1 allowed the generation of one hundred sets of input parameters for the considered model. The annual gas and electricity consumptions obtained each time by Simply are presented Figure 4.4.

On this figure, the actual measurements of gas and electricity annual consumptions obtained thanks to the DLE are depicted in red. The distributions of consumptions obtained with Simply for both electricity and gas are not centred around these values, thus indicating a modelling bias of the tool for this given building.

The variations observed on the computed electricity and gas annual consumptions are due to the variation of the input parameters through the uncertainty quantification. An illustration of

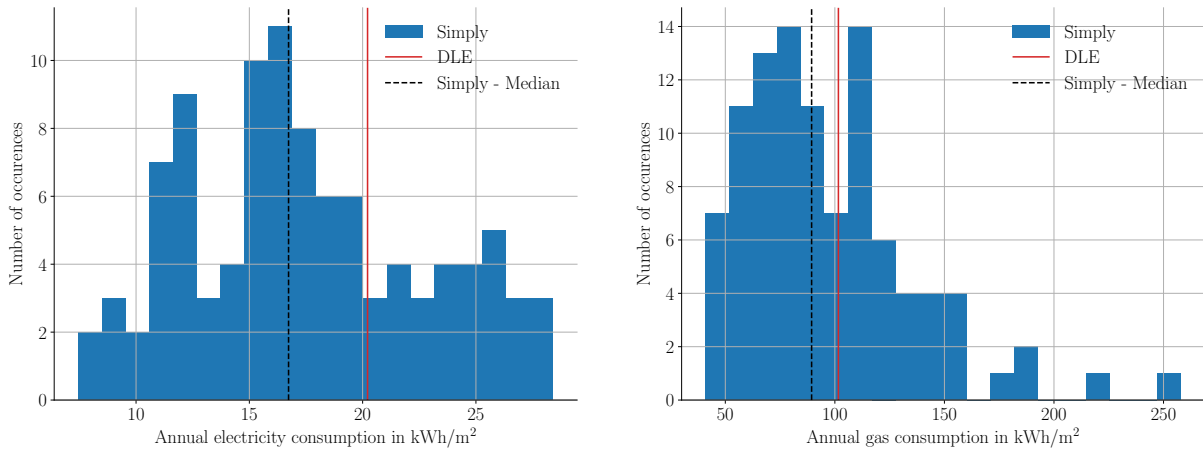


Figure 4.4: Distributions of the annual gas and electricity consumptions obtained by Simply on the test building

this phenomenon is provided in Figure 4.5. In this figure, the hypotheses that were made at the uncertainty quantification step should be visible on the marginal distributions of the parameters. If not, this reveals a problem in the numerical implementation of the uncertainty quantification.

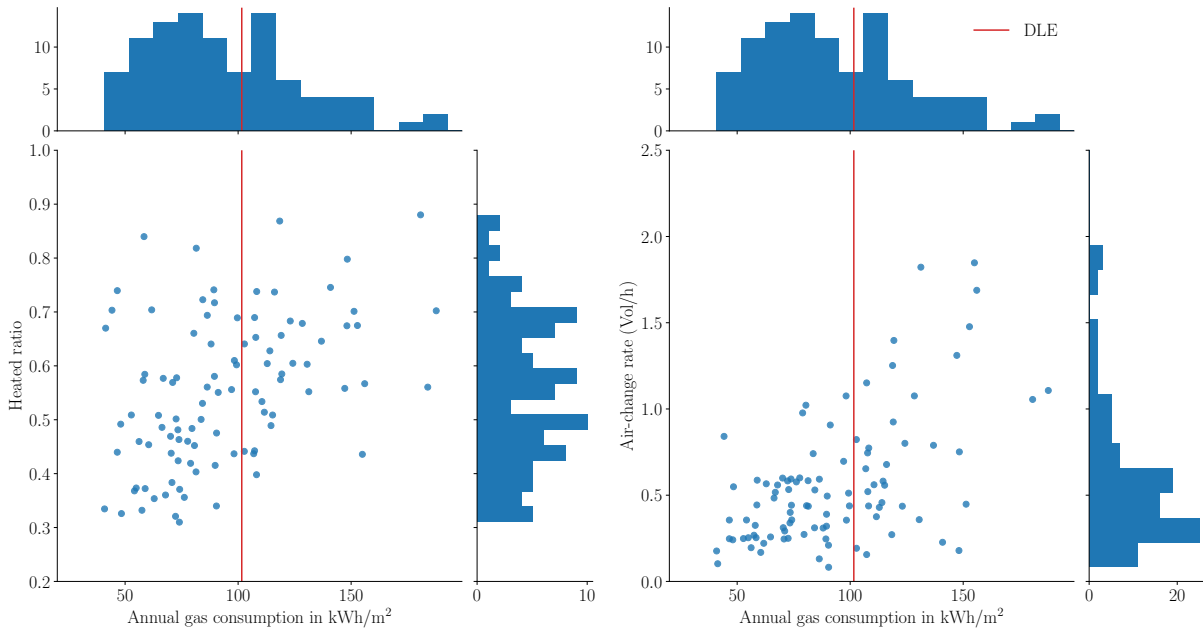


Figure 4.5: Variation of the annual gas consumption with respect to the heated ratio and to the air-change rate

This kind of representation when applied to all the uncertain parameters is a first step towards the diagnosis step of the validation procedure since it enables to identify uncertain variables that have a great impact on the computed consumptions. The diagnosis step is discussed more in depth in Chapter 5, but it is interesting at this step to provide this type of representations to enable the modeller to verify that the uncertainty quantification is not off grounds.

Metrics applied to the annual consumptions distributions Once the modeller has checked thanks to Figures 4.4 and 4.5 that the model behaves as expected, metrics to properly quantify its error should be defined.

First, let us define the Percentage of Error (PE) as per Equation 4.8.

$$PE(x) = \frac{x_{estimated} - x_{measured}}{x_{measured}} \cdot 100 \quad (4.8)$$

The distributions of the percentage of error obtained on both the gas and electricity annual consumptions are depicted Figure 4.6.

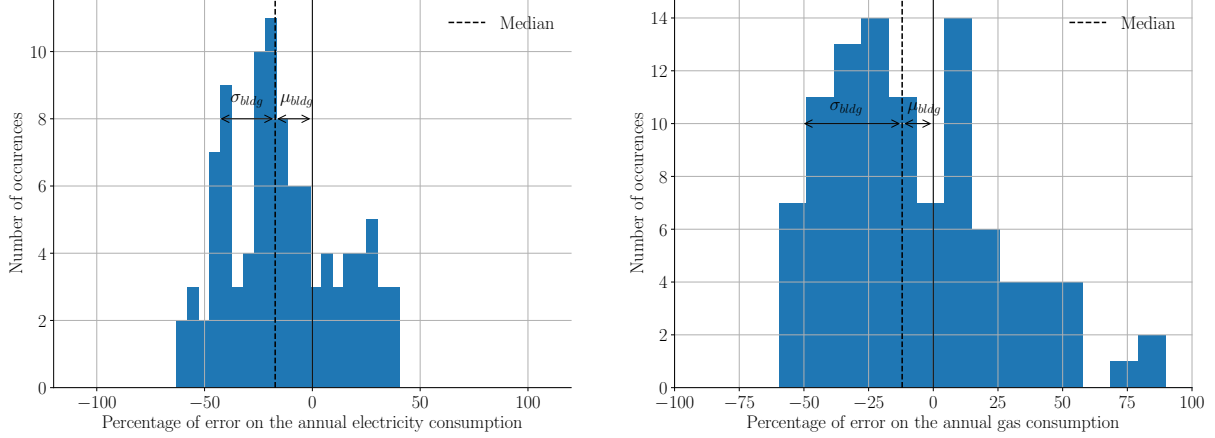


Figure 4.6: Distributions of the percentage of error obtained by Simply on the annual gas and electricity consumptions of the test building

We define then the **residual** obtained at the building scale, μ_{bldg} , as **the median of the distribution of the percentage of error**. Similarly, we define the standard deviation obtained at the building scale, σ_{bldg} , as per Equation 4.9.

$$\sigma_{bldg} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (PE_i - \mu_{bldg})^2} \quad (4.9)$$

with n the number of values in the distribution

When considering the annual electricity consumptions distribution obtained for the test building, Simply displays a residual $\mu_{bldg_{elec}}$ of -17% and a standard deviation, $\sigma_{bldg_{elec}}$, of 27%. Although a slightly better residual, $\mu_{bldg_{gas}}$, of -12% is obtained for the gas, the standard deviation, $\sigma_{bldg_{gas}}$, is higher (38%). This means that the distribution of the error on the gas consumptions is more centred than the distribution of the error on electricity consumptions, but it is also more spread out, thus having more configurations for which the error is very large.

4.2.2.2 Validation at the sample scale

Now that metrics have been defined to quantify the results at the building level, attention can be turned to the sample level.

The distribution of the annual electricity and gas consumptions obtained at the sample scale are depicted in Figure 4.7.

In this case, the distribution depicted for Simply is the distribution of the median consumption obtained on each building. When considering the distribution of the DLE for electricity, two groups of buildings can be identified:

- those heated with electricity;

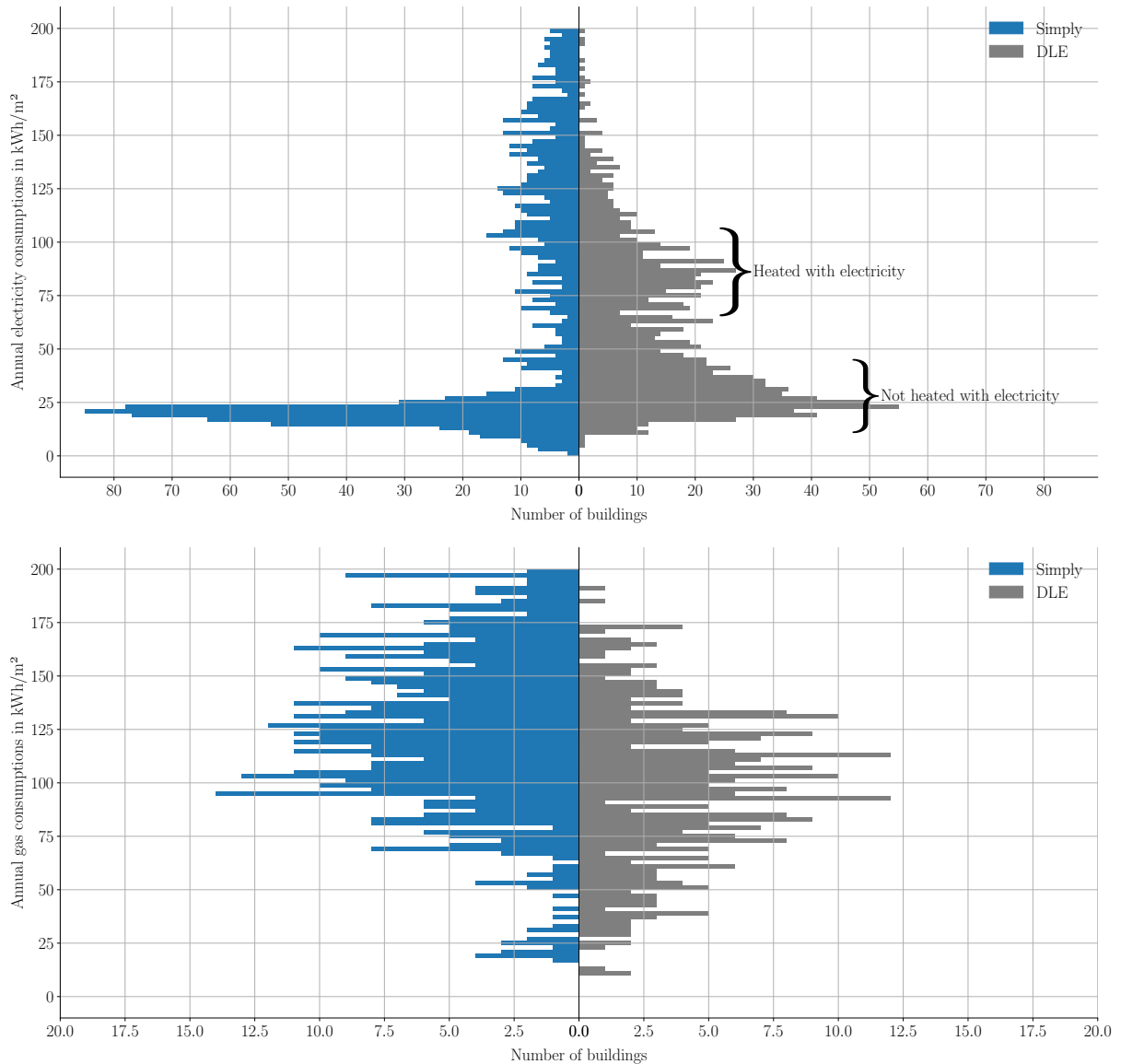


Figure 4.7: Distributions of the annual electricity and gas consumptions obtained at the sample scale

- those which consume electricity only for their appliances.

This trend is reproduced in a more pronounced way by Simply.

When considering the gas, it appears that Simply tends to compute higher gas consumptions than the DLE.

Bias and standard deviation at the sample scale Since a metric, μ_{bldg} , was defined to assess the residual at the building scale, the distribution of the residual obtained on each building of the sample is depicted for both gas and electricity on Figure 4.8.

Both distributions seem to be almost centred around zero, but to properly assess this **bias** a metric, μ_{sample} , is defined as **the median value of these residuals distributions**. A **standard deviation coefficient**, σ_{sample} , is also defined in the same way as σ_{bldg} . When considering electricity, $\mu_{sample_{elec}}$ is worth 5% and $\sigma_{sample_{elec}}$ is worth 105%. This means that the residual displayed by Simply on each building of the validation dataset has an almost unbiased

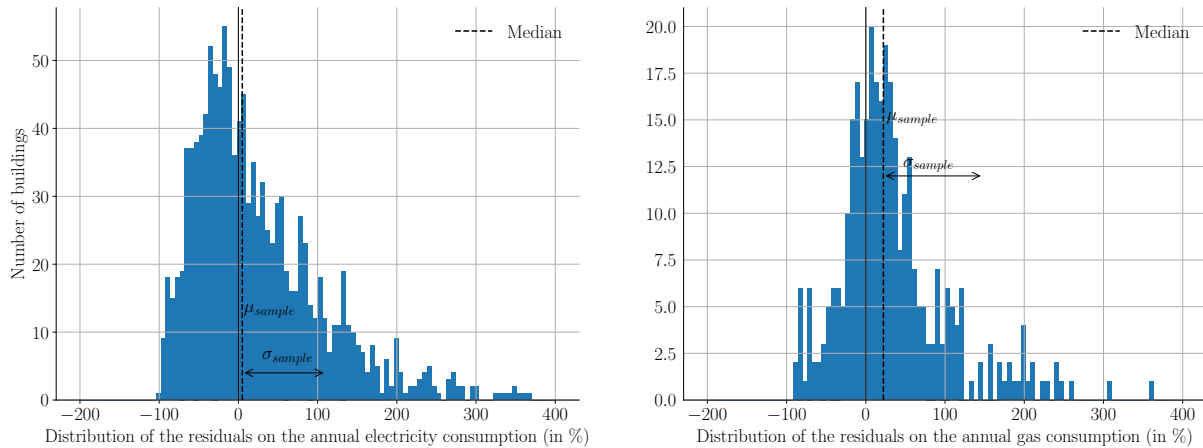


Figure 4.8: Distributions of the residual on the annual electricity and gas consumptions obtained at the sample scale

distribution, but this same distribution displays a very high standard deviation meaning that many buildings have a median percentage of error that is very far from zero.

Regarding gas, the bias, $\mu_{sample_{gas}}$, is much higher with a value of 22%, and the standard deviation, $\sigma_{sample_{gas}}$, with a value of 125%, is even higher than the standard deviation obtained for electricity. The highest value of $\mu_{sample_{gas}}$ indicates namely that Simply tends to overestimate the consumptions related to the heating and DHW needs. This corroborates the observation made on Figure 4.7.

The very high values displayed by both $\sigma_{sample_{gas}}$ and $\sigma_{sample_{elec}}$ indicate that the model adequacy to the buildings real behaviours can vary a lot depending on the selected buildings. The reasons for such behaviour are investigated at the diagnosis step in Chapter 5.

The HPD region and the empirical coverage The two metrics introduced above enable to give a good picture of the residuals distribution. However, these metrics do not allow a threshold to be easily set to create a category of buildings for which the model is considered wrong. Such threshold should be fixed while considering the level of confidence required by the use case.

Another metric is then introduced based on the Highest Posterior Density (HPD) region. This mathematical tool is comparable to a confidence interval and its use through the validation process is encouraged by Moser and Oliver [63]. A clear definition is provided by Murphy [13] and is depicted Figure 4.9. The main principle is that the distributions obtained after an uncertainty propagation can be regarded as probability densities. The highest posterior density region for a given level of confidence α computes then the region of these densities with the most probable points that gather $(1 - \alpha) \cdot 100$ of the total mass of the given pdf.

It is then possible to compute an HPD region for each building of the validation dataset. An example is provided in Figure 4.10 where the HPD region obtained for the test building studied in the previous part is displayed.

It can be noticed that, in this case, the measured consumptions of gas and electricity both lie in the HPD region computed for a confidence level of 95%. The HPD region enables then to set a threshold regarding the UBEM performance on a given building: the measurement is contained in the HPD region or the measurement is out of the HPD region.

Following this idea, an additional metric is proposed at the sample scale: the **empirical coverage**. The empirical coverage is simply the percentage of buildings of the validation dataset for

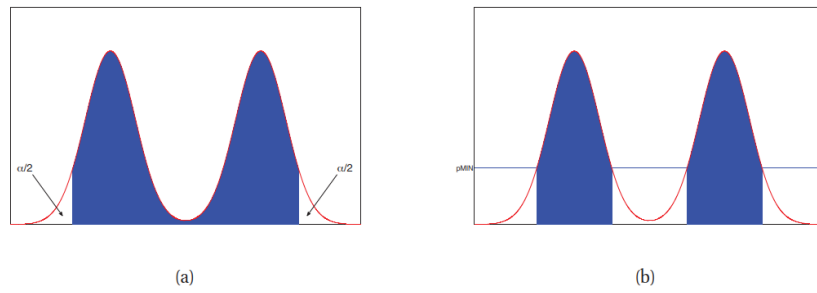


Figure 4.9: Difference between central interval (a) and HPD region (b) for a multimodal pdf - figure taken from Murphy [13]

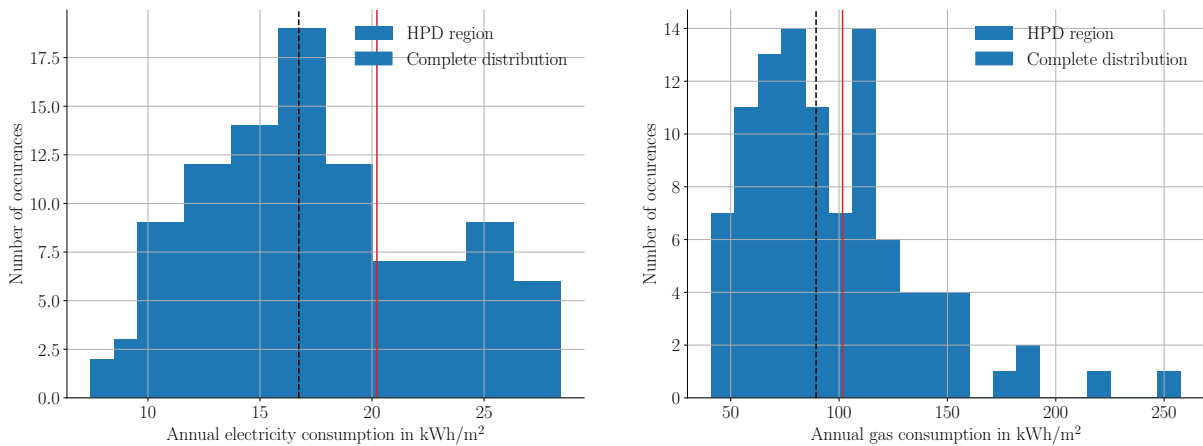


Figure 4.10: HPD region computed for a confidence level of 95% on the annual gas and electricity consumptions of the test building

which the measured value lies within the HPD region. For an HPD region computed for a level of confidence of 95%, the empirical coverage should be around 95%, that is to say that around 95% of the buildings should have their measured value within their HPD region. Normally, the computed value of the empirical coverage should always be equal or below the given level of confidence. Otherwise, it would reveal a misrepresentation of the uncertainty.

This metric enables then to assess how a tool behaves with regard to a given level of confidence.

In the case of Simply, it displays an empirical coverage of 61% for the electricity and of 82% for the gas for a level of confidence of 95%.

Figure 4.11 offers an illustration of the empirical coverage applied to the validation dataset.

On this figure, the median consumption obtained for each building along with its HPD region is displayed by a dot and a line. Points for which the measurements lie within the HPD region (the line intersects the first bisector) are set in blue whereas the buildings that do not comply with this requirement are set in red.

When taking a closer look at the results on electricity, two trends can be identified:

- A first group of points can be found below the first bisector with a common median value for the computed electricity consumption. These buildings are buildings using electricity for appliances only, and the common median value comes from the way uncertainty on

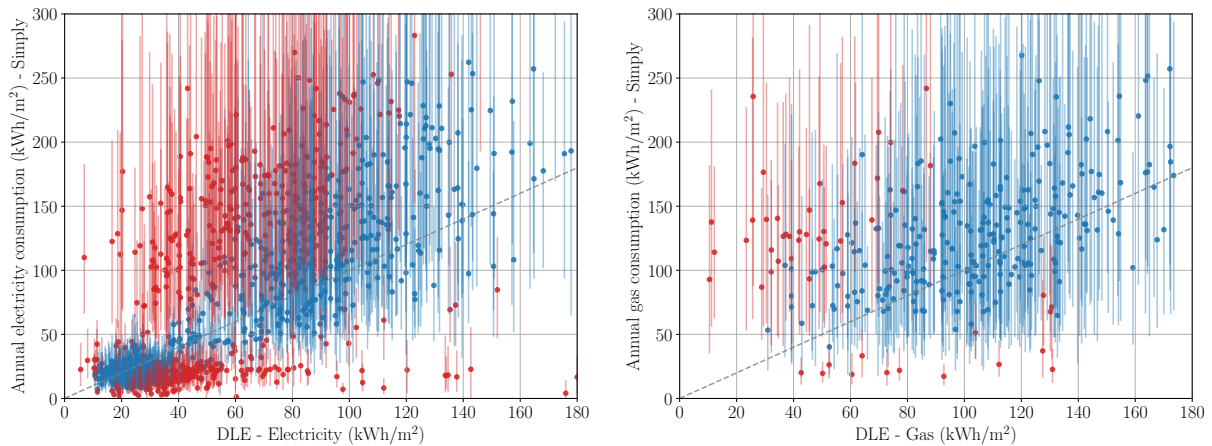


Figure 4.11: Distribution of the computed consumptions with a level of confidence of 95% in regard with the DLE

appliances consumptions was modelled (thanks to the Panel ElecDom survey [113]).

- A second group of points lie above the first bisector. These points surely identify to buildings heated with electricity. This would indicate that Simply overestimate the heating needs of the buildings.

In both cases, part of the error on these buildings could come from the fact that each building was considered having a single energy carrier for heating. Since it is not always the case as explained previously (dwellings of the same buildings can be heated with electricity and with gas), it would explain an overestimation of the electricity consumptions when building were considered heated with electricity only and an underestimation of the electricity consumption in the other case.

A similar finding on heating needs can be retrieved from the gas distribution.

The results presented Figure 4.11 are the results obtained when computing an HPD region with a level of confidence of 95% but it is interesting to study the performance of the tool when considering different levels of confidence. Thus, the results obtained for three other levels of confidence (80%, 60% and 50%) are displayed Figure 4.12.

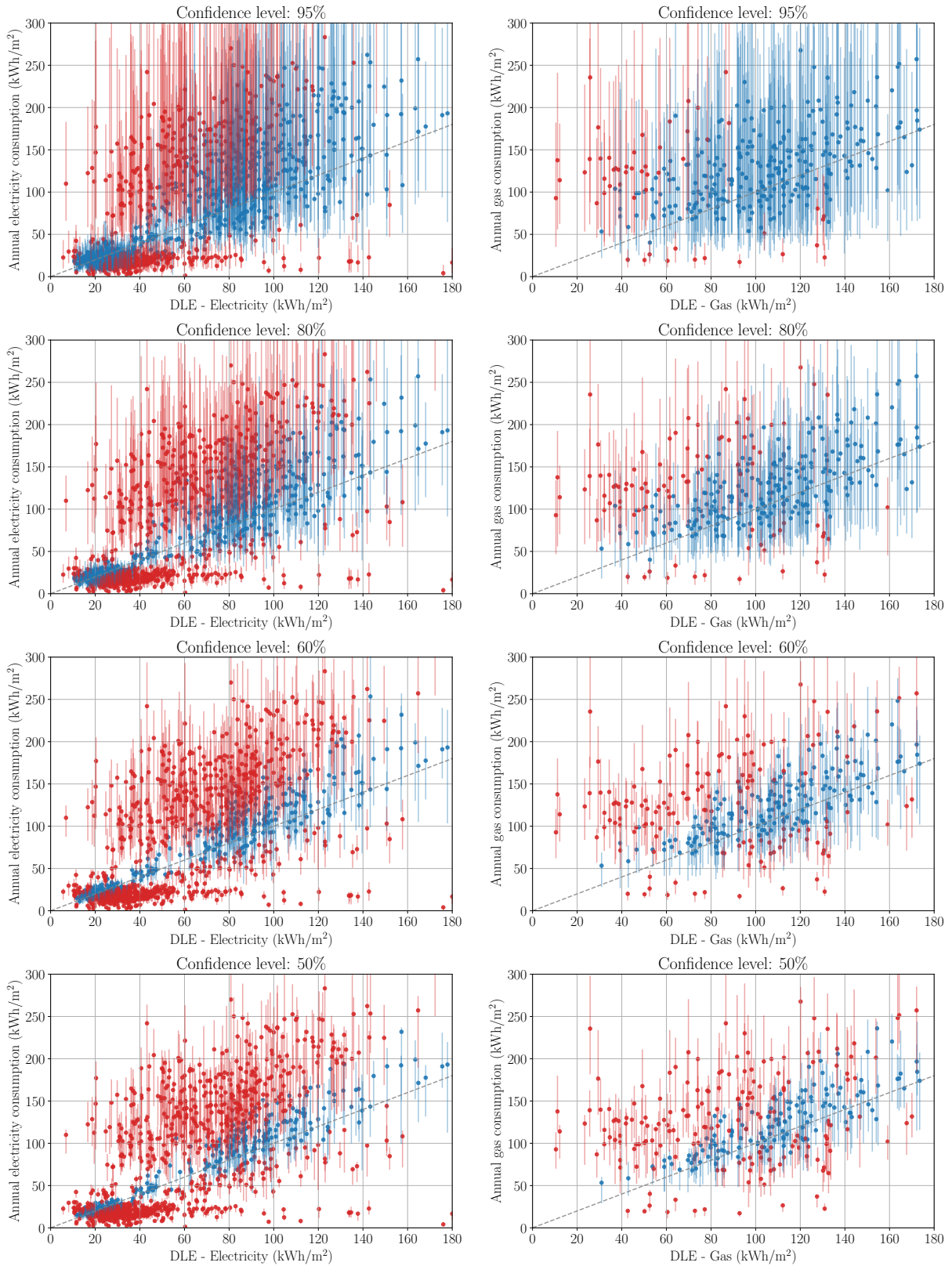


Figure 4.12: μ_{bldg} and the HPD region of each building against its DLE

On this figure, it appears clearly that the number of buildings for which the measurement does not fall in the HPD region increases rapidly when the level of confidence decreases. This trend is summarized by the evolution of the empirical coverage obtained for the four different level of confidence depicted in Figure 4.13.

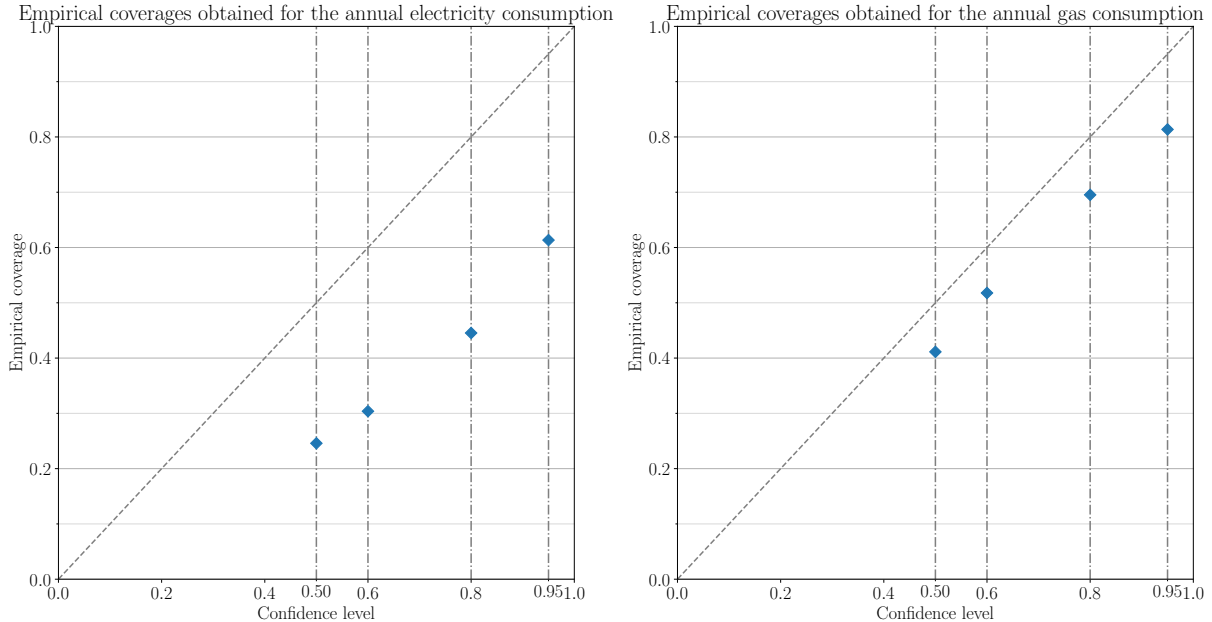


Figure 4.13: Evolution of the empirical coverage for different levels of confidence

Here, the four values displayed for both gas and electricity should always be as close as possible to the first bisector since it would mean that, for a given level of confidence α , the tool obtains an empirical coverage close to α .

It appears clearly on this figure that Simply performs better when predicting the annual consumptions of gas than the consumptions of electricity. Since the gas consumptions take into account solely the heating and DHW needs, this illustrates a rather good behaviour of the thermal model implemented in Simply.

4.2.2.3 A convergence study to verify the consistency of the results

The results discussed in the previous section have been obtained as a result of an uncertainty analysis. This uncertainty analysis was run with one hundred configurations for each building of the validation dataset. This sample size is quite small compared to what is usually recommended for uncertainty analysis [114].

In any case, a convergence study should always be led hand in hand with an uncertainty analysis as advocated by Soize [94]. Since Simply requires less computation time than other tools such as Dimosim, the convergence study is conducted in this manuscript on this tool only and the conclusions are applied as is to the other tools reviewed in the next part. Nonetheless, a thorough modeller should verify that the convergence stands for each tool in the validation process.

The results obtained by Simply on this convergence study are introduced Figure 4.14.

On this figure, the empirical coverage obtained for four levels of confidence is plotted against the number of samples used in the uncertainty analysis. Although there is a distinct difference between the results obtained with 1500 samples and the results obtained with 100 samples, there is no major shift of the trend. It appears here that the uncertainty analysis is rapidly converging,

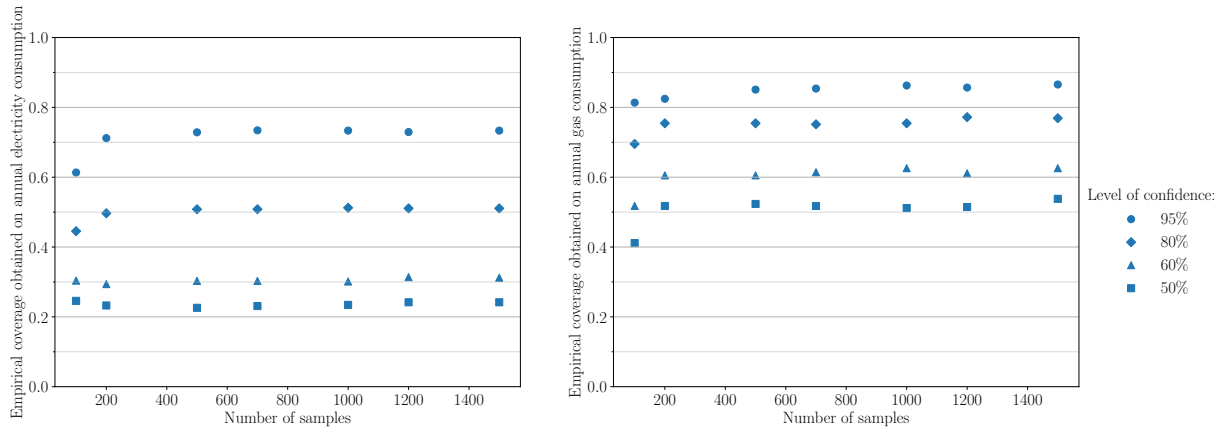


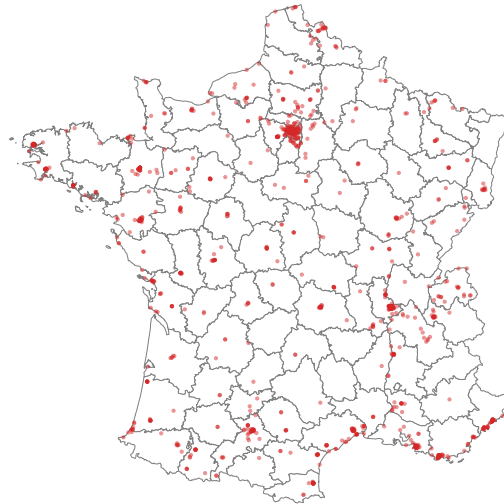
Figure 4.14: Evolution of the empirical coverage obtained by Simply for different number of samples

and even if better and more robust results could be obtained with a higher number of samples, the choice to use only 100 samples is deemed acceptable.

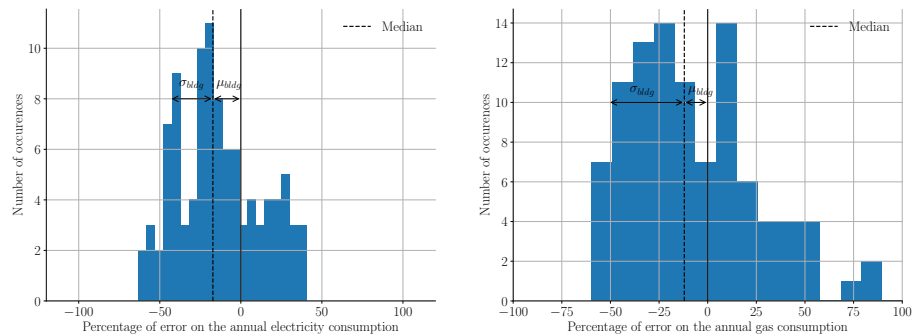
Now that the main steps of the assessment stage of the validation procedure have been detailed, the same method can be applied to other UBEM tools. A recap of the main steps is proposed hereafter before tackling the assessment of two other UBEM tools.

4.2.2.4 Summary of the assessment step applied to Simply

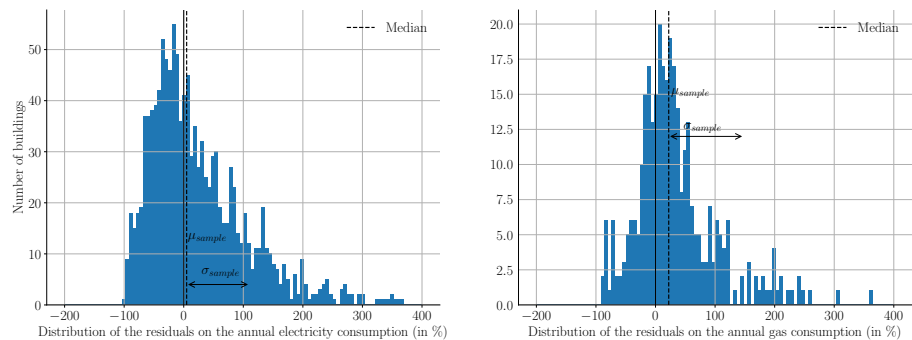
1. Select two KPI: the annual gas and electricity consumptions at the building scale.
2. Define a validation dataset to assess these KPIs:



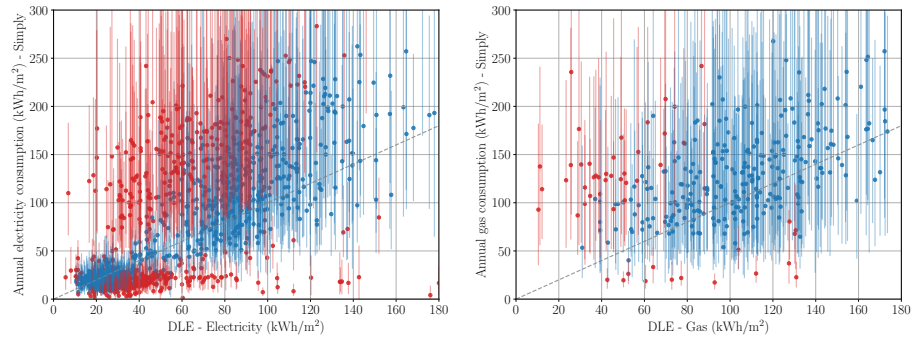
3. Quantifying and propagating uncertainty:
 - Defining probability density functions to characterize the inputs uncertainty;
 - Sampling the pdf;
 - Running the simulation for each sample.
4. Computing the distribution of the percentage of error for each building:



5. Retrieving the residual, μ_{bldg} , obtained on each building.
6. Computing the bias, μ_{sample} and the standard deviation, σ_{sample} obtained by the tool at the sample scale.



7. Computing the HPD region for each building:



8. Computing the empirical coverage.

9. Interpreting the three main indicators:

- μ_{sample} ;
- σ_{sample} ;
- The empirical coverage.

4.3 Applying the validation procedure to other UBLEM tools

Now that metrics have been defined for the assessment step, the proposed validation methodology can be applied to other tools.

In this section, the results obtained at the assessment step by two UBLEM tools, Calomel and Dimosim, are presented. Dimosim was introduced in Chapter 3, and Calomel is a regulatory model used in the French EPC framework.

These two tools were chosen for different reasons. Regarding Dimosim, it is a well-established UBLEM tool, and it seemed particularly interesting to test the validation framework on an historical UBLEM. When considering Calomel, it enabled to test the validation framework with a model that is more detailed than Simply but still less detailed than Dimosim. It seemed thus interesting to study the validation of tools with different levels of detail.

A brief introduction to Calomel is provided in the next part before discussing the results obtained at both the building level and the validation dataset level.

4.3.1 Brief description of Calomel

As mentioned above, Calomel is a regulatory model, therefore, many standard assumptions are made regarding parameters such as the heating set point or the occupancy. All these assumptions are described extensively in Calomel's documentation [115].

Moreover, Calomel normally predicts only gas and electricity consumptions that are related to the heating and DHW needs. As the DLE measure consumptions from all uses, the consumption of household appliances has been added to the electricity consumption calculated with Calomel as was done previously for Simply.

A comparison of the main inputs parameters required by Simply and Calomel is provided in Table 4.6. The main difference is that the air-change rate, the set point temperature and the DHW needs cannot be changed in Calomel since these are not exposed parameters that the modeller can easily modify.

4.3.2 Introducing the results on the test building

The distributions of the annual gas and electricity consumptions obtained by Dimosim, Calomel and Simply on the test building are displayed Figure 4.15.

When taking a closer look at the distributions obtained for the annual electricity consumption, the three distributions seem to shift from the left to the right with Simply underestimating the consumptions, Dimosim overestimating the consumptions by the same order of magnitude as Simply and Calomel widely overestimating the consumptions. A similar behaviour can be observed on the gas distributions.

The metrics defined in section 4.2.2.1 to assess the performances of UBLEM tools at the building scale are now to be applied to the three tools.

First, the distribution of the percentage of error obtained on the test building by each tool is computed and displayed Figure 4.16.

The residual, $\mu_{bl dg}$, for the electricity ranges from -17% for Simply to 64% for Calomel with Dimosim having an in-between value of 21% . For the gas, all tools show a better performance with lower absolute value of the residual. When considering the standard deviations obtained

	Simply	Calomel
Thermal properties and systems	Walls U-value	
	Roof U-value	
	Windows U-value	
	Air-change rate	Cannot be changed
	Set point temperature	Cannot be changed
	Heating system	
	Heating system efficiency	Cannot be changed
	DHW system	
Building geometry	DHW system efficiency	Cannot be changed
	Roof area	
	Windows-to-wall ratio	
	Building heated volume	
	Total floor area	
Weather data	Walls area	
	Outdoor air temperature	
	Solar gains	
Others	DHW needs	Cannot be changed
	Appliances electricity consumption	Manually added as for Simply
	Internal gains	Manually added as for Simply

Table 4.6: Comparison of the input parameters required by both Simply and Calomel

on gas and on electricity, the standard deviation obtained by the tools are generally higher for the gas than for the electricity.

Results obtained on the test building by the three tools are summarized Table 4.7.

	DLE electricity (kWh/m ²)	Median electricity consumptions (kWh/m ²)	DLE gas (kWh/m ²)	Median gas consumptions (kWh/m ²)	$\mu_{bldg_{elec}}$ (%)	$\sigma_{bldg_{elec}}$ (%)	$\mu_{bldg_{gas}}$ (%)	$\sigma_{bldg_{gas}}$ (%)
Simply		17		89	-17	26	-12	39
Calomel	20	33	102	130	64	11	28	32
Dimosim		25		108	21	16	7	59

Table 4.7: Results obtained with the three tools on the test building

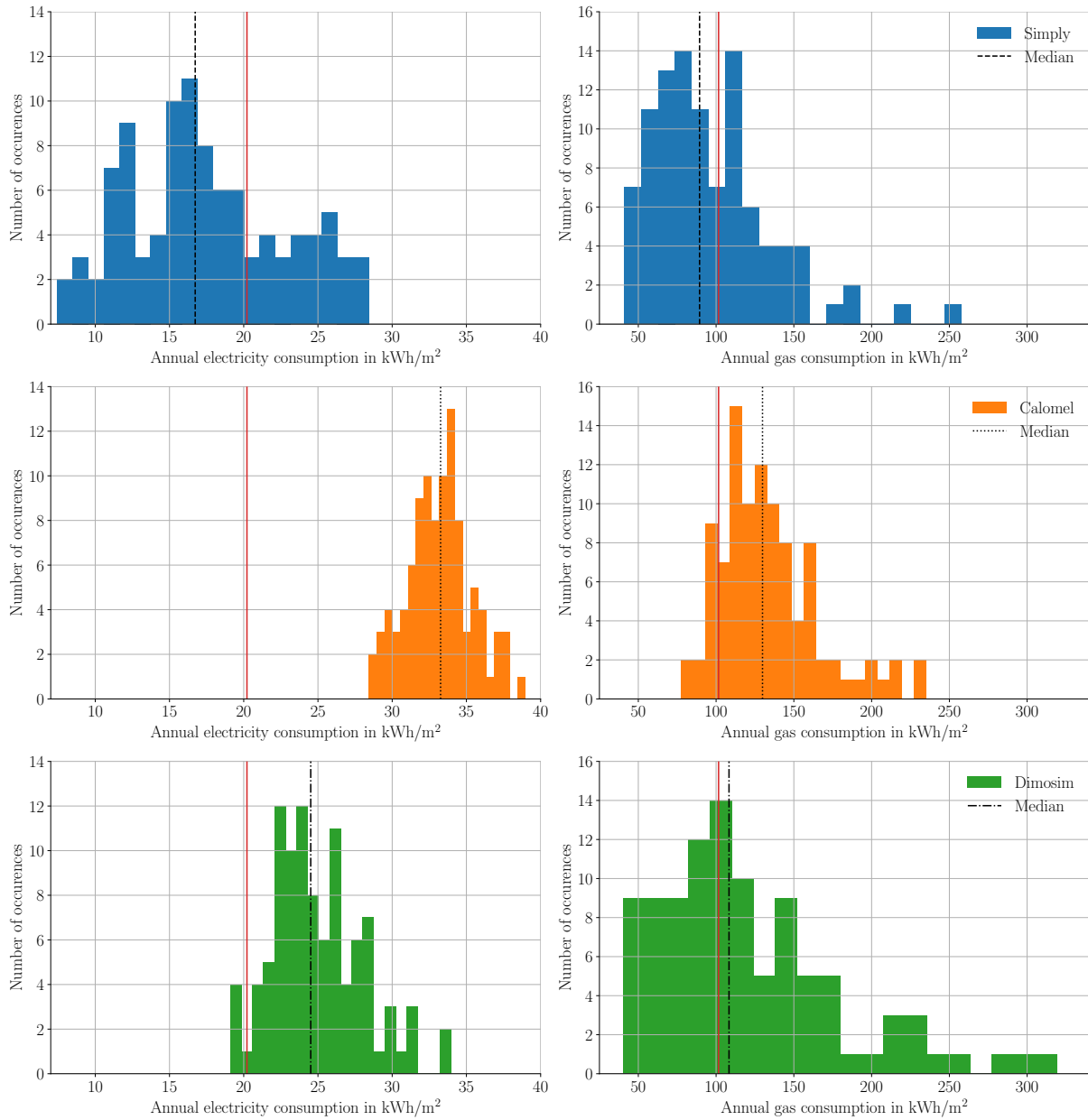


Figure 4.15: Distributions of the annual gas and electricity consumptions obtained by Simply, Calomel and Dimosim on the test building

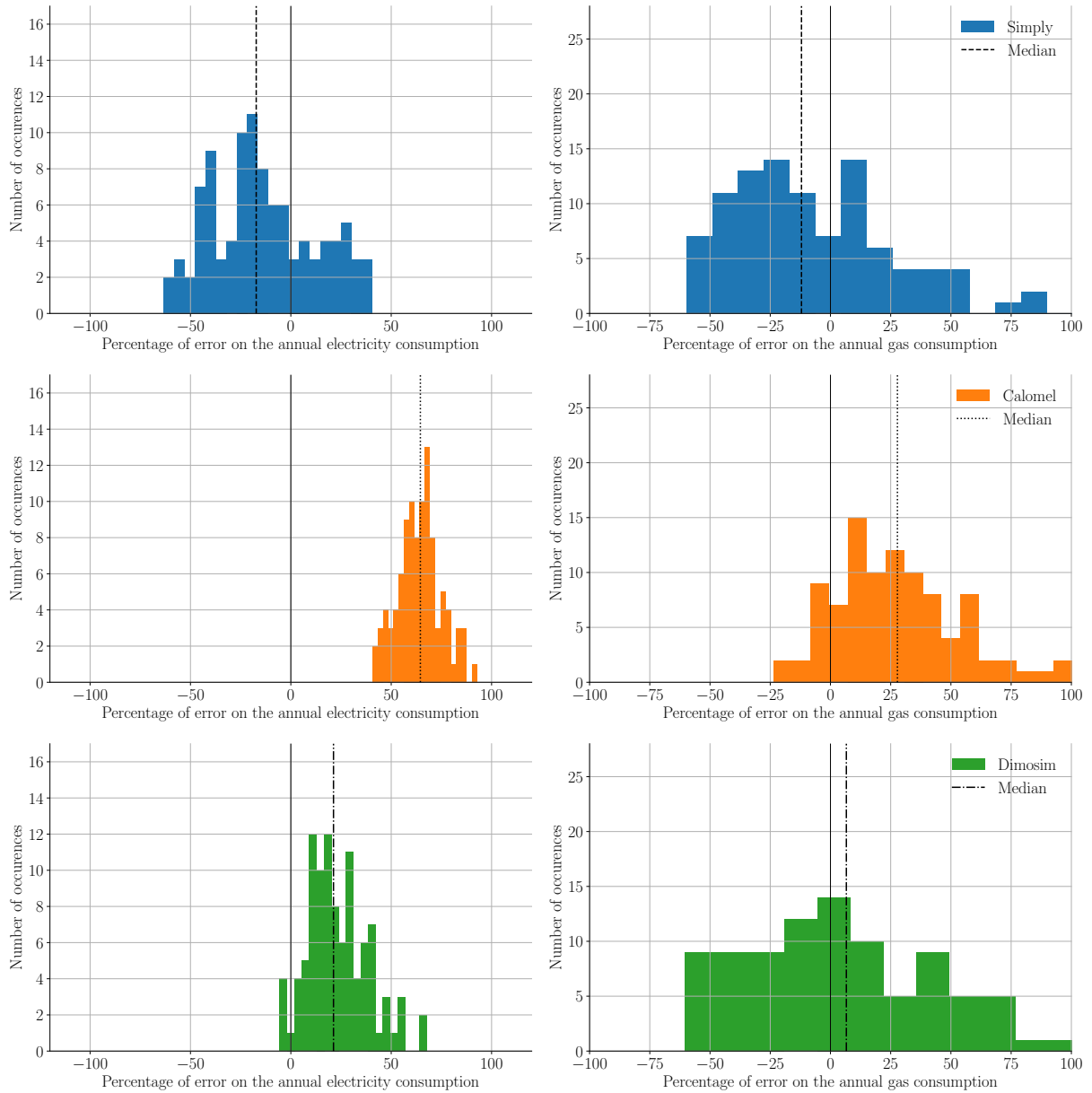


Figure 4.16: Distributions of the percentage of error obtained by Simply, Calomel and Dimosim on the annual gas and electricity consumptions of the test building

4.3.3 Introducing the results on the entire validation dataset

The focus is now set on the tools performances at the scale of the entire validation dataset. First, the distribution of the residuals obtained on all buildings by each tool is plotted Figure 4.17.

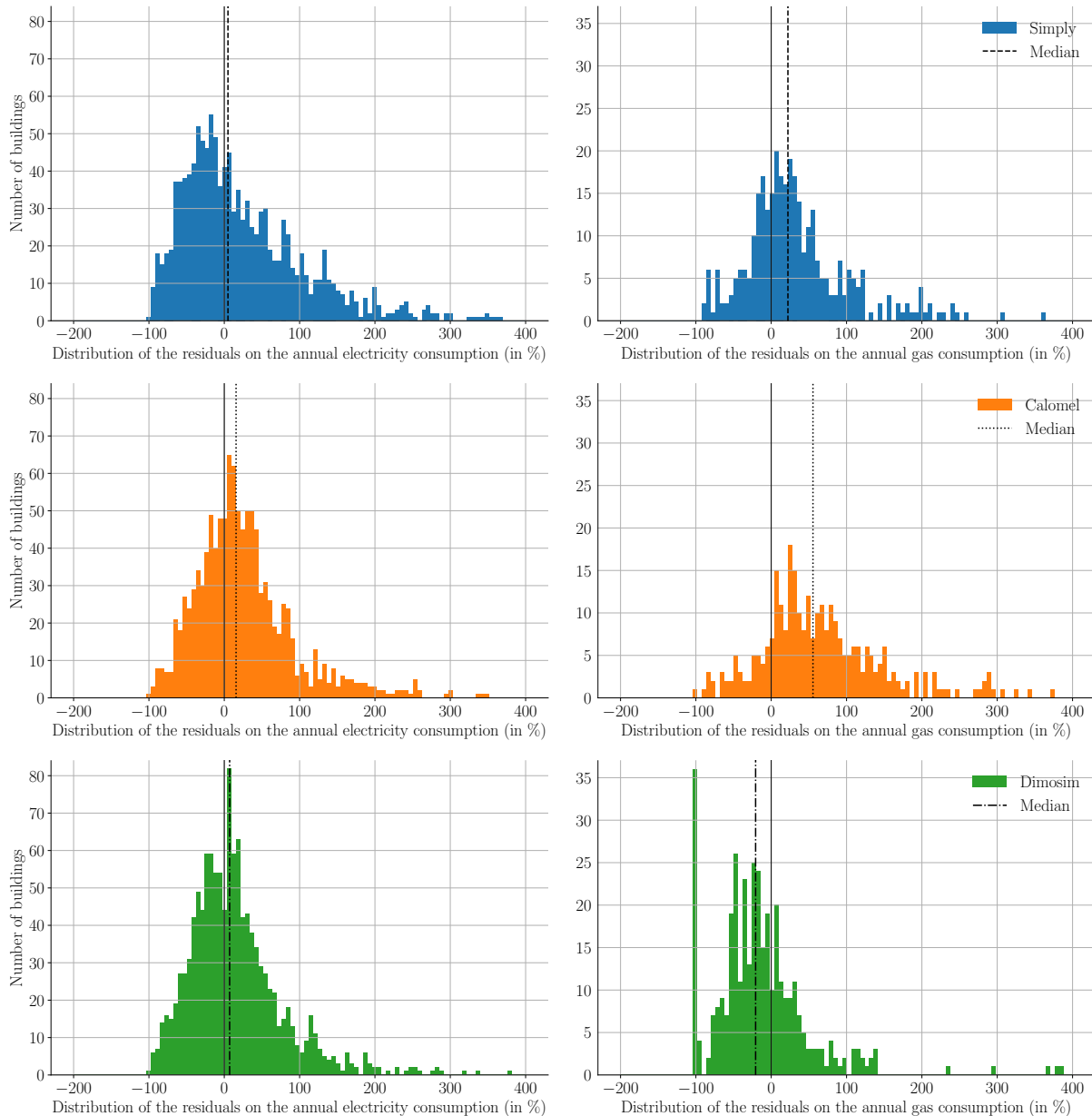


Figure 4.17: Distributions of the residuals obtained on the validation dataset by Simply, Calomel and Dimosim

The distributions obtained by the three tools for the annual electricity consumptions are almost centred in zero with a bias, $\mu_{sample_{elec}}$, of 5% for Simply, 7% for Dimosim and 16% for Calomel. When looking at the residual distribution obtained on the gas, median values, $\mu_{sample_{gas}}$, range from -21% for Dimosim to 55% for Calomel. In the case of the distribution obtained by Dimosim for the gas, 36 buildings (i.e approximately 3% of the validation dataset) display a residual of -100%. These buildings are buildings for which a gas measurement was provided by the DLE but for which no gas consumption was computed by the tool. These are usually buildings for which there was a wrong allocation of the heating system during the model parameterization. This phenomenon does not appear for Calomel and Simply since they retrieve the information

on the heating system from different parameters of the EPC. The bias obtained by Dimosim on the gas would probably be closer to zero if this phenomenon was corrected.

Taking a closer look at the standard deviation on the gas, it ranges from 49% for Dimosim to 127% for Calomel, with Simply displaying an intermediate value of 98%. There is a large gap between the standard deviation obtained by Simply and the one obtained by Dimosim. It reveals that the residual distribution obtained by Dimosim is less spread out than the residual distribution obtained with Simply. Dimosim should then give way to fewer buildings with very high residuals.

As was done in Section 4.2.2.2, the HPD region was computed for each building simulated with each tool, along with the empirical coverage of each tool.

The HPD region and the median consumptions computed by each tool on each building are displayed Figure 4.18.

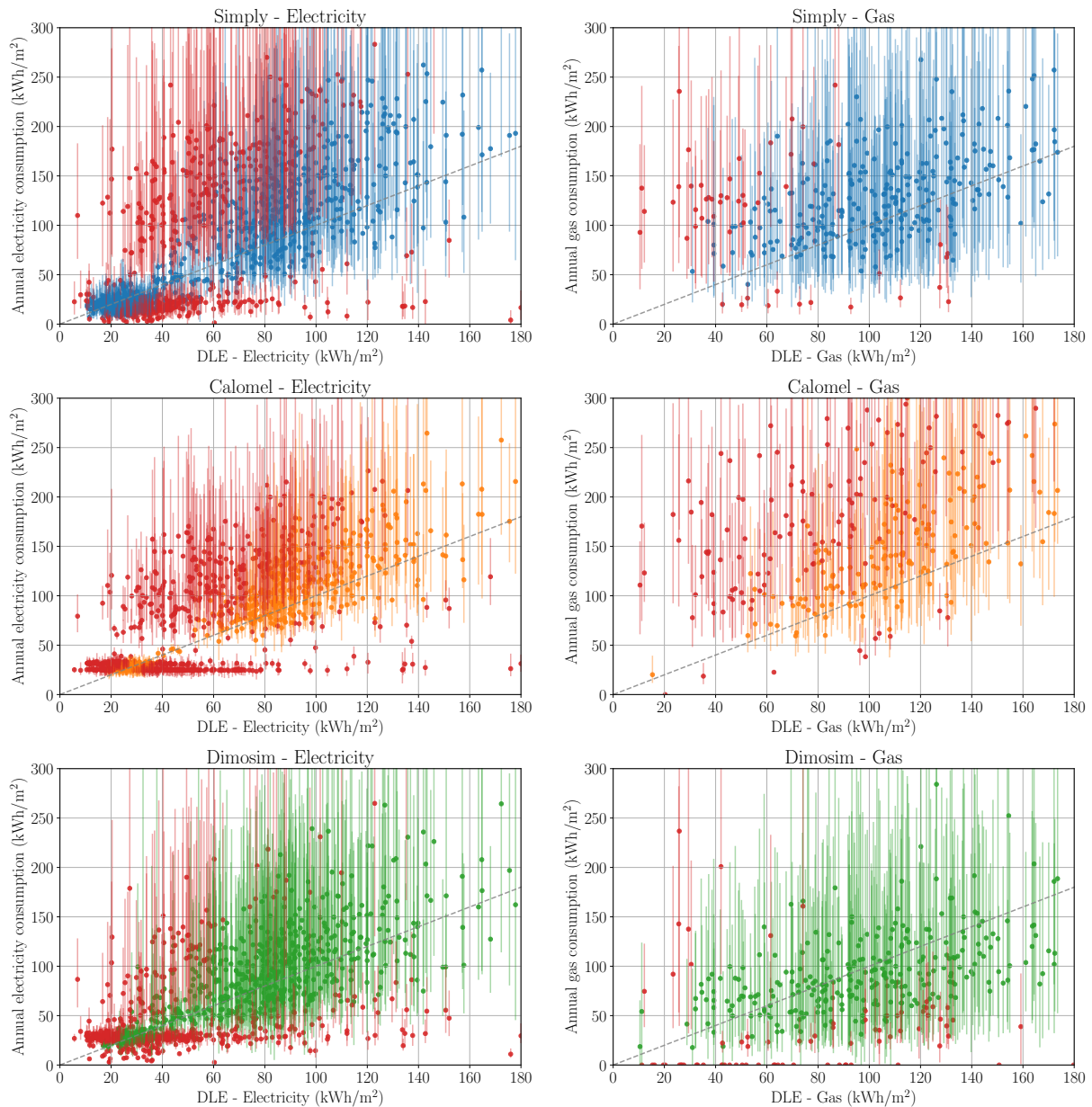


Figure 4.18: Distributions of the median consumptions and HPD region obtained by Simply, Calomel and Dimosim with a 95% confidence level

The empirical coverage was also computed for different confidence level for all the tools and the results are plotted Figure 4.19.

When looking at the empirical coverage obtained for the electricity consumptions, Dimosim and Simply are very close to each other and closer to the first bisector than Calomel, thus displaying a better performance than Calomel on this given metric. Surprisingly enough, Simply outperforms all the other tools on the gas consumptions. This result can be seen as quite surprising given the very low level of detail of this model. However, as it is a very simple thermal model, it requires less input data than Dimosim and may therefore be less prone to divergence due to missing or uncertain data. A more detailed analysis of these differences is carried on at the diagnosis step of the validation framework which is discussed in the next chapter.

To finish with, the results obtained by the three tools at the sample level are summarized in Table 4.8

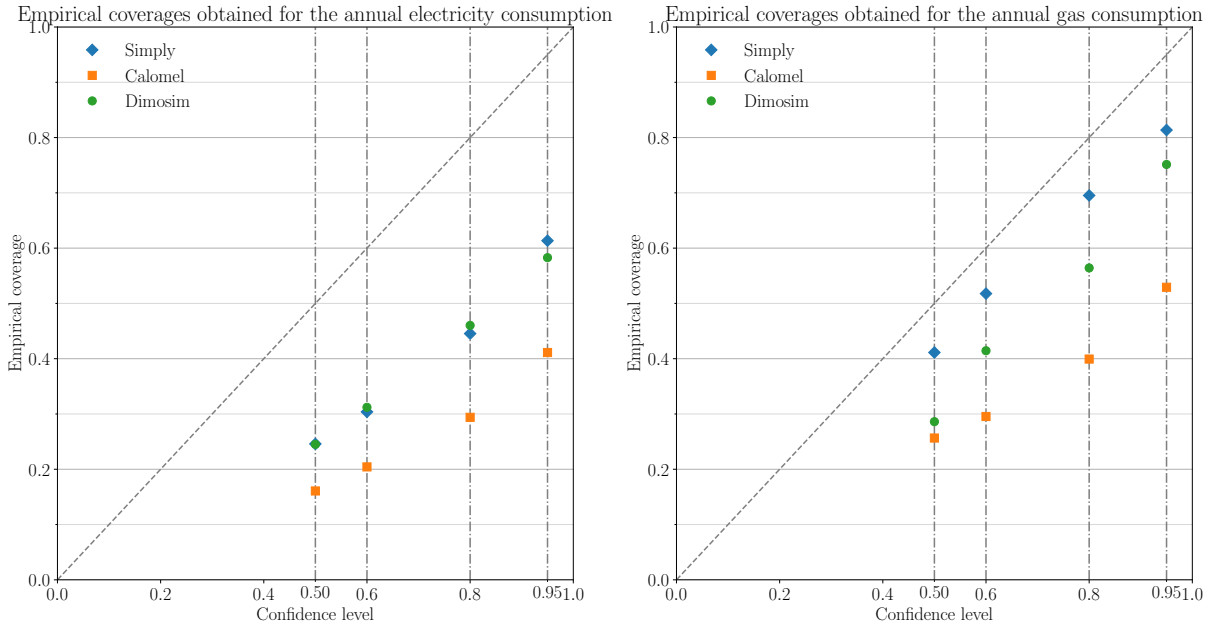


Figure 4.19: Evolution of the empirical coverage for different confidence levels for Simply, Calomel and Dimosim

	Empirical coverage - electricity	Empirical coverage - gas	$\mu_{sample_{elec}}$ (%)	$\sigma_{sample_{elec}}$ (%)	$\mu_{sample_{gas}}$ (%)	$\sigma_{sample_{gas}}$ (%)
Simply	0.61	0.81	5	105	22	98
Dimosim	0.58	0.75	7	75	-21	49
Calomel	0.41	0.53	16	78	55	127

Table 4.8: Results obtained with the three tools at the sample level

Conclusion

In this chapter, the assessment stage of the validation methodology was detailed.

It starts with the quantification of the uncertainty. This first step consists in identifying the uncertainty related to the data used in the validation dataset, and in defining probability distributions representing the uncertainty on each input parameters. Distinction was made between independent and correlated inputs.

Once the uncertainty has been quantified, its propagation consists in running the model for each set of input parameters. Metrics to assess the obtained results are then first defined at the building scale before being studied at the sample scale. Thus, we defined the residual at the building scale, μ_{bldg} , which is the median value of the distribution of the percentage of error obtained on each building. Two metrics were defined at the sample scale to quantify the distribution of the residuals obtained on the entire validation dataset: the bias, μ_{sample} , and the standard deviation, σ_{sample} . Finally, to assess the behaviour of the tool with regard to the expected confidence level, a last metric, the empirical coverage, was defined.

The validation procedure was then applied to two other tools: Dimosim and Calomel. The metrics defined above were applied to these two tools and enabled an inter-model comparison

between Simply, Dimosim and Calomel.

Dimosim and Simply display better results on the electricity annual consumptions than Calomel, while Simply takes the lead on the gas annual consumptions.

The assessment stage defined in this chapter has made it possible to quantitatively evaluate UBEM tools, but a validation procedure cannot be limited to the assessment of the models and must also propose indicators to help the modellers in the diagnosis of their tools' weaknesses. The diagnosis stage of the validation procedure is then presented in the next chapter.

Chapter 5

Diagnosis: Understanding the weaknesses of UBEM

Contents

5.1	Presentation of the diagnostic stage applied to Simply	95
5.1.1	Identification of parameters influencing the error at the sample scale . .	95
5.1.2	Identification of parameters influencing the error at building scale . . .	98
5.1.3	Identification of buildings with a similar behaviour towards the error . .	101
5.1.4	Summary of the diagnostic stage applied to Simply	105
5.2	Applying the diagnostic stage to other tools	105
5.2.1	Identification of parameters influencing the error at the sample scale . .	105
5.2.2	Identification of parameters influencing the error at building scale . . .	106
5.2.3	Identification of buildings with a similar behaviour towards the error . .	107
5.3	Studying the influence of different levels of detail on the tools per- formances	111
5.3.1	Evolution of the simulated consumptions with regard to data availability	115
5.3.2	Assessment stage applied to three levels of data availability	118
5.3.3	Diagnostic stage applied to three levels of data availability	121
5.3.4	Comparing the influence of different levels of details on three tools' per- formances	125

Introduction

In the previous chapter, a procedure to assess UBEM tools based on uncertainty was introduced along with a set of metrics to properly quantify the tools performances. These metrics enabled to give a score to each tested tool. This score is rarely as high as the modeller would expect. A stage of diagnostic is then necessary following the assessment stage to understand where the weaknesses of the tools are located and how to improve the tools' score.

Several reasons could explain a bad score, among which are the following:

- The input data is of poor quality ("garbage in, garbage out");
- The enrichment process is faulty;
- The physical model is faulty;
- The hypotheses made for the uncertainty quantification are inappropriate.

These reasons can be grouped in three error sources:

- **the data;**
- **the UBEM itself;**
- **the application of the validation framework.**

It is difficult to distinguish between these three sources, but to help gain some clarity on the subject, several sub-questions can be studied :

1. **Is it possible to identify parameters that have an influence on the error at the sample scale?**
2. **Is it possible to identify parameters driving the error for each building?**
3. **Is it possible to identify groups of buildings with a similar error behaviour?**

None of these questions can strictly distinguish between the three error sources by itself, but the results obtained for each one when compared with the others should help identify trends and to understand if one of the sources (the data, the UBEM or the application of the validation framework) is more responsible for the error than the two others.

The results obtained by Simply, Calomel and Dimosim are studied in the light of these three sub-questions. Since the same validation procedure was applied to the three tools with the same enrichment process, the comparison of the behaviour displayed by each of these tools on each of the sub-questions should help understand more precisely the influence of the physical model on the error.

UBEM users are often faced with different levels of data availability from one case study to another, therefore modifying the way the physical model of the UBEM is parameterized. To understand how this change in the model's parameterization can influence the results of a UBEM, a study is led here with different levels of data availability. The approach chosen here was to remove the uncertainty from the data that was deemed available, therefore this study also enables to question the way the validation framework was applied, namely by challenging the uncertainty quantification step.

5.1 Presentation of the diagnostic stage applied to Simply

The Percentage of Error (PE) was defined in the Chapter 4 to measure the error made by the tool for each simulated configuration of a given building. Distributions of percentage of error were obtained for each building such as the one displayed in Figure 5.1.

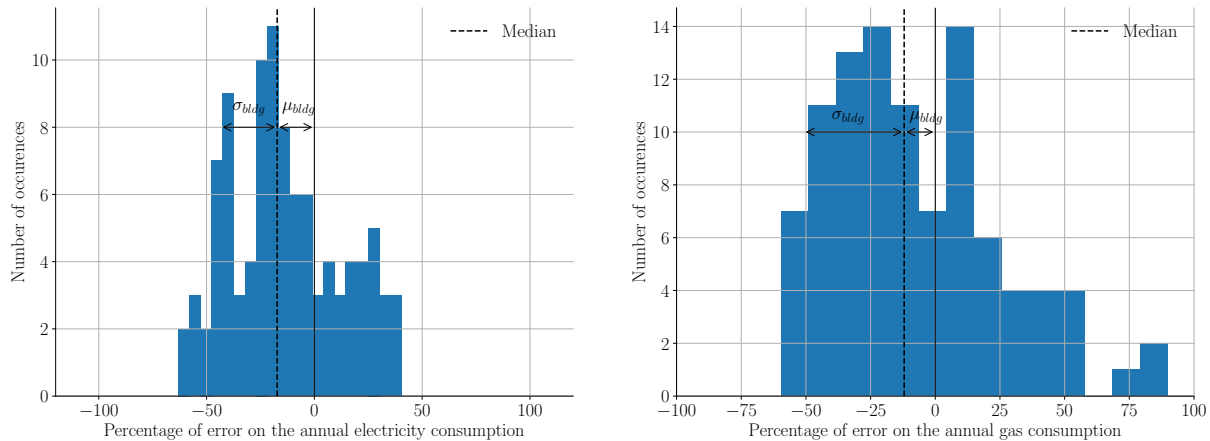


Figure 5.1: Distributions of the percentage of error obtained by Simply on the annual gas and electricity consumptions for the test building

Each of these distributions was characterized thanks to two metrics:

- μ_{bldg} , the residual defined as the median of the distribution of the percentage of error;
- σ_{bldg} , the standard deviation of the distribution of the percentage of error.

Studying these two metrics should help enlighten the three questions mentioned above and provide some insights regarding the behaviour of the tool towards the error.

5.1.1 Identification of parameters influencing the error at the sample scale

In this part, we try to identify parameters characterizing the error obtained on the whole validation dataset. To do so, the distribution of residuals obtained on the whole sample is studied. The variation of the residual from one building to another should be explained by parameters that characterize each building such as the energy carrier for heating or the construction year.

To identify a potential relationship between the residual and these parameters, the mutual information was computed between the residuals obtained by Simply for all buildings of the validation dataset and the main characteristics of each building. Some parameters characterizing a building were deemed uncertain in the validation procedure, such as the walls U-value. Since the distribution of the walls U-value used for the uncertainty propagation may differ from one building to another, each building was characterized by the mean of its walls U-values used in the uncertainty propagation. The same approach was applied to the main uncertain parameters.

The mutual information obtained between the residuals, for both electricity and gas, and the characteristics of the buildings is shown in Figure 5.2.

Two parameters exhibit a value of mutual information with the residual that is much higher than the other parameters: the energy carrier for heating and the energy carrier for DHW.

This means that there is a noticeable relationship between these two parameters and the residual, and that the variation of the residual could be explained based on the values of these two

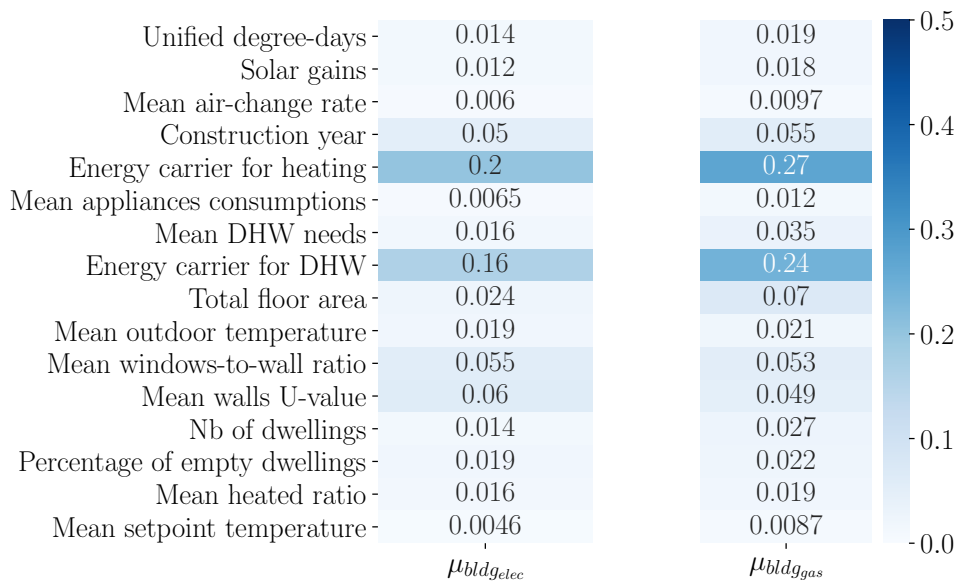


Figure 5.2: Mutual information computed between the residual, μ_{bldg} , and buildings characteristics

parameters. This insight is illustrated in Figure 5.3. On this graph, the values of μ_{bldg} have been plotted against the values of the two influential parameters and against the values of a non-influential parameter (the unified degree-days). The difference of influence on the error is clearly visible between the energy carrier for heating and DHW and the unified degree-days.

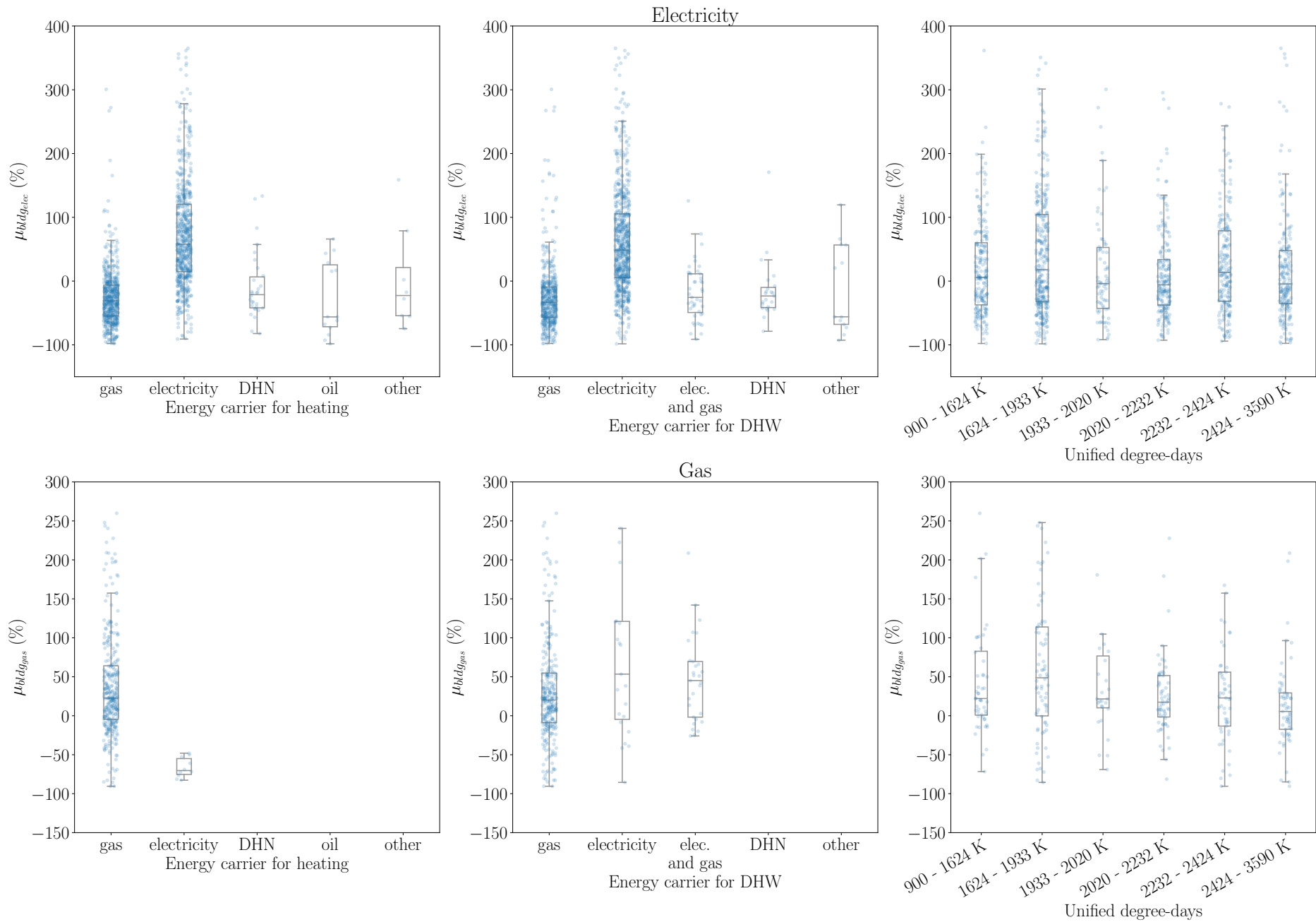


Figure 5.3: Relationship between uncertain parameters and the residual

This first step, the identification of parameters influencing the error at the sample scale, has then enabled to identify two buildings characteristics that could play a role in the error: the energy carrier for heating and the energy carrier for DHW. Regarding these two parameters, they were considered as certain during the uncertainty quantification. The energy carrier for heating and the energy carrier for DHW are provided by the EPC of each building and are generally regarded as the most reliable information that can be retrieved from the EPC since diagnosticians that make the EPC have generally access to the heating and DHW systems of the dwelling they survey.

At this stage, the influence of these two parameters on the error could be explained by two different causes:

- The heating and the DHW energy carriers provided by the EPC are not consistent with the actual consumptions of gas and electricity provided by the DLE;
- The UBEM displays an error that depends on the part of the model that drives the energy consumption.

Considering that the data is in this case regarded as quite reliable, the error is most likely to come from the second cause. This hypothesis is conformed by the Figure 5.3 where the residual distribution for electricity for buildings heated with gas seems to be much narrower than the one for buildings heated with electricity. This would mean that the part of the tool modelling the consumptions of appliances generates a smaller error than the part modelling the heating needs.

In order to understand more precisely what explains the error, the parameters influencing the error for each building are then studied.

5.1.2 Identification of parameters influencing the error at building scale

Now that parameters have been identified at the sample scale as influencing the distribution of the residuals, a focus is offered here to understand what explains the variation of the percentage of error displayed by each building such as the one presented Figure 5.1.

Identifying parameters influencing the error is usually done using classical approaches of sensitivity analysis. However, since the correlations among input parameters were taken into account during the uncertainty quantification, these approaches are not applicable as is here [65]. Instead, the mutual information is used to identify parameters with the greatest influence on the error. The procedure used to identify influencing parameters for each building is first introduced on the test building that was used in Chapter 4. Next, it is applied to every building of the validation dataset.

5.1.2.1 Identifying parameters with an influence on the error for a given building

The distributions of the percentage of error obtained with Simply on the test building for electricity and gas consumptions is provided in Figure 5.4 as a reminder. As it was detailed in Chapter 4, the test building is located in Rennes, France, and its heating and DHW needs are covered thanks to a standard collective gas boiler (see Table 4.5 for a summary of its properties).

With this in mind, the goal is now to identify uncertain parameters explaining the variation of the distributions depicted in Figure 5.4.

To do so, the mutual information between the percentage of error obtained for each simulated configuration of the building and the values of the uncertain parameters for the same configuration is computed.

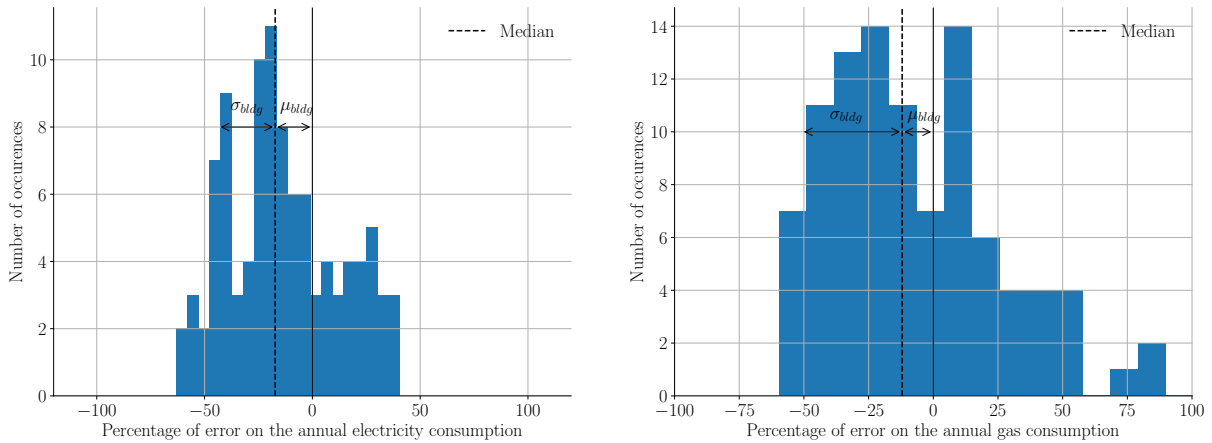


Figure 5.4: Distributions of the percentage of error obtained by Simply on the annual gas and electricity consumptions of the test building

The results are shown in Figure 5.5. The reader attention is drawn to the fact that the mutual information presented here is not normalised unlike mutual information tables presented before. This means that there is no upper bound for the possible values of mutual information. The normalisation enables to set an upper bound at 1 for the mutual information which makes it easier to analyse when considering a mutual information vector by itself. However, the normalisation is made based on the dataset used to compute the mutual information and is therefore dependent on the dataset related to each individual building.

Since it would be interesting to compare the ranking of the parameters influencing the most the error from one building to another, it requires to work with non-normalised mutual information in order to be able to compare the obtained values among the buildings of the validation dataset.

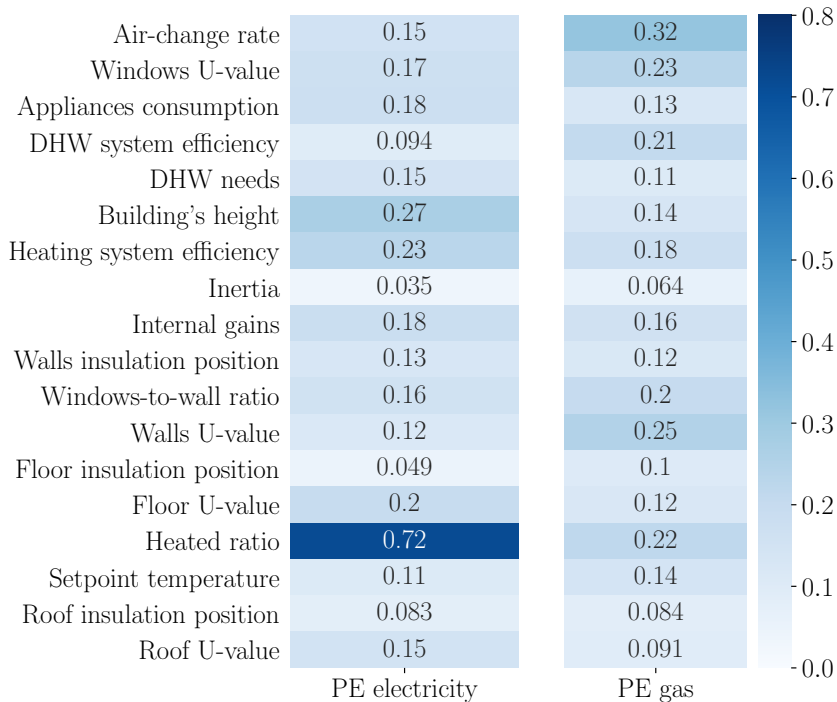


Figure 5.5: Mutual information between the percentage of error and uncertain parameters

When working with non-normalised mutual information, it may be harder to discriminate be-

tween meaningful parameters and random noise. In this case, the modeller's knowledge of the model should be used to set a threshold.

For instance, when looking at Figure 5.5, the percentage of error on electricity is highly linked to the heated ratio. Since the test building is not heated with electricity, it may seem surprising at first. However, the value provided for appliances consumption is in kWh/m²/year and is then multiplied by the total floor area of the building to obtain the actual electricity consumption related to appliances. The total floor area depends on the heated ratio which scales up or down the building geometry at the beginning of the simulation. This explains why this heated ratio appears with such a high value of mutual information.

Conversely, since the building is not heated with electricity, Equation 4.6 clearly shows that the heating system efficiency cannot influence the electricity consumption (there is no auxiliary consumptions considered by Simply). On this vector of mutual information, it can then be determined that values below 0.23 are the image of random noise and should not be considered. This threshold is linked to this particular vector of mutual information, but its value would likely be different for another vector.

When looking at the vector of mutual information obtained for the error on the gas consumptions, the four most influencing parameters are:

1. The air-change rate;
2. The walls U-value;
3. The windows U-value;
4. The heated ratio.

This seems well appropriate since the building is heated with gas.

As described for the vector for electricity, it is also possible here to define a lower bound below which the mutual information values are random noise. In the case of Simply, the floor U-value is not a parameter of the model as it can be seen Equation 4.5. The values of mutual information below 0.12 may then be considered in this example as random noise.

For this given building, the parameters that seem to drive the error are pretty consistent with influencing parameters that can be found in the literature when sensitivity analysis is applied to Building Energy Models. This conclusion would indicate that the physical model behaves as expected and that the error comes from the fact that the main parameters are not well known. In this case, the error could originate from the input data, from the enrichment process or from the way uncertainty was modelled.

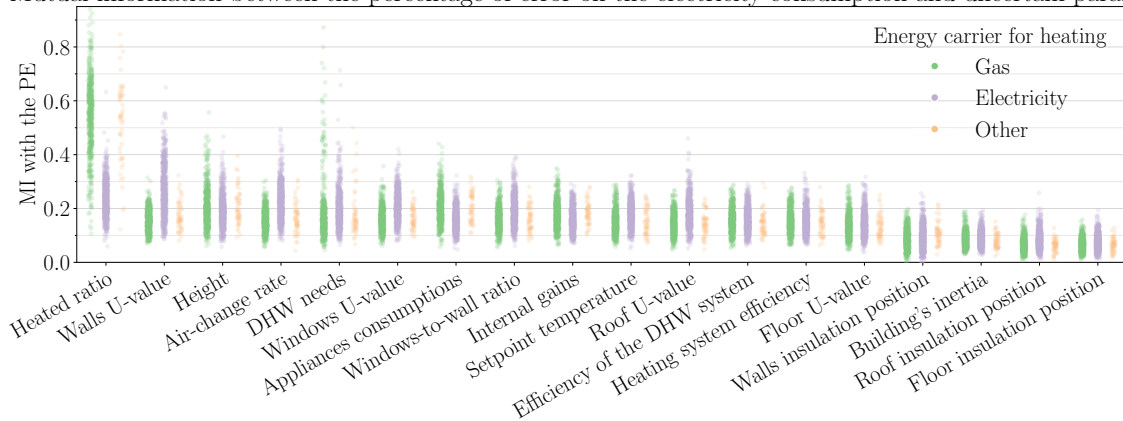
5.1.2.2 Identifying parameters with an influence on the error for every building of the validation dataset

Now that a procedure to identify parameters influencing the error for a given building has been described, it can be applied to every building in the validation dataset. By doing so, it should be possible to reinforce the conclusion outlined above or to discard it.

The mutual information values obtained by all buildings between the percentage of error and uncertain parameters are displayed in Figure 5.6.

Since the energy carrier for heating was identified in section 5.1.1 as a parameter strongly influencing the residual, the buildings were separated based on their heating energy carrier. The

Mutual information between the percentage of error on the electricity consumption and uncertain parameters



Mutual information between the percentage of error on the gas consumption and uncertain parameters

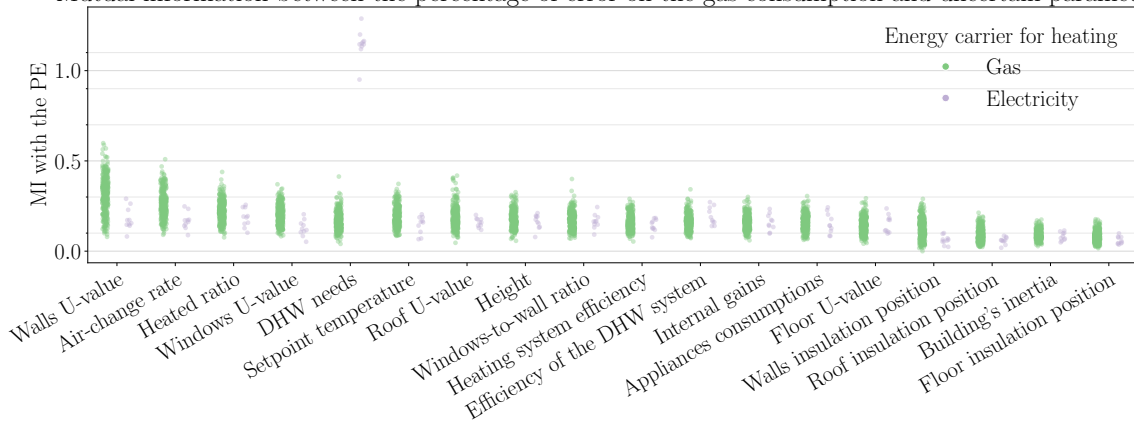


Figure 5.6: Mutual information between the percentage of error on the electricity and gas consumptions and the uncertain parameters for all buildings of the validation dataset

distribution of values obtained by all buildings of the validation dataset for the mutual information between the percentage of error and each uncertain parameter enables then to identify parameters influencing the error, depending on the heating energy carrier of the buildings.

For instance, it appears that the heated ratio is the parameter with the greatest influence on the error on electricity consumptions for all the buildings that are not heated with electricity. Conversely, the error on electricity consumptions is mainly influenced by the walls U-value for the buildings that are heated with electricity.

When looking at the vector of mutual information obtained for the error on the gas consumptions, very high values of mutual information can be observed for the DHW needs for buildings heated with electricity. It probably corresponds to buildings not heated with gas but for which the energy carrier for DHW is gas. In this case, it seems normal to have a high link between the error on the gas consumptions and the DHW needs.

5.1.3 Identification of buildings with a similar behaviour towards the error

To go further in the analysis, groups of buildings with a similar behaviour towards the error are identified. Parameters with the greatest influence on the error could then be determined for each group of buildings.

To create groups of elements within a dataset, it is a common practise to reduce the dimensionality in order to be able to visualise the dataset in a 2D plane. To do so, the well-known approach

of Principal Components Analysis (PCA) was selected. Two main reasons dictated this choice:

- The eigenvectors obtained through a PCA and used to project the dataset in a space of reduced dimension are a linear combination of the input parameters that can be interpreted;
- The PCA is a well-known approach, already used in the building energy modelling field (such as in [57]), and using a common statistical tool should help with the diffusion of the proposed validation framework.

The main idea behind the PCA is to create a new basis of vector and to identify the share of variation of the dataset explained by each vector of the new basis. If a few vectors explain the majority of the dataset variation, it is possible to reduce the number of vectors of the basis to the ones that are deemed meaningful, and thus to reduce the dimension of the space in which the dataset is projected. To verify that, the eigenvalues associated with each eigenvector of the new basis are plotted in Figure 5.7. These values are directly interpretable as the share of the dataset variation explained by each associated vector [116]. It seems then legitimate to reduce the number of eigenvectors to two in this case.

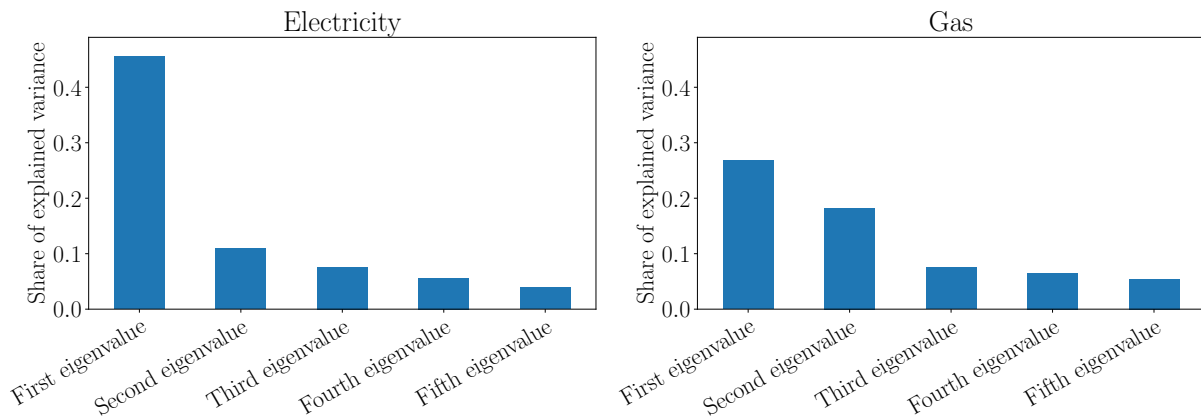


Figure 5.7: Share of variance explained by the first five eigenvectors obtained via PCA

The projection of the dataset in the plane formed by the first two eigenvectors obtained with the PCA are displayed in Figure 5.8.

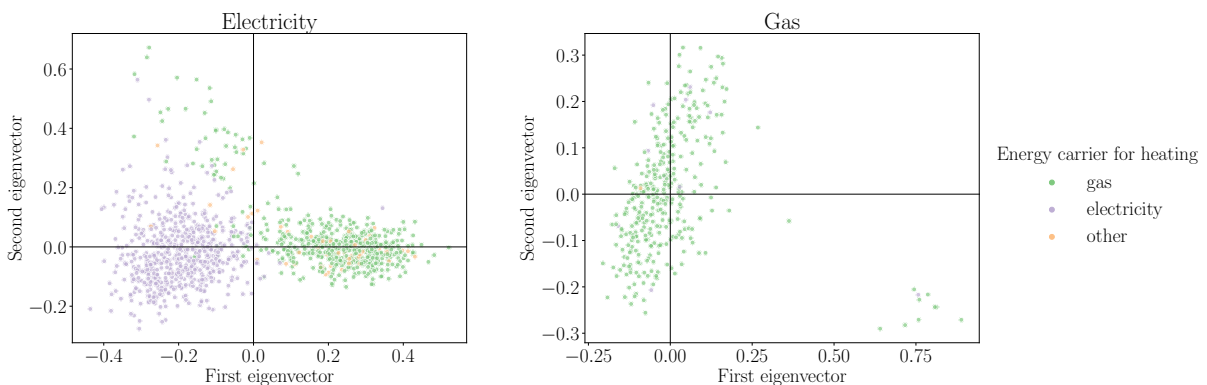


Figure 5.8: Projection of the mutual information table on the two-dimension space obtained via PCA - coloured based on the heating energy carrier

Each point in the scatter plot corresponds to a vector of mutual information for a given building. When looking at Figure 5.8 displaying the error on the electricity consumptions, the colour scheme clearly highlights two groups: buildings heated with electricity and buildings not heated

with electricity. It may be noticed that electrically heated buildings are located on the left-hand side of the first eigenvector axis, whereas the other buildings are mainly located on the right-hand side of this axis.

It is interesting then to take a look at the eigenvectors which are a linear combination of the initial parameters. The eigenvectors obtained for electricity and gas are displayed Figure 5.9.

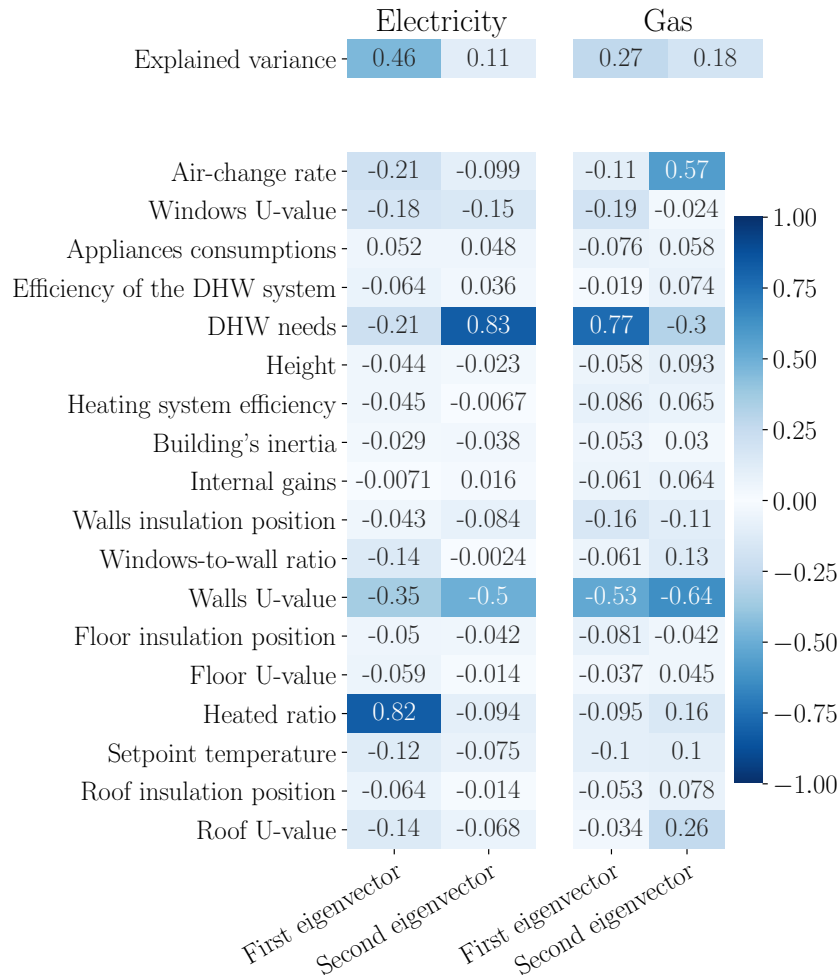


Figure 5.9: Eigenvectors obtained via PCA

Let us first analyse the results obtained with the PCA for the error on the electricity consumptions. Among the two eigenvectors, the first one explains the highest part of the variance of the vectors of mutual information. It is then studied first.

This first eigenvector displays values that are both positive and negative. The parameters with positive values characterise the buildings located along the positive side of the first eigenvector axis. On the contrary, the parameters with negative values characterise the buildings located along the negative side of the axis.

Back to the ranking of uncertain parameters in the first eigenvector obtained for electricity, there is one very high positive value: the heated ratio. This means that buildings that are not heated with electricity have an error on the electricity consumptions linked to the heated ratio. Since the heated ratio influences greatly the building total floor area which is in turn used to compute the electricity consumption for appliances in Simply, this result is considered convincing.

Regarding parameters with negative values, the four most important ones are:

- The walls U-value;
- The DHW needs;
- The air-change rate;
- The windows U-value.

These four parameters have then an influence on the error on the electricity consumptions for electrically heated buildings. This seems plausible too, considering the thermal model used in Simply.

Let us consider now the second eigenvector obtained for electricity, it explains a lower part of the variance, but it may nonetheless provide interesting insights. In this case, only two parameters stand out: the DHW needs with a positive value and the walls U-value with a negative value. To explain these two values it is worthy to look at the scatter plot with a different colour scheme. In Figure 5.10, the buildings have been coloured depending on the energy carrier for domestic hot water instead of the energy carrier for heating.

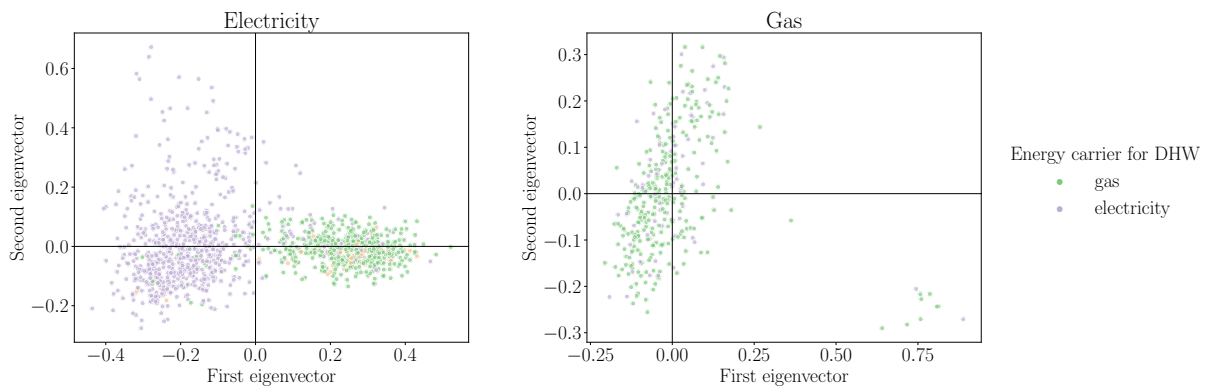


Figure 5.10: Projection of the mutual information table on the two-dimension space obtained via PCA - coloured based on the DHW energy carrier

A subgroup of buildings appears when comparing Figure 5.10 and Figure 5.8: buildings heated with gas but using electricity as energy carrier for domestic hot water. These buildings are located in the top left corner of the scatter plot obtained for electricity, and it seems then normal for these buildings to have an error on the electricity consumption influenced by the DHW needs. For the buildings electrically heated and using electricity for DHW, it seems that the walls U-value is still the most influencing parameter.

The focus is now set on the analysis of the PCA applied to the error on gas consumptions. In this case, two groups of buildings are also visible although the buildings are not evenly divided between these two groups. The first eigenvector for the gas displays the same kind of values as the second eigenvector for electricity: the DHW needs and the walls U-value have high absolute value though with opposite signs. It seems that the small group of buildings located on the right-hand side of the scatter plot have an error on gas consumptions mainly explained by the DHW needs. The conclusion is not as clear as it was for electricity since there is no clear difference between Figure 5.8 and Figure 5.10 on the energy carrier used for heating or DHW. This group of buildings may correspond to a group of buildings with very low heating needs and for which the error on gas consumptions would be mainly driven by the DHW needs.

In addition, the second eigenvector displays positive values for the air-change rate and the roof U-value, and negative values for the walls U-value and the DHW needs. This could illustrate a variety of behaviour towards the error within the group of buildings heated with gas.

The analysis led here indicates no major issue as to the behaviour of the physical model: the parameters influencing the error for each group of buildings seem quite natural given the common characteristics of each group of building. This would then put the blame on the data, on the enrichment process or on the way uncertainty was modelled, i.e. the way the validation framework was applied. Applying the same analysis to other tools and with different levels of data availability should help distinguish between these possibilities.

5.1.4 Summary of the diagnostic stage applied to Simply

In this section, the diagnostic stage of the validation framework developed in this manuscript was introduced and applied to a first UBEM tool, Simply.

This diagnostic stage articulates itself around three main points:

1. Identifying parameters that have an influence on the error at the sample scale;
2. Identifying parameters driving the error for each building;
3. Identifying groups of buildings with a similar behaviour towards the error.

These three points should help identify which cause is the most responsible for the error between the data, the UBEM and the application of the validation framework.

When studying the parameters with an influence on the error at the sample scale, the energy carrier for heating and the energy carrier for DHW stood out. This indicated that the UBEM has a different behaviour towards the error depending on the part of the model at stake.

The parameters driving the error at the building scale were then identified for each building of the validation dataset. They are quite distinct depending on the heating energy carrier of the buildings.

Groups of buildings with similar behaviour towards the error could then be defined. The buildings naturally grouped by heating and DHW energy carriers and parameters with an influence on the error could be retrieved for each group.

The parameters obtained are quite consistent with the expected energy behaviour of each building depending on the heating energy carrier. This would indicate that the physical model behaves in an appropriate way for each group of buildings and that the error is more likely then to originate from the data, from the enrichment process or from the uncertainty quantification.

5.2 Applying the diagnostic stage to other tools

Now that the diagnostic procedure has been introduced, it can be applied to Calomel and Dimosim, and results obtained for the three tools can be compared. As it was done for Simply, the parameters influencing the error at the sample scale are studied first. Then, the parameters influencing the error at the building scale are identified and groups of buildings with similar behaviour towards the error are distinguished.

5.2.1 Identification of parameters influencing the error at the sample scale

In this part, the distribution of residual obtained on the whole validation dataset is studied. The mutual information between the residual and the buildings characteristics is computed for the three tools and the results are provided Figure 5.11.

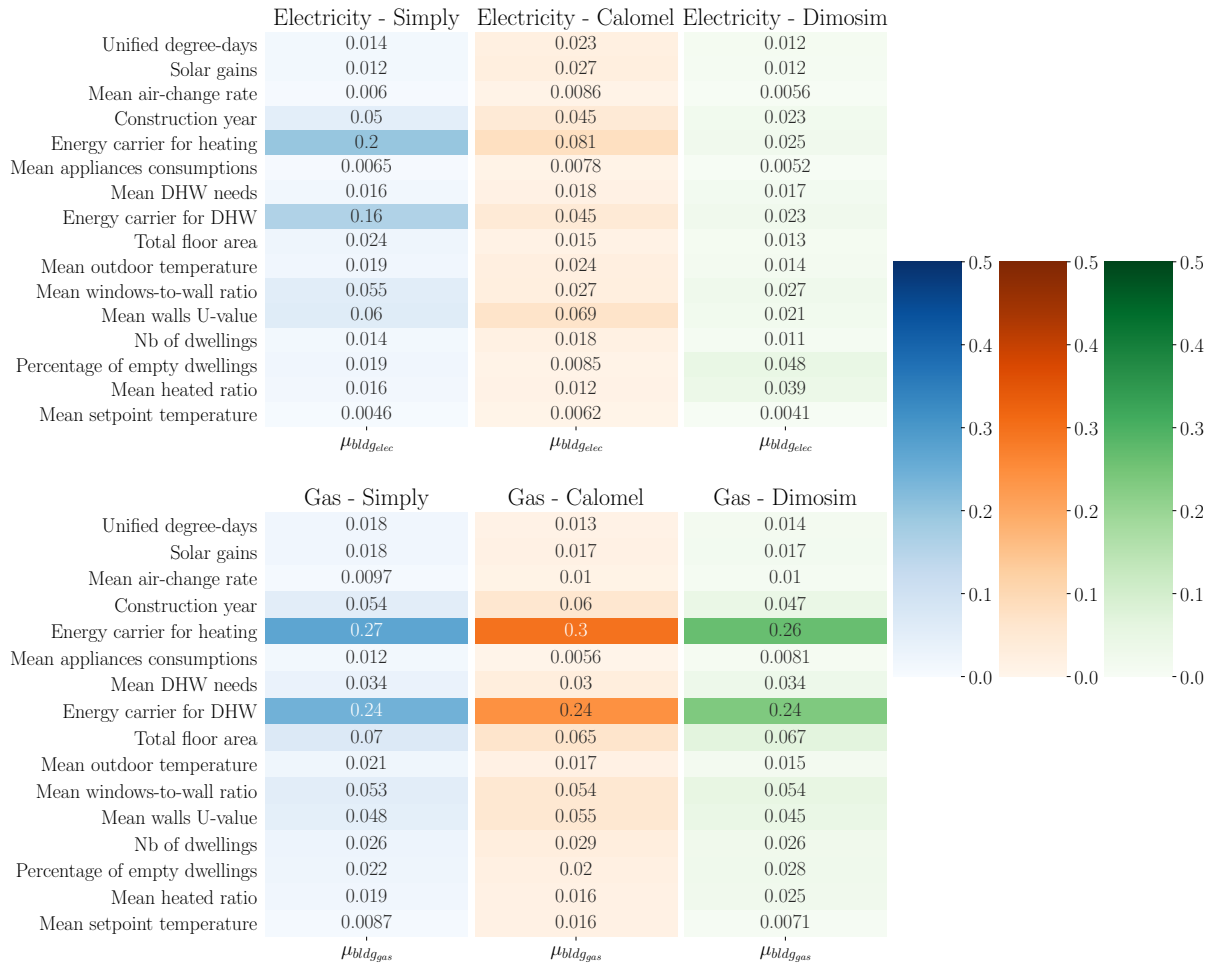


Figure 5.11: Mutual information computed between the residual and building’s characteristics

When looking at the mutual information obtained for the error on the gas consumptions, the three tools display a very similar behaviour; all depicting high values of mutual information between the heating energy carrier and the residual and between the DHW energy carrier and the residual.

However, there is no such thing when considering the values of mutual information obtained for the residual on electricity consumptions. The values of mutual information displayed by both Calomel and Dimosim are all very low and no parameter stands out. It is then difficult to draw any conclusion regarding the parameters that could influence the median error on electricity for these two tools. The next steps are all the more important in this case, and should help gain a better understanding of these tools’ behaviour.

5.2.2 Identification of parameters influencing the error at building scale

As it was explained above, a vector of mutual information is computed for each building between the uncertain parameters characterizing the building and the percentage of error. These vectors are depicted for the whole validation dataset and for the three tools in Figure 5.12 and in Figure 5.13. Once again, the buildings were split and coloured depending on their heating energy carrier.

It can be noticed immediately that the three tools display very different values of mutual information. For instance, for the error on the electricity consumptions, Simply displays a high value of mutual information between the percentage of error and the heated ratio for buildings that

are not heated with electricity whereas, in the case of Calomel and Dimosim, it is the appliances consumptions factor that stands out. As for the gas, the strongest relationship with the error seems to come from the walls U-value for Simply and Calomel while the air-change rate seems to be more influential in the case of Dimosim.

5.2.3 Identification of buildings with a similar behaviour towards the error

In order to determine groups of buildings with a similar behaviour towards the error, a principal component analysis is performed for each tool. The results obtained by the three tools are summarized in Figure 5.14 and in Figure 5.15.

The first observation that can be done on this figure is that the results of the PCA is very different from one tool to another. In most cases, the two first eigenvectors explain a great share of the initial dataset variance, but in two cases, for Calomel on the error on electricity and for Dimosim on the error on gas, the second eigenvector should not be considered since the associated share of explained variance is too low.

When focusing on the results obtained for electricity, Dimosim displays three groups of buildings, the two most important ones being the buildings heated with electricity and the buildings heated with gas. These two groups are spread almost perpendicular to each other. The group of electrically heated buildings is spread along the second eigenvector axis while the group of buildings heated with gas is spread along the first eigenvector axis. Two similar groups can be noticed for Calomel, but their spread is solely explained by the first eigenvector, their distribution along the second axis being meaningless since the share of variance explained by the second eigenvector is negligible. In the case of Simply, a similar trend to the one observed for Dimosim is noticeable although less pronounced.

When considering the values of the eigenvectors obtained for electricity, the appliances' consumption stands out at first glance from the first eigenvectors of Dimosim and Calomel, while it is the heated ratio for Simply. When considering the second eigenvectors, the DHW needs stand out for both Dimosim and Simply. The complete absence of this parameter in the eigenvectors obtained with Calomel is a good signal since the DHW needs could not be modified in Calomel as explained in Chapter 4 (see section 4.3.1 for more details). This parameter being kept at a pre-set value in this tool, it could not explain the variance of the error obtained by Calomel at building scale.

The attention is now turned to the error on gas. The scatter plots obtained for Dimosim and Simply are very similar to one another while the points displayed for Calomel seem more randomly spread out. The first eigenvectors obtained for both Dimosim and Simply show the highest value for the same parameter (the DHW needs). However, the similarity stops there. Simply displays then high values for the walls U-value, followed by the windows U-value, whereas for Dimosim, the air-change rate and the heated ratio take jointly the second rank. When looking at the eigenvectors obtained for Calomel, they both display high absolute values for the thermal parameters (walls, floor, roof U-values and insulation position).

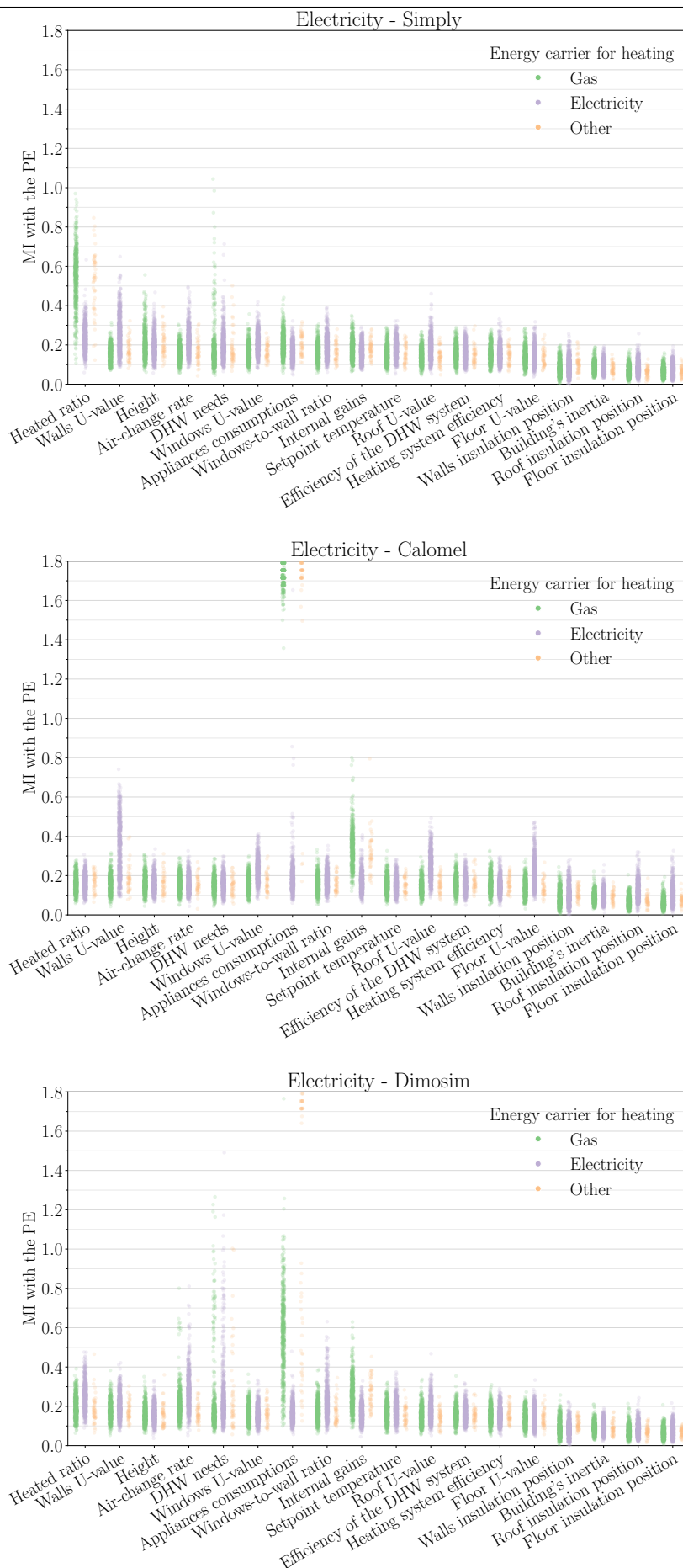


Figure 5.12: Vectors of mutual information between uncertain parameters and percentage of error obtained on electricity consumptions for the buildings of the validation dataset and for three UBEM

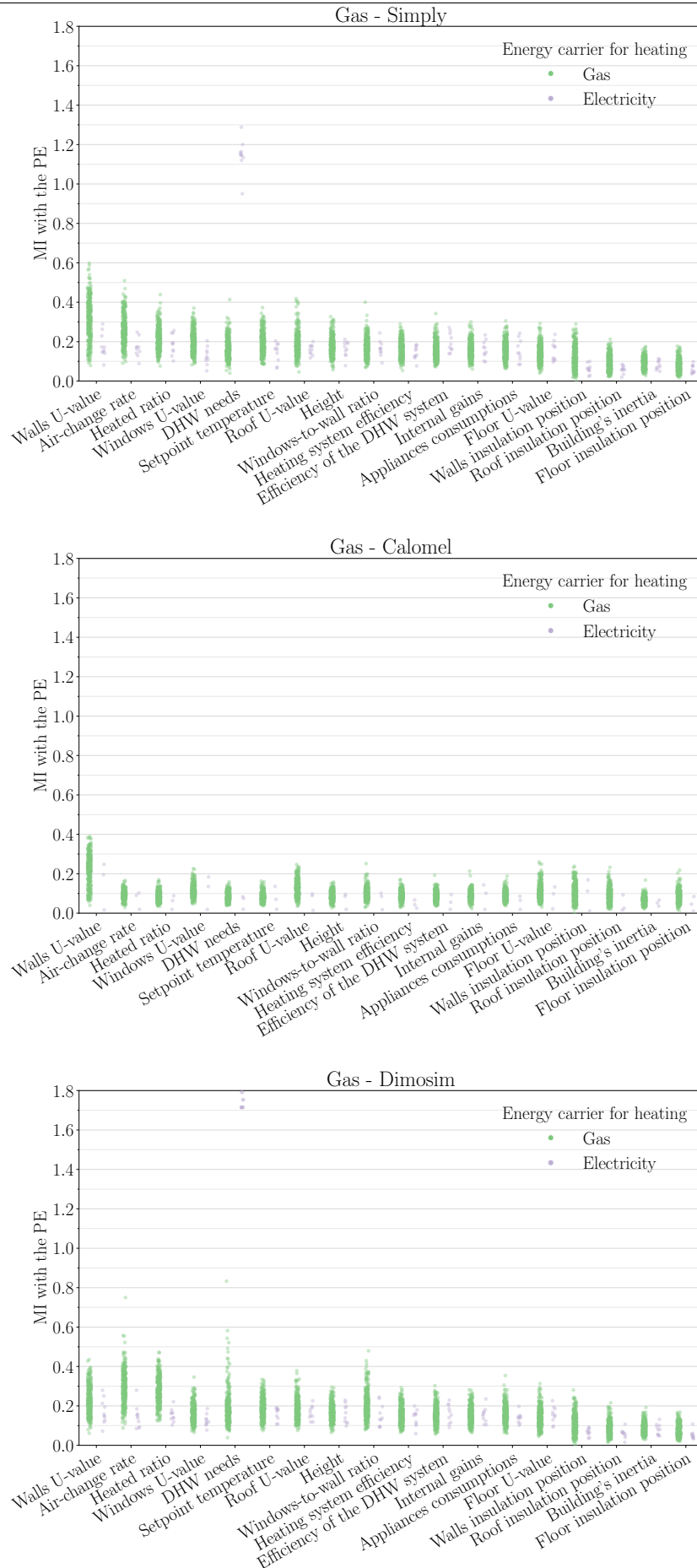


Figure 5.13: Vectors of mutual information between uncertain parameters and percentage of error obtained on gas consumptions for the buildings of the validation dataset and for three

UBEM

Taking a step back, this analysis of the error on both gas and electricity enabled to identify three configurations of buildings that could have a different behaviour towards the error. First, thanks to the PCA applied to the three tools for the error on electricity consumptions, two groups of buildings could be distinguished:

- Buildings for which the appliances consumption drives the error on electricity;
- Buildings for which the thermal needs drive the error on electricity.

Next, with the PCA applied to the error on gas, the category of buildings whose thermal needs were identified as driving the error on energy consumption could be subdivided into two categories:

- Buildings for which the error on the energy consumption is driven by the DHW needs;
- Buildings for which the error on the energy consumption is driven by the heating needs.

This behaviour observed for the gas can also be noticed for the error on electricity once the buildings for which the error is driven by the appliances' consumption have been set aside. For instance, the same order of influencing parameters (DHW needs, air-change rate, heated ratio) is visible in the second eigenvector for electricity and the first eigenvector for gas obtained with Dimosim.

For these three configurations of buildings, the parameters with the greatest influence on the error could be determined. These results are summarized in Table 5.1.

Buildings with error on energy consumption driven by:	Parameters influencing the error on					
	Electricity consumption			Gas consumption		
	Simply	Calomel	Dimosim	Simply	Calomel	Dimosim
Appliances consumption	Heated ratio	Appliances consumption	Appliances consumption	-	-	-
DHW needs	DHW needs	n/a	DHW needs	DHW needs	n/a	DHW needs
Heating needs	Walls U-value	Walls U-value	Air-change rate and Heated ratio	Walls U-value	Walls U-value	Air-change rate and Heated ratio

Table 5.1: Results obtained thanks to the analysis on the error variance

The group of buildings for which the variance of the error was driven by the DHW needs could not be studied for Calomel since this tool does not allow for the parameterisation of DHW needs making it impossible to analyse the uncertainty on this input.

The diagnostic stage applied to the three tools enabled then to identify groups of buildings with a similar behaviour towards the error and to identify influencing input parameters for each

group. Among the reasons indicated in introduction of this chapter that could explain the tools' deviation from reality, this diagnostic stage enabled to rule out at least one of them. Since the identified trends were pretty similar from one tool to another, it indicates that the main error is likely to originate from another source than the physical models. The remaining causes for the error are then:

- The data quality;
- The enrichment process;
- The way uncertainty was accounted for.

It is rather difficult to rule out the data quality since it is difficult to obtain a reference that could enable to characterize the quality level of the data. However, it is possible to study how the enrichment process and the uncertainty modify the tools' behaviour. The impact of these two processes are on the model parameterisation. In order to study the impact of the model parameterisation, the assessment and the diagnostic procedures were applied to the three tools for two other levels of available input data. This work is introduced hereafter.

5.3 Studying the influence of different levels of detail on the tools performances

In this section, the impact of data availability on UBEM performances is discussed. To do so, simulations were run using Simply for three levels of data availability. For each of these levels, an increasing share of information was retrieved from the validation dataset. The information taken into account for each level and for each input parameter is detailed in Table 5.2.

It can be summarized as follows:

- Level 0: no information is retrieved from the EPC available on each building;
- Level 1: the heating and DHW systems of each building are retrieved from the available EPC;
- Level 2: the heating and DHW systems, the roof, walls and windows U-value and the windows-to-wall ratio are retrieved for each building from the available EPC (i.e. there is no uncertainty modelled on these parameters).

The results displayed in the previous sections were obtained for the level 1 of data availability.

The share of uncertainty is then reduced when the level of data availability increases and so is the influence of the enrichment process since more and more parameters are known and thus not set through the enrichment process.

A first analysis of the evolution of the electricity and gas consumptions obtained for the three levels of data availability is discussed below before applying the validation procedure.

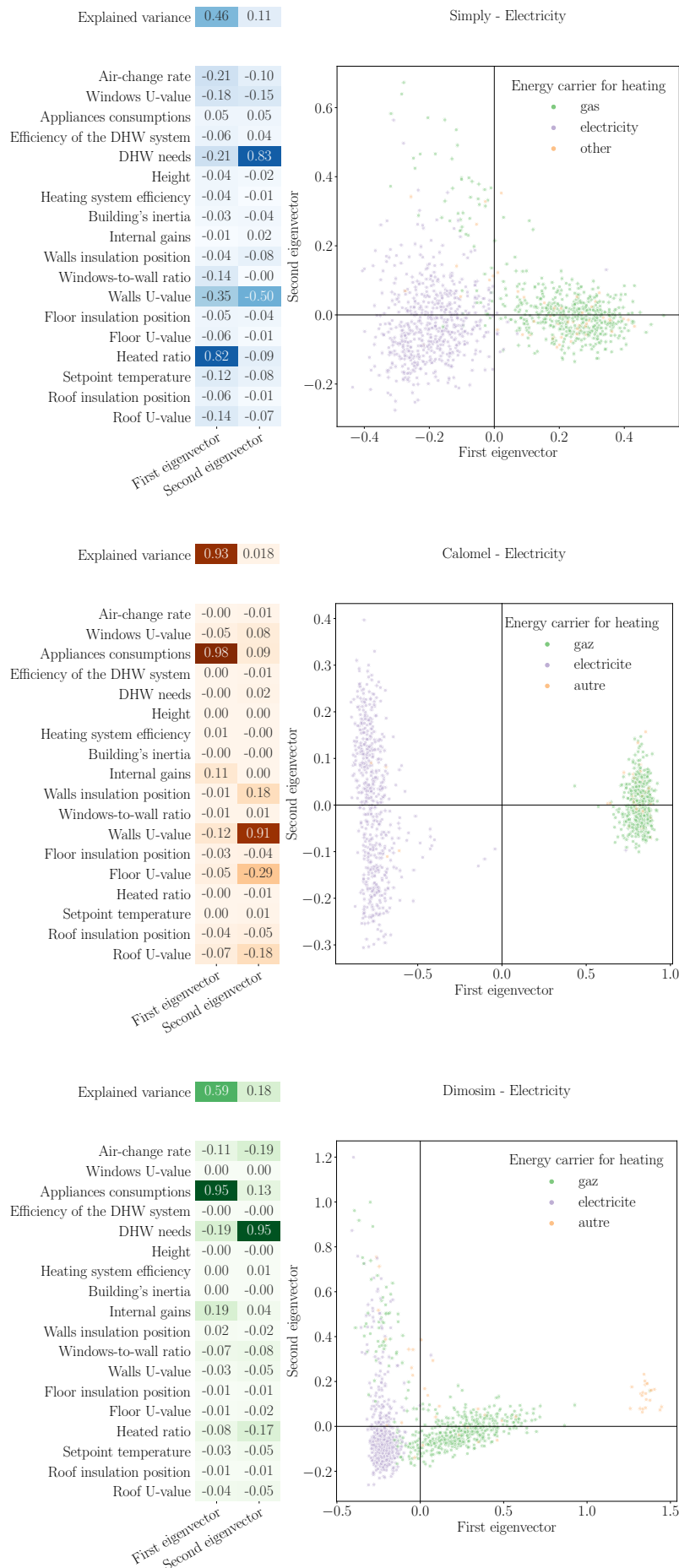


Figure 5.14: Results of principal components analysis applied to vectors of mutual information for the error on electricity consumptions for Simply, Calomel and Dimosim

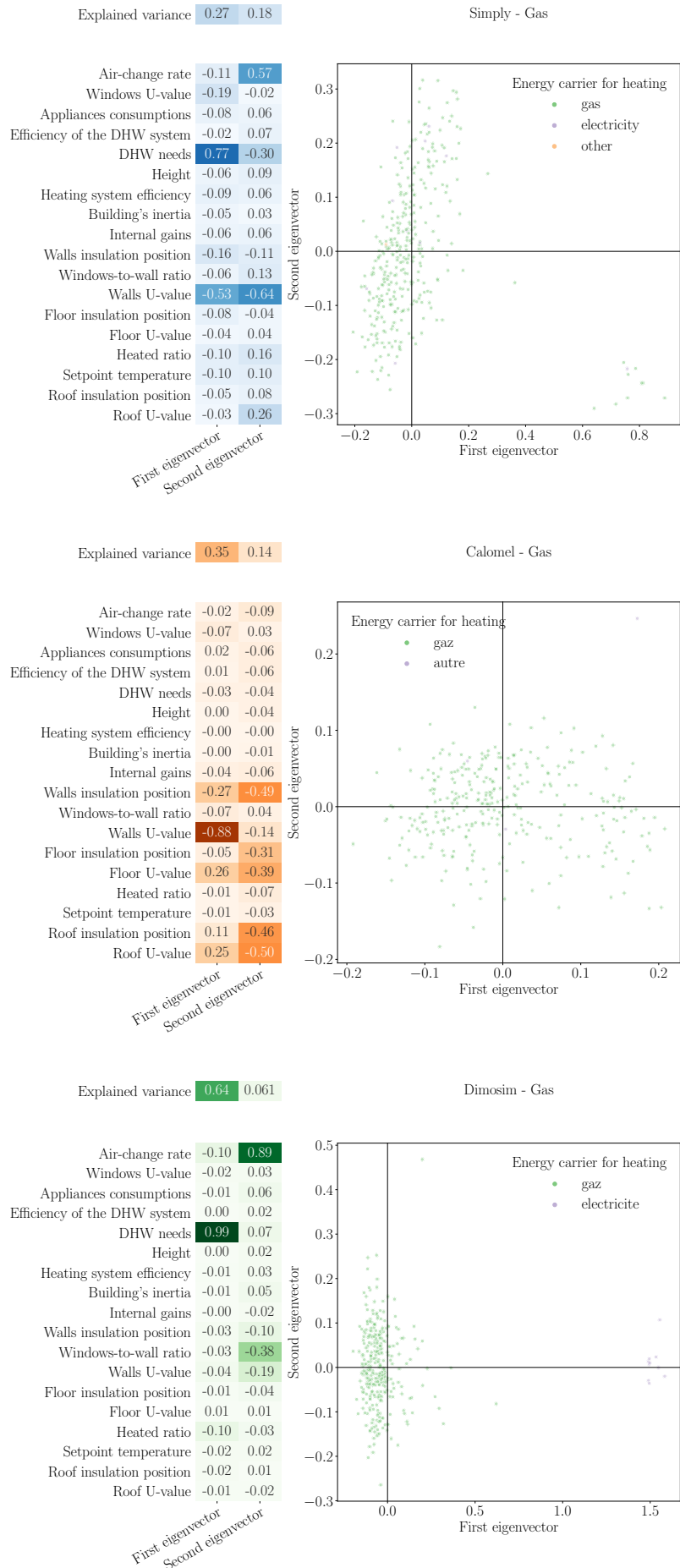


Figure 5.15: Results of principal components analysis applied to vectors of mutual information for the error on gas consumptions for Simply, Calomel and Dimosim

		Level 0	Level 1	Level 2
Thermal properties and systems	Walls U-value	Unknown	Unknown	Retrieved from the building's EPC - no uncertainty
	Roof U-value	Unknown	Unknown	Retrieved from the building's EPC - no uncertainty
	Windows U-value	Unknown	Unknown	Retrieved from the building's EPC - no uncertainty
	Air-change rate	Unknown	Unknown	Unknown
	Set point temperature	Unknown	Unknown	Unknown
	Heating system	Unknown	Retrieved from the building's EPC - no uncertainty	Retrieved from the building's EPC - no uncertainty
	Heating system efficiency	Unknown	Unknown	Unknown
	DHW system	Unknown	Retrieved from the building's EPC - no uncertainty	Retrieved from the building's EPC - no uncertainty
	DHW system efficiency	Unknown	Unknown	Unknown
Building geometry	Roof area	Retrieved from BD TOPO [®] with uncertainty	Retrieved from BD TOPO [®] with uncertainty	Retrieved from BD TOPO [®] with uncertainty
	Windows-to-wall ratio	Unknown	Unknown	Retrieved from the building's EPC - no uncertainty
	Building heated volume	Retrieved from BD TOPO [®] with uncertainty	Retrieved from BD TOPO [®] with uncertainty	Retrieved from BD TOPO [®] with uncertainty
	Total floor area	Retrieved from BD TOPO [®] with uncertainty	Retrieved from BD TOPO [®] with uncertainty	Retrieved from BD TOPO [®] with uncertainty
	Walls area	Retrieved from BD TOPO [®] with uncertainty	Retrieved from BD TOPO [®] with uncertainty	Retrieved from BD TOPO [®] with uncertainty
Weather data	Outdoor air temperature	Retrieved from weather file - no uncertainty	Retrieved from weather file - no uncertainty	Retrieved from weather file - no uncertainty
	Solar gains	Retrieved from weather file - no uncertainty	Retrieved from weather file - no uncertainty	Retrieved from weather file - no uncertainty
Others	DHW needs	Unknown	Unknown	Unknown
	Appliances electricity consumption	Unknown	Unknown	Unknown
	Internal gains	Unknown	Unknown	Unknown

Table 5.2: Known and uncertain parameters for the three levels of data availability

5.3.1 Evolution of the simulated consumptions with regard to data availability

In this section, a first outlook of the simulated consumptions obtained with Simply for each building and for each level of data availability is discussed. To do so, the distribution of the mean and of the standard deviation of the consumptions obtained for each building and for each level of data are displayed Figure 5.16.

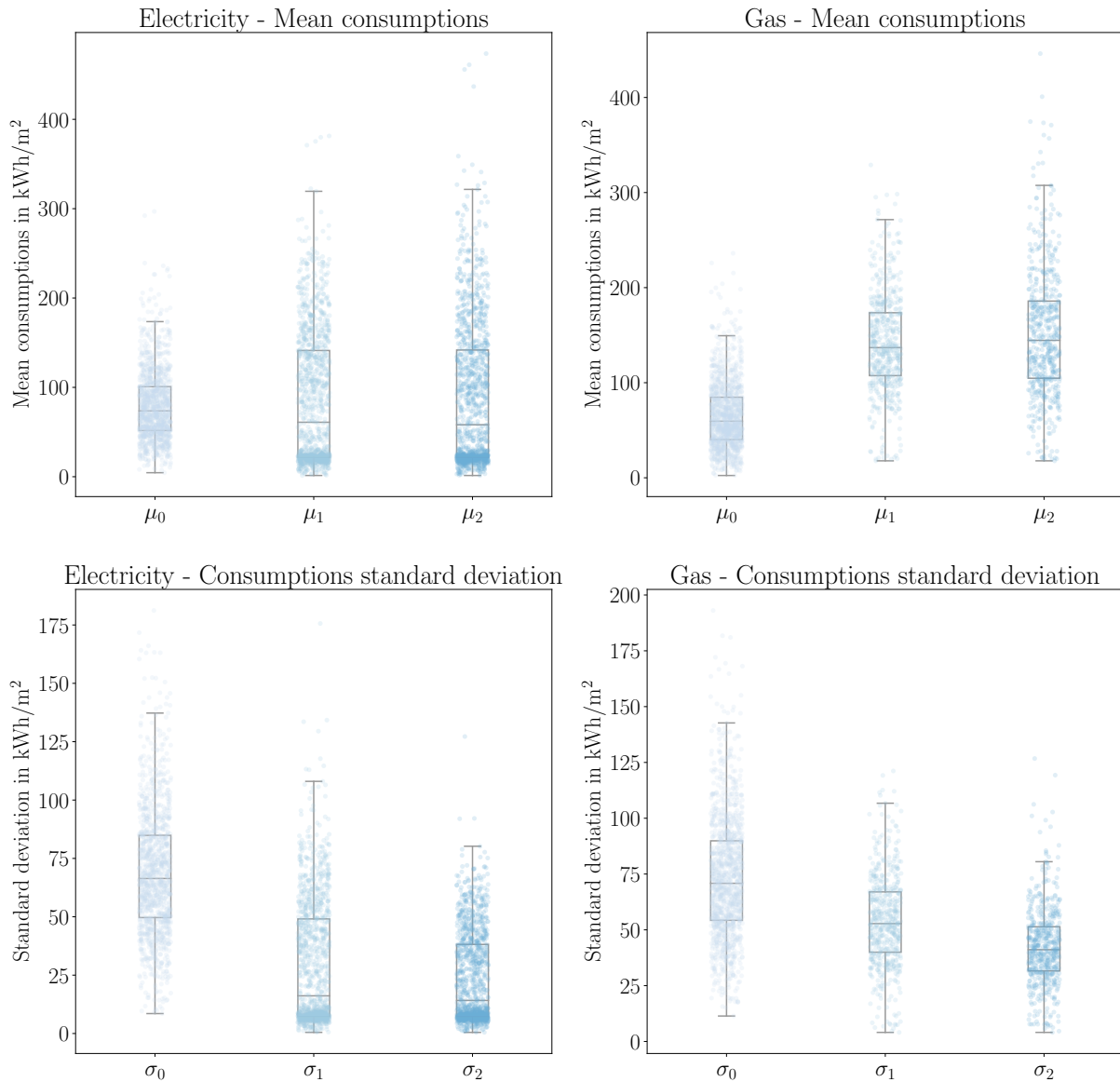


Figure 5.16: Evolution of the mean and standard deviation of the consumptions obtained for each building between the three levels of data availability

First, as expected, the standard deviation of the consumptions obtained on each building decreases rapidly when more information is added. This is reassuring regarding the quality of the uncertainty quantification.

Then, it can be noticed that the distribution of the mean consumptions is more spread out when shifting from the level 0 of data availability to the level 1. This phenomenon is a natural consequence of the knowledge or lack of knowledge of the heating and DHW systems and highlights the proper modelling of uncertainty on these two parameters. It is easier to understand when

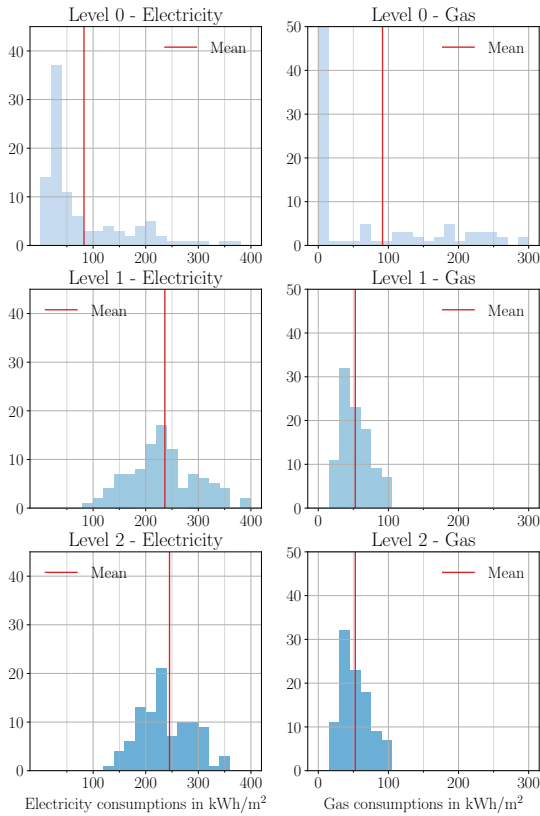


Figure 5.17: Evolution of the gas and electricity consumptions obtained for a building heated with electricity

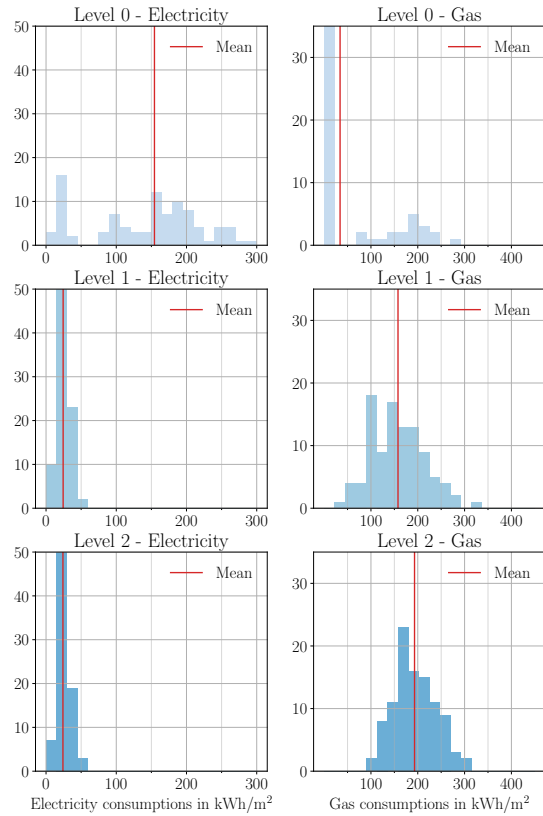


Figure 5.18: Evolution of the gas and electricity consumptions obtained for a building heated with gas

considering the energy consumptions displayed by two buildings, one heated with electricity and one heated with gas. As displayed Figure 5.17, the mean of the electricity consumptions obtained for the building heated with electricity increases when switching from level 0 to level 1. Similarly, the mean of the gas consumptions decreases for this same building. This is due to the fact that for level 0, the heating and DHW systems are unknown. Thus, for some configurations, the buildings will have electric heater and for other configuration, it will have a gas boiler. The average of the electricity consumptions on all the configurations is then naturally lower than in the case where all the configurations are set with a electric heater. Conversely, it can be seen Figure 5.18, that the mean electricity consumption for a building heated with gas decreases whereas its mean gas consumption increases when switching from level 0 to level 1. The differences of the mean consumptions between level 0 and level 1 for these two buildings explain, when generalized to the whole sample, the fact that distribution of mean consumptions appears to be more spread out in level 1 than in level 0.

Looking back at Figure 5.16, the distribution of mean consumptions obtained at level 2 seems to be shifted towards the high consumptions compared to the distribution obtained at level 1. The difference between the two levels lies in the thermal properties: in level 1, the walls, roof and windows U-value are not fixed and a probable set of values is sampled from a Bayesian network trained on the database of EPC for each building configuration whereas in level 2, these values are kept fixed for all the configurations of a given building. This difference challenges then the enrichment process used here. When comparing the two distributions of mean consumptions, they should be similar as long as the training set used to train the Bayesian network is similar to the validation dataset. The two datasets are compared in Figure 5.19. A bias is visible with an over-representation of old buildings in the validation dataset compared to the training set.

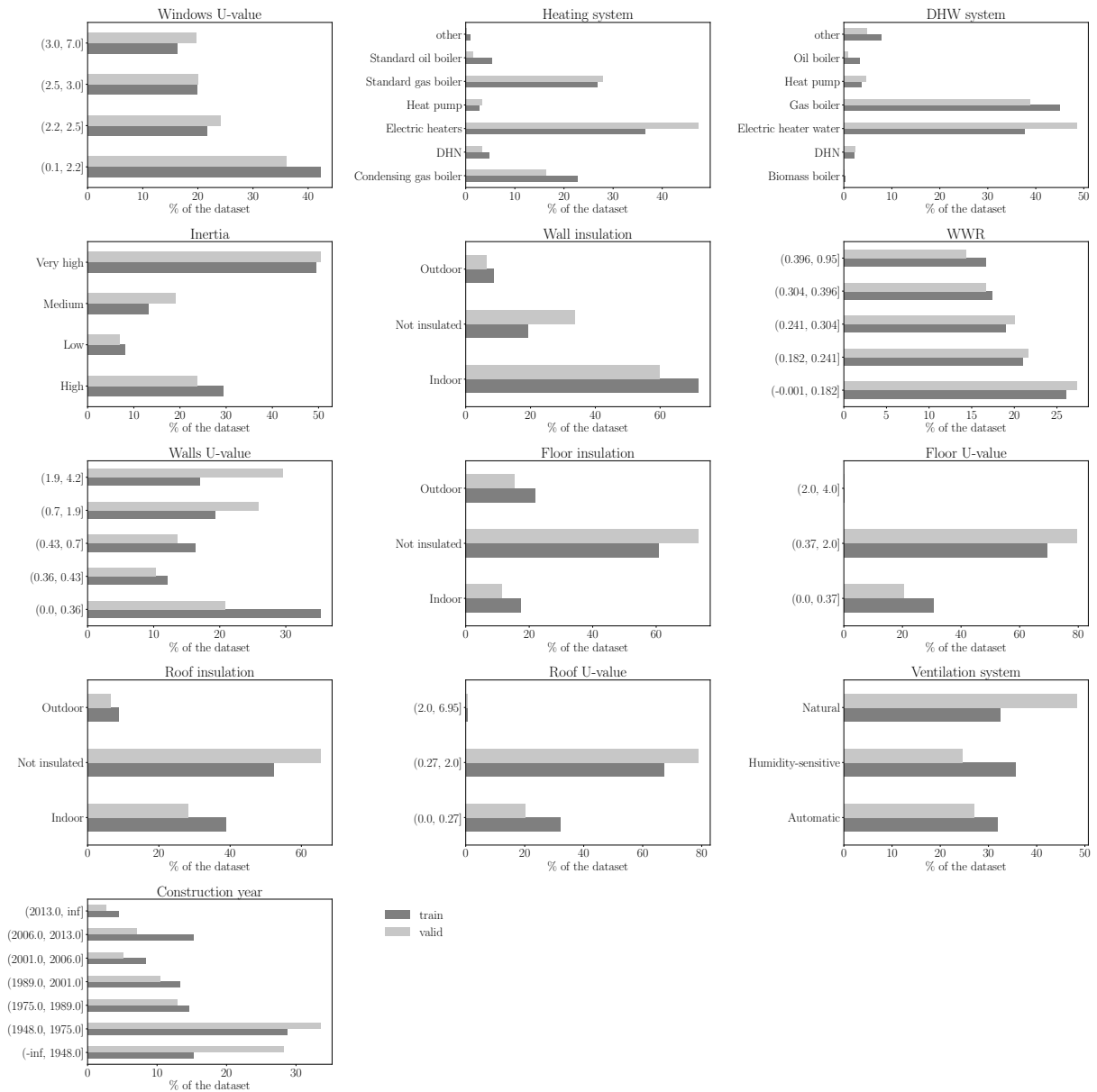


Figure 5.19: Comparison between the validation dataset and the training set used to train the Bayesian network

This difference induces more poor values of U-values at level 2 compare to level 1 which explains higher mean consumptions for the level 2.

Now that the evolution of the consumptions obtained by Simply on the three levels of data availability has been discussed, the assessment procedure can be applied to each set of simulations. The variation of performances displayed by Simply from one level of data availability to another could then be assessed. The conclusions of this study are introduced in the following section.

5.3.2 Assessment stage applied to three levels of data availability

The metrics introduced in Chapter 4 were applied here to the consumptions obtained with Simply for the three levels of detail. The distributions of the residuals obtained for the three levels on electricity and gas are displayed respectively in Figure 5.20 and Figure 5.21.

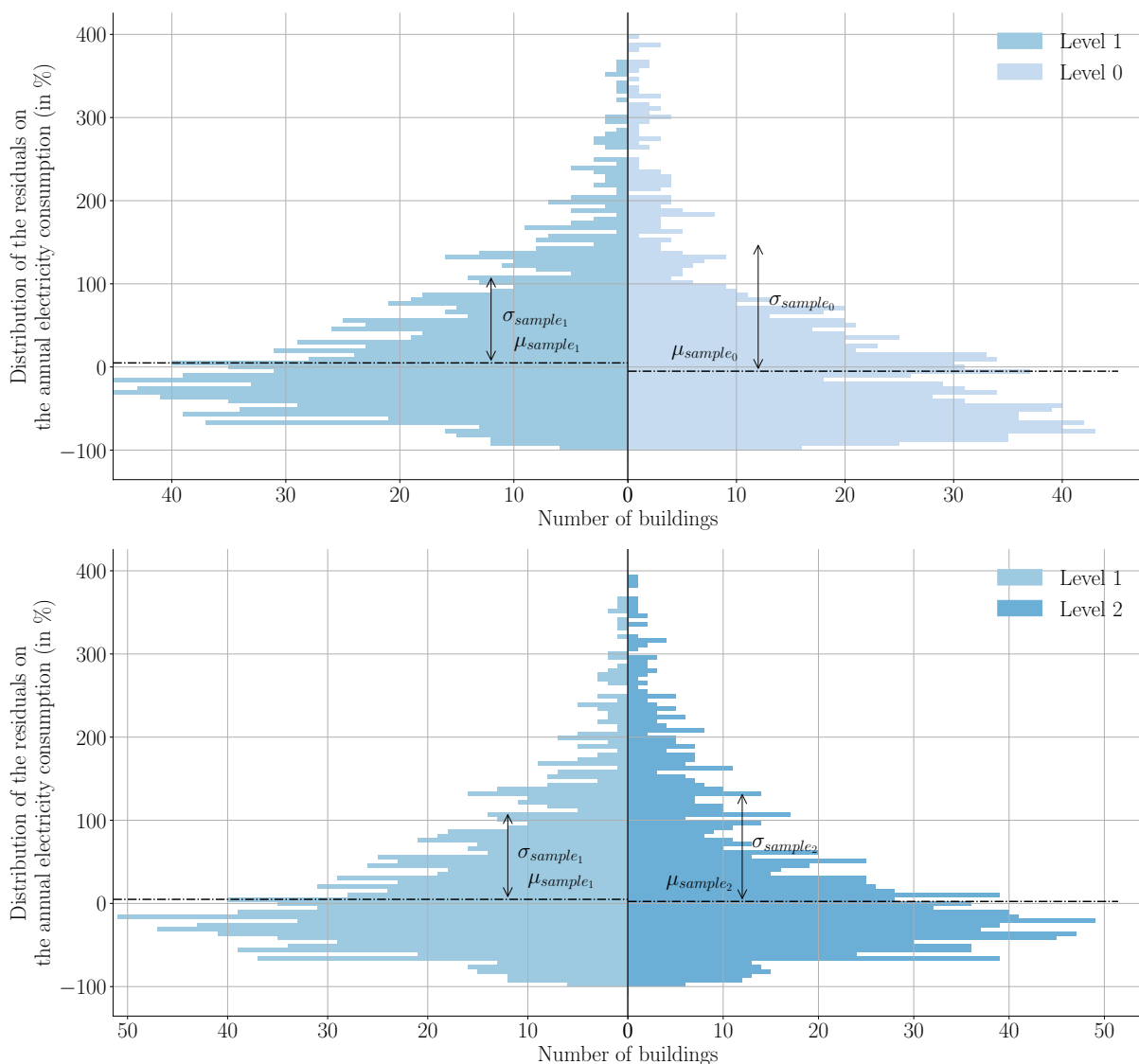


Figure 5.20: Distributions of residuals obtained with Simply for the three levels of data availability on electricity

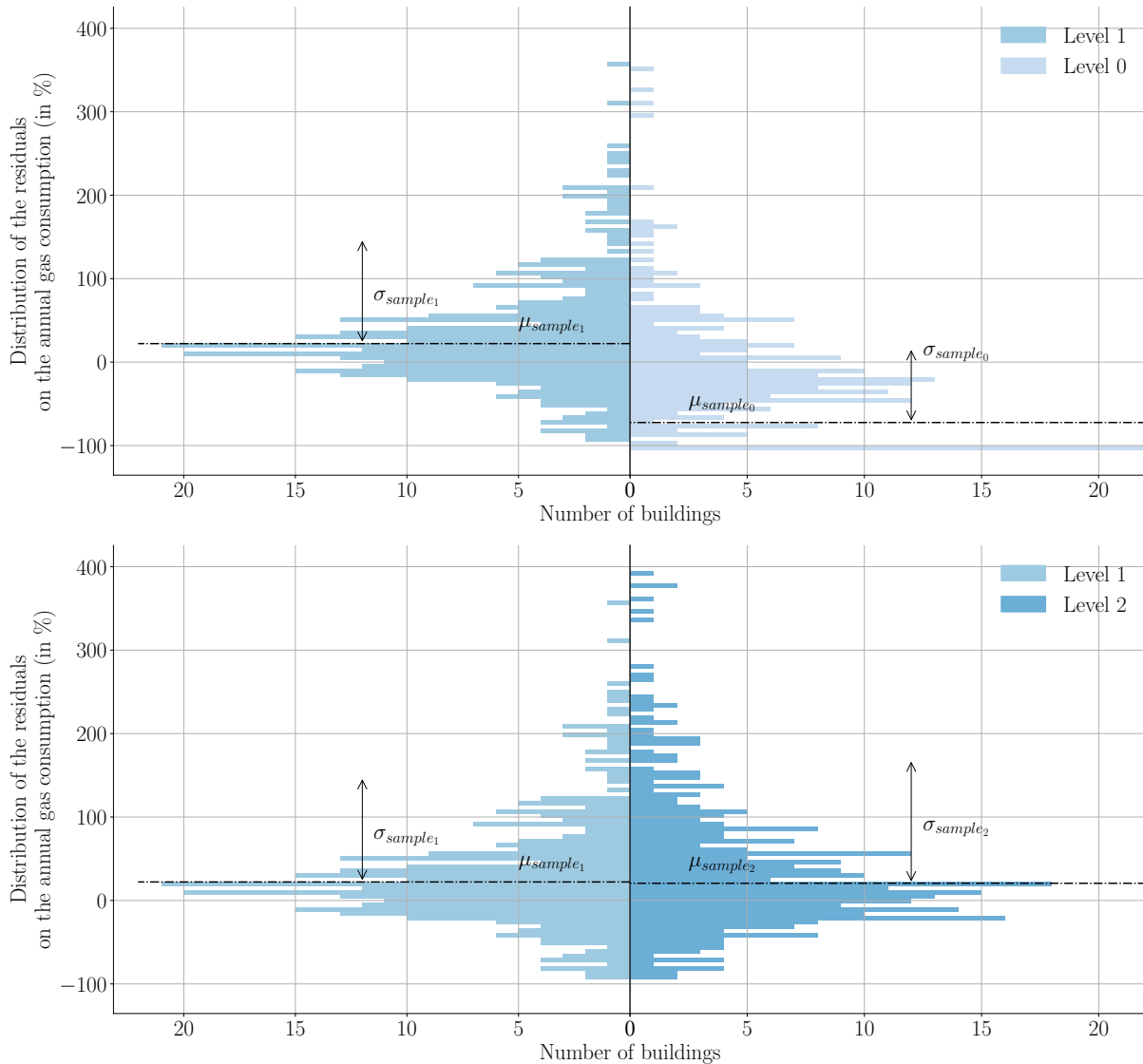


Figure 5.21: Distributions of residuals obtained with Simply for the three levels of data availability on gas

Looking at the distributions of residuals obtained for electricity, the three distributions are almost centred around zero. It seems that the quantity of available data has then no impact on the median of these distributions, μ_{sample} . However, the standard deviation of these same distributions, σ_{sample} , seems to be more sensitive to the data availability: it decreases between level 0 and level 1, but increases between level 1 and level 2. This behaviour is quite surprising, it would have been expected for the standard deviation to decrease gradually when the quantity of available data increases. On the contrary, it is observed here an increase of the standard deviation between level 1 and level 2, meaning that there are more buildings with a high residual when the thermal properties are retrieved directly from the EPC available for each building than when uncertainty is modelled on thermal properties. Two main reasons could explain this phenomenon:

- The EPC provide a good picture of the building stock but taken individually, each EPC struggles to adequately describe the building it is supposed to assess.
- By fixing the thermal properties and thus reducing the modelled uncertainty, another bias is revealed that was not visible before.

It is at the moment difficult to arbitrate between these two possibilities and both should be kept under the spotlights for the rest of the analysis.

When it comes to the residual distributions for gas, a clear improvement of the bias, μ_{sample} , is visible between level 0 and level 1, and it stabilizes between level 1 and level 2. Regarding the standard deviation, it steadily increases from level 0 to level 2. This trend would advocate in favour of the second reason (another bias) explaining the behaviour of standard deviation on electricity: with more parameters set to a fixed value, a bias becomes more visible. The argument behind is that the difference between level 0 and level 1 is the heating and DHW systems which are known for each building in level 1. The heating and DHW systems are the parameters that should be the most reliable among all the data provided by the EPC since these systems are reviewed by the auditors making the EPC of the building. An increase in standard deviation between level 0 and level 1 is then more likely to originate from a bias being revealed rather than from issues on all EPC regarding the buildings heating and DHW systems.

Since the level of data availability changes the number of parameters that are deemed uncertain, it is interesting to compare the HPD region obtained by the buildings between the three levels. This comparison is shown Figure 5.22. As it was done in Chapter 4, each point represents the median of the computed consumptions for a building plotted against its DLE and the vertical line displays the associated HPD region. The buildings for which the DLE lies within the HPD region (i.e. the vertical line intersects the first bisector) are set in blue whereas buildings for which the DLE is outside the HPD region are set in red.

It appears on this figure that the number of buildings for which the DLE is outside the HPD region increases when switching from level 0 to level 1 and from level 1 to level 2. In the meantime, it seems that the simulated consumptions have a stronger correlation with the actual consumption provided by the DLE when the level of data availability increases. For instance, when looking at the graph displaying HPD regions for electricity with a level 0 of data availability, the electricity consumptions computed with Simply seem to have the same distribution independently of the real consumptions provided by the DLE. However, when considering the same graph for level 2, a bias is clearly visible, but it also seems that the simulated consumptions increase when the consumptions provided by the DLE increase.

To properly quantify this phenomenon, the mutual information between the median simulated consumptions and the DLE has been computed for electricity and gas for each level of data availability. The results are presented in Table 5.3

	MI - Electricity	MI - Gas
Level 0	0.03	0.04
Level 1	0.20	0.07
Level 2	0.19	0.09

Table 5.3: Mutual information between measurements and simulated consumptions for the three levels of data availability

A steady increase of the mutual information between the simulated and measured gas consumptions can be noticed. Considering electricity, a huge increase in mutual information is visible between level 0 and level 1, but it then deteriorates between level 1 and level 2. This means that the uncertainty is so high at level 0 that it completely dominates the simulated consumptions making the tool unable to adequately simulate each building. The buildings all show a very large variation in simulated consumptions, which artificially increases the number of buildings catching the DLE within their HPD region, but this shows that with so little information available, the modellers should never try to simulate consumptions at building scale without taking

uncertainty into account, as a single simulation for each building would most likely be wrong.

This conclusion is reinforced by the empirical coverage computed for the three levels of data availability and presented Figure 21.

The empirical coverage obtained for the level 0 for electricity displays a value above the first bisector for the confidence level of 80% which indicates a complete domination of the results by the uncertainty. As expected after reviewing Figure 5.22, the empirical coverage is also better for level 1 than for level 2.

In order to gain a better understanding of the differences between the three levels of data availability that have been assessed here, the diagnostic stage is applied to the three sets of simulation in the following section.

5.3.3 Diagnostic stage applied to three levels of data availability

5.3.3.1 Identification of parameters influencing the error at the sample scale

As it was done previously, the diagnostic stage starts with the identification of the parameters that influence the most the error at the sample scale. The mutual information between the residuals and characteristics of each building was computed for the three levels of data availability and the results are presented Figure 5.24.

The heating and DHW systems stand out as influencing parameters for both level 1 and level 2. However, since these two parameters were deemed uncertain at level 0, their influence on the simulations at level 0 cannot be assessed here. It can be noticed that no other parameter takes the lead at level 0 to explain the residuals. It is not surprising given the homogeneity of the consumptions distributions noticed in Figure 5.22.

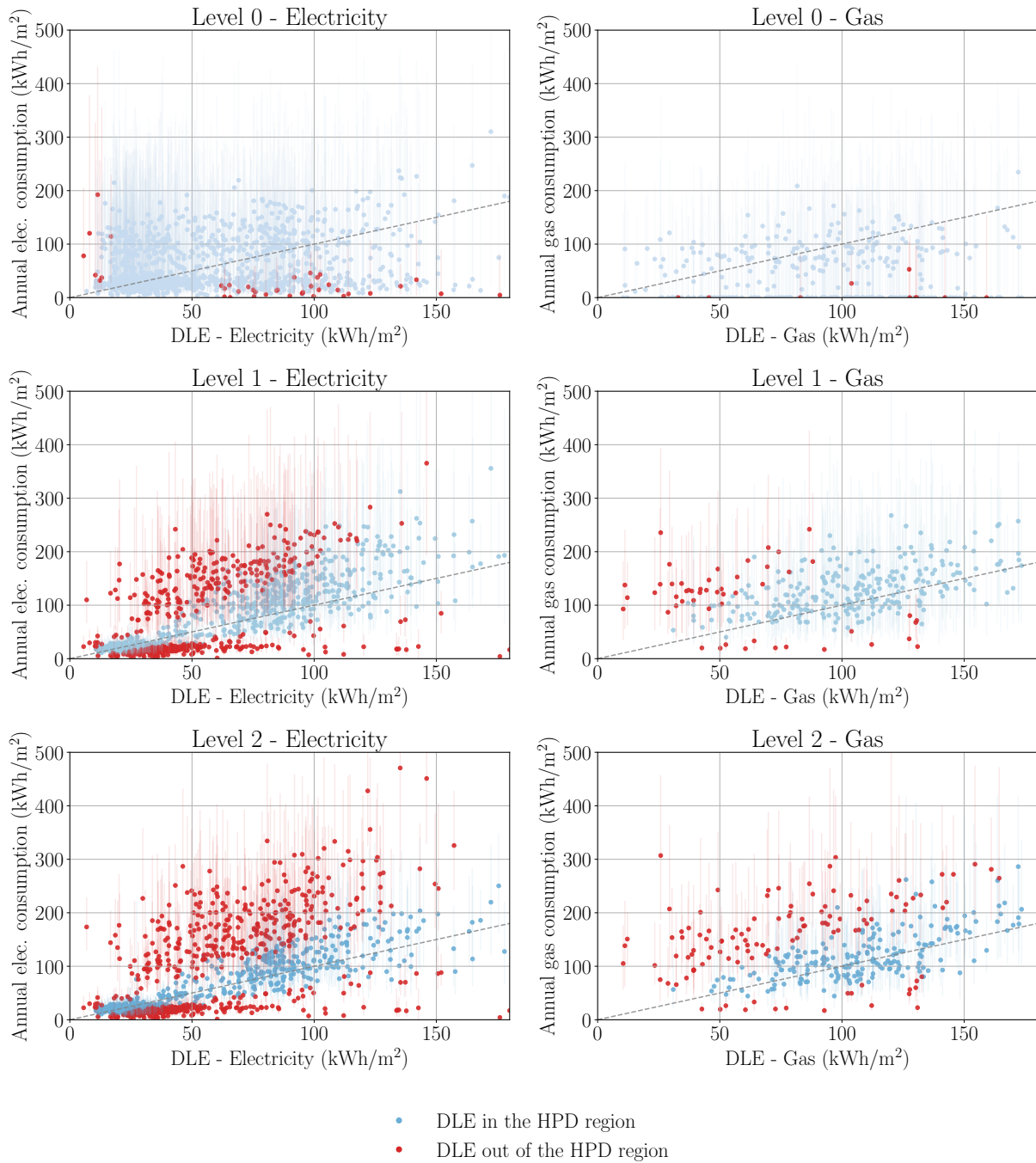


Figure 5.22: μ_{bldg} and the HPD region of each building against its DLE for three levels of data availability

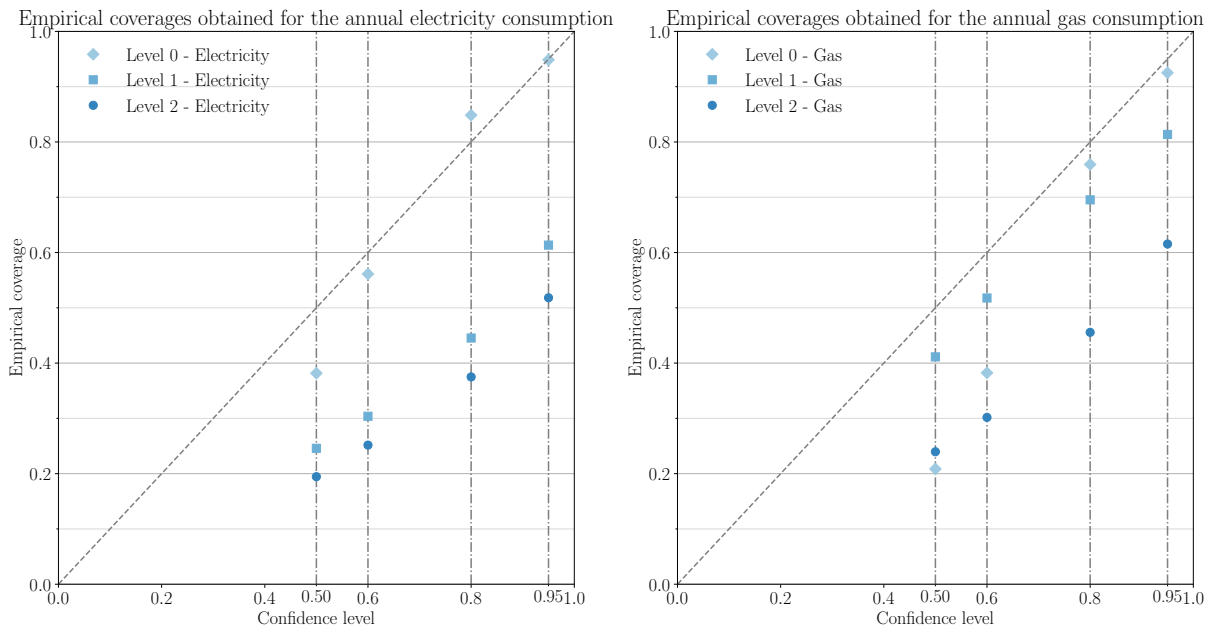


Figure 5.23: Evolution of the empirical coverage depending on the data availability

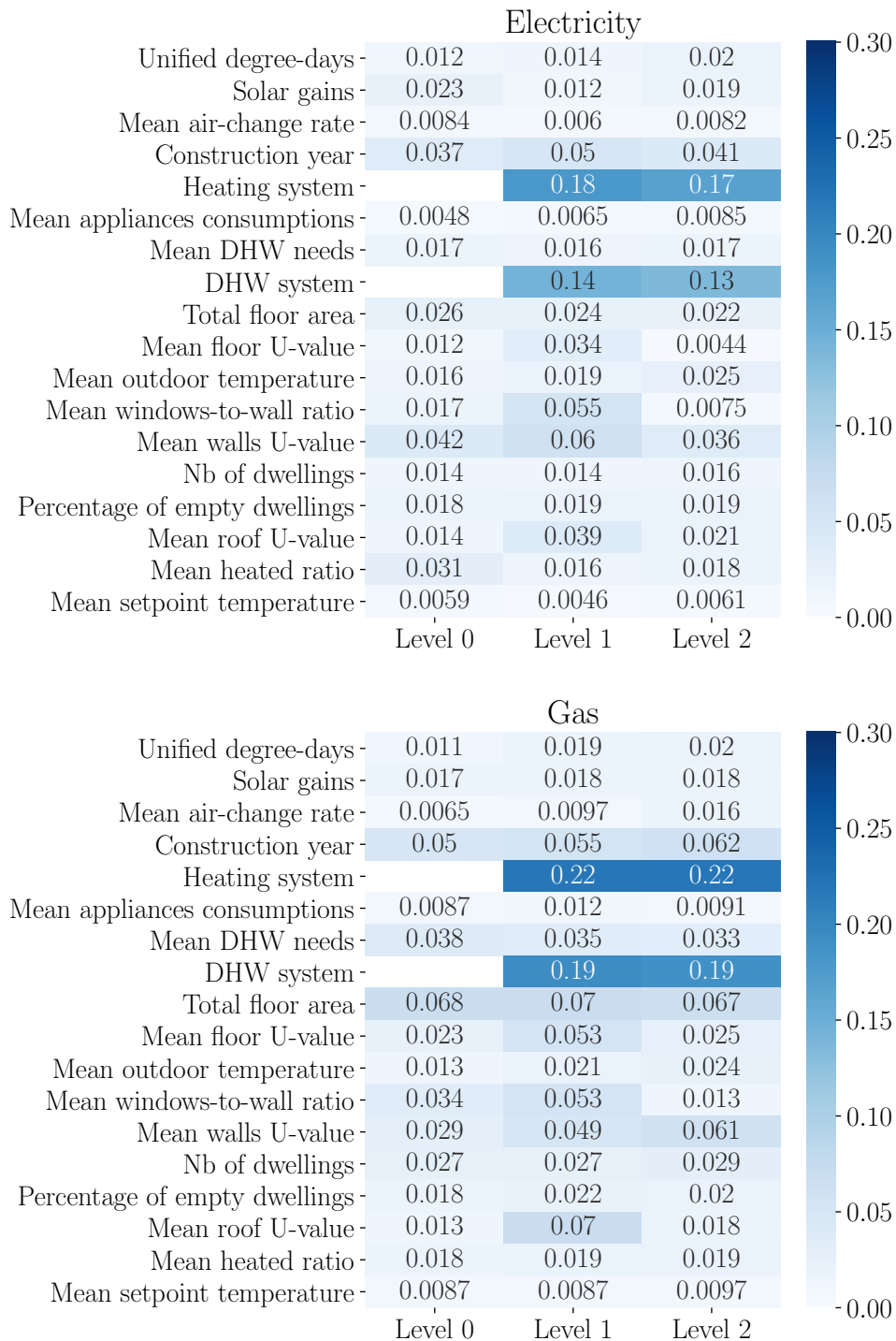


Figure 5.24: Mutual information between the residual and building's characteristics for three levels of data availability

5.3.3.2 Identification of parameters influencing the error at building scale

This first analysis on the residual is followed by an analysis of the variance of the error at building scale. Vectors of mutual information were computed for each building between the percentage of error and the uncertain parameters. These vectors are depicted for the three levels of details and for gas and electricity on Figure 5.25 and Figure 5.26.

The colour scheme in this case is based on the real heating energy carrier of each building. Looking at the vectors obtained for level 0, it appears clearly that buildings display the same kind of mutual information independently of their heating energy carrier. This makes sense as this parameter was deemed unknown at this level. However, it can be noticed for this same level that all buildings display a very high value for mutual information between their percentage of error and the heating system and between the percentage of error and DHW system. In this case too, the heating and DHW energy carriers drive the error. Comparing level 1 and level 2, there is no clear difference at this point. Influencing parameters observed for level 2 seems to be in the same ranking as they were for level 1.

5.3.3.3 Identification of buildings with a similar behaviour towards the error

To verify the trends highlighted above, a PCA was run for each level of data availability and the results are introduced hereafter in Figure 5.27 and in Figure 5.28.

It can be noticed immediately that the results of the PCA are very different in level 0 from what they are for level 1 and level 2. In levels 1 and 2, groups of buildings based on the heating energy carrier are clearly visible whereas the buildings seem more randomly spread out in the plane for level 0. Studying the eigenvectors, the heating and DHW systems stand out as the most influencing parameters in the first eigenvectors for gas and electricity obtained at level 0 by Simply. However, the parameters identified as influencing in section 5.1.2.2, the DHW needs, the walls U-value and the heated ratio, are noticeable in the second eigenvector for electricity of level 0. On the contrary, they do not stand out at all in the eigenvectors obtained at level 0 for the gas. It is interesting to see that they appear then clearly even in the eigenvectors obtained at level 1, once the heating and the DHW systems have been fixed. Similarly, the walls U-value is kept at a fixed value in level 2 which reveals two other influencing parameters behind the DHW needs and the heated ratio: the set point temperature and the air-change rate.

These parameters could then explain the bias highlighted in Figure 5.22 for level 2. Having a better knowledge of the DHW needs, the appliances consumptions, the set point temperature and the air-change rate could then improve the results obtained for level 2.

To strengthen this insight, the same work was led on Calomel and Dimosim, and the results are displayed in the following section.

5.3.4 Comparing the influence of different levels of details on three tools' performances

In this part, the aim is not so much to study each tool's results in depth, but rather to analyse whether the trends identified above are confirmed when applying the same procedure to other tools. For this reason, not all figures related to Calomel and Dimosim's results will be introduced here, but the reader who would like to study in detail the results of each of these tools will find all the required figures in the Appendix.

First, the HPD region along with the residual obtained by Dimosim and Calomel for each building are plotted against their DLE for the three levels of data availability in Figure 5.29 and Figure

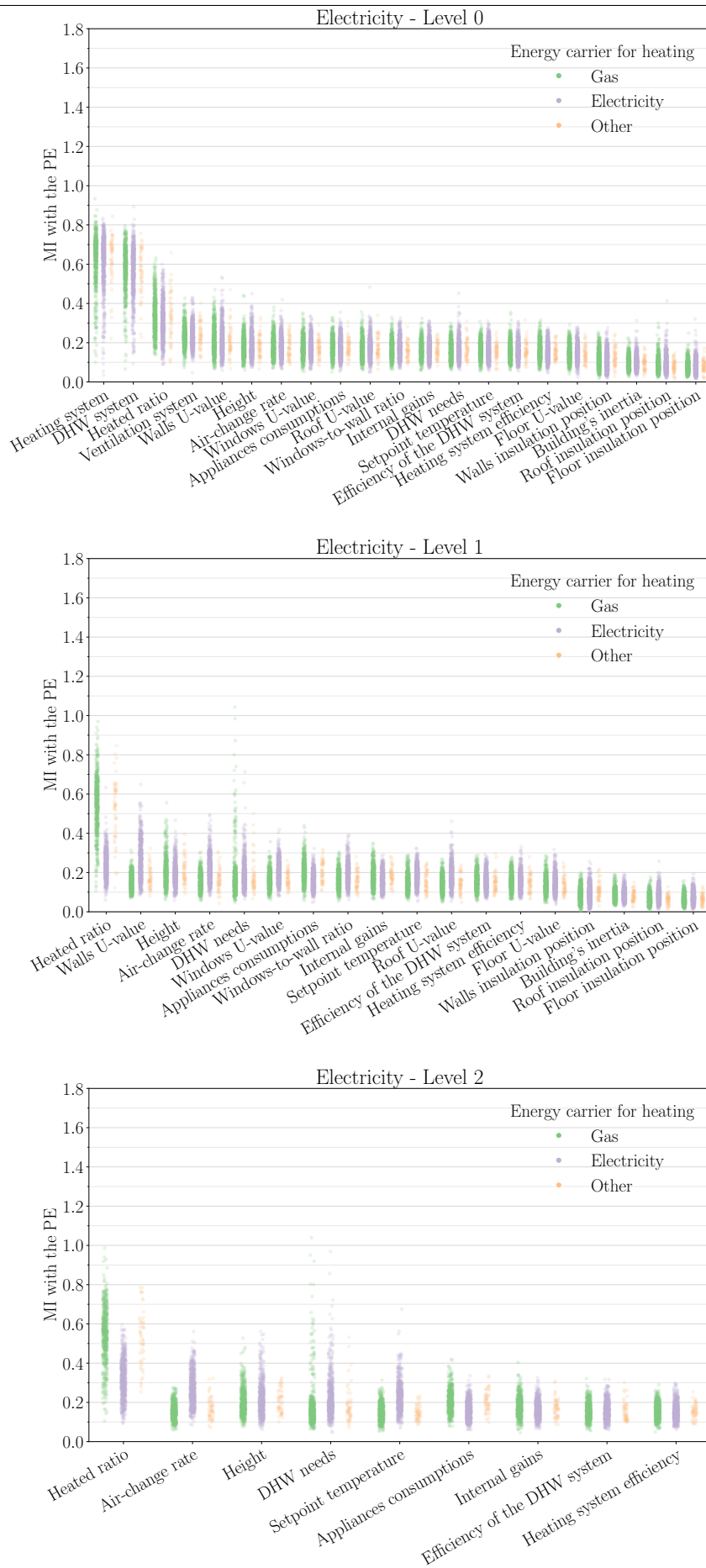


Figure 5.25: Vectors of mutual information between uncertain parameters and percentage of error on electricity consumption obtained for the buildings of the validation dataset and for three levels of data availability

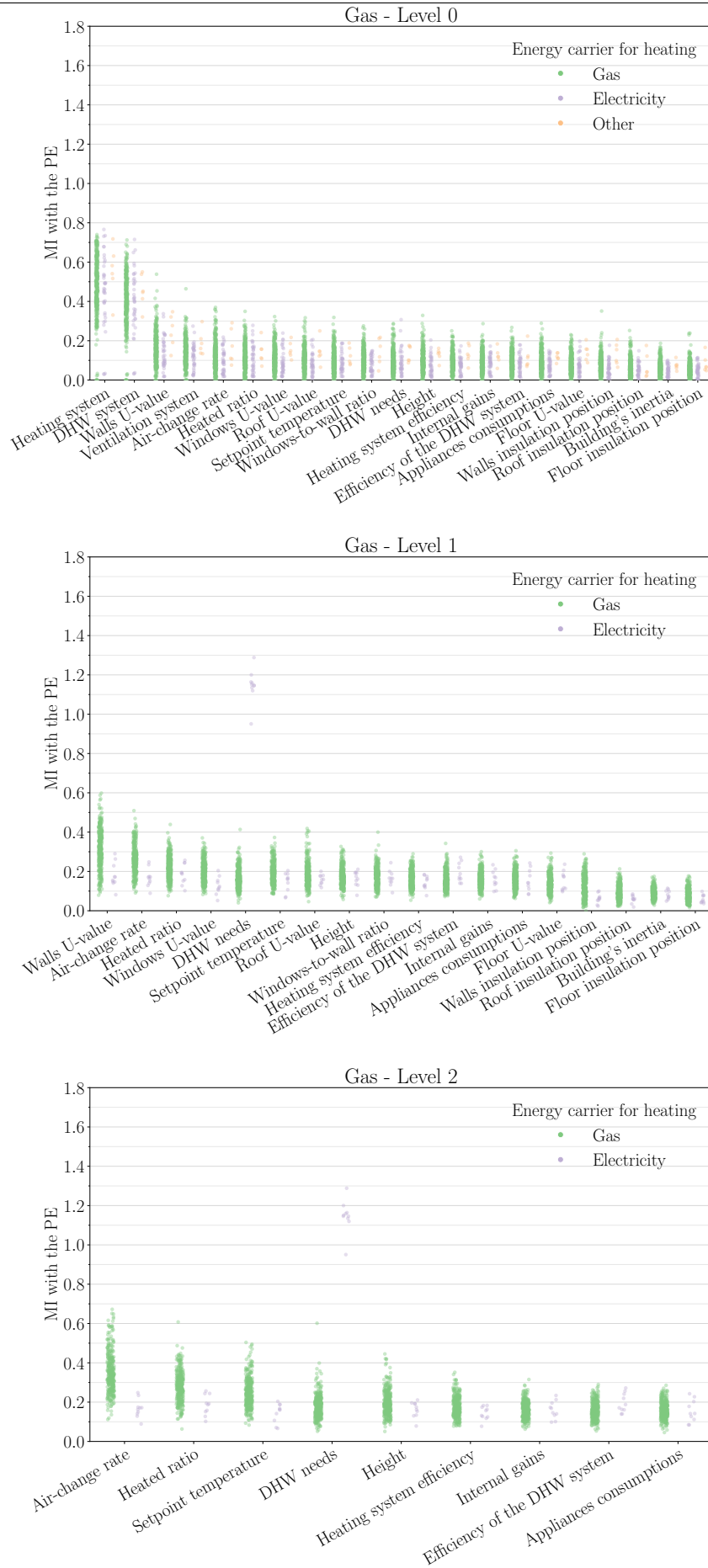


Figure 5.26: Vectors of mutual information between uncertain parameters and percentage of error on gas consumption obtained for the buildings of the validation dataset and for three levels of data availability

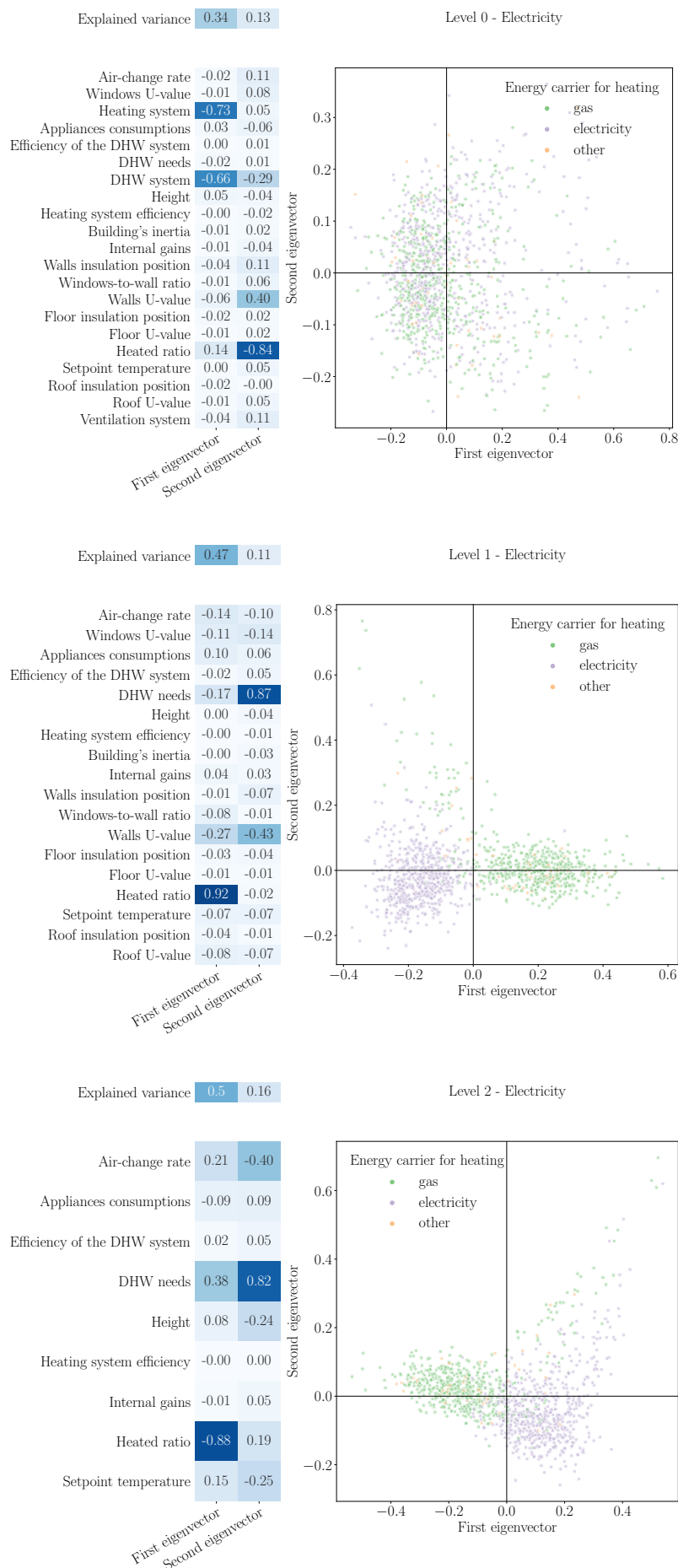


Figure 5.27: Results of principal components analysis applied to vectors of mutual information for the error on electricity consumption for three levels of data availability

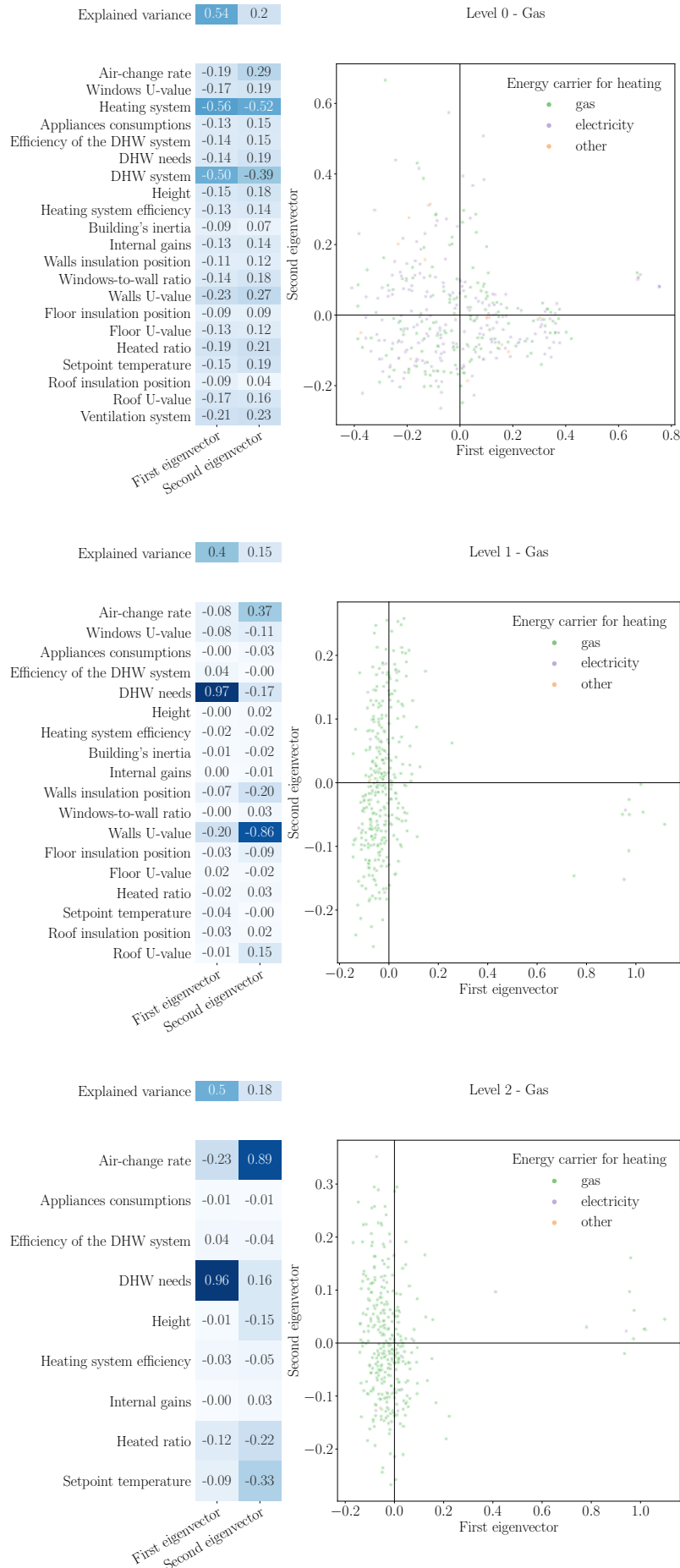


Figure 5.28: Results of principal components analysis applied to vectors of mutual information for the error on gas consumption for three levels of data availability

5.30. The apparition of a bias, becoming stronger as the level of data availability increases, is immediately visible for Calomel whereas the trend is less distinct for Dimosim. Since fewer parameters are considered uncertain in Calomel's simulations, the HPD region displayed by each building is significantly smaller than the one displayed in the case of Dimosim. Consequently, the number of buildings for which the HPD region does not intersect with the DLE is higher for Calomel than for Dimosim. It is especially visible for the gas for the level 2 of data availability where almost all the buildings simulated with Calomel display a DLE outside the HPD region. In this case, uncertainty was modelled on only three input parameters for Calomel while it was modelled on nine parameters for Dimosim.

Mutual information between residuals and buildings characteristics was then computed for both Calomel and Dimosim, and results are displayed respectively Figure 5.31 and Figure 5.32.

In both cases, there is no significant parameters standing out to explain the residuals obtained on electricity. However, the heating and the DHW systems both are significant when it comes to the residual on gas for levels 1 and 2. Regarding level 0 of data availability, no parameter appears to have an influence on the residual. As it was observed for Simply, the HPD region and the residual obtained with Dimosim for each building seem to be completely independent of the actual measured consumption of the building which explains that no significant parameters was identified here. However, when looking at the values obtained with Calomel for electricity, two groups of buildings are to be distinguished: a group of buildings with a residual around 25 kWh/m^2 and another group with residuals contained in the interval $[30 \text{ kWh/m}^2, 200 \text{ kWh/m}^2]$. This phenomenon comes from the way the uncertainty was modelled on the heating and DHW systems. These systems were predicted from a Bayesian network for each configuration of each building, based on what was known from the building (in this case, the construction year). The probability distribution of having heating system fuelled with gas or with electricity is then slightly different from a building to another depending on the construction year. Some buildings have then a higher probability to have a heating system fuelled with electricity than to have a heating system with a different energy carrier. These buildings are then the ones with a residual contained in the interval $[30 \text{ kWh/m}^2, 200 \text{ kWh/m}^2]$. It is most likely that this phenomenon would be visible for Simply and Dimosim as well if their bias was as important as the one shown by Calomel.

Since the study on the residuals proved to be insufficient to explain the error at level 0, vectors of mutual information were computed between the percentage of error of each building and its uncertain parameters for both Calomel and Dimosim. The results are shown respectively Figures 5.33 and 5.34 and Figures 5.35 and 5.36.

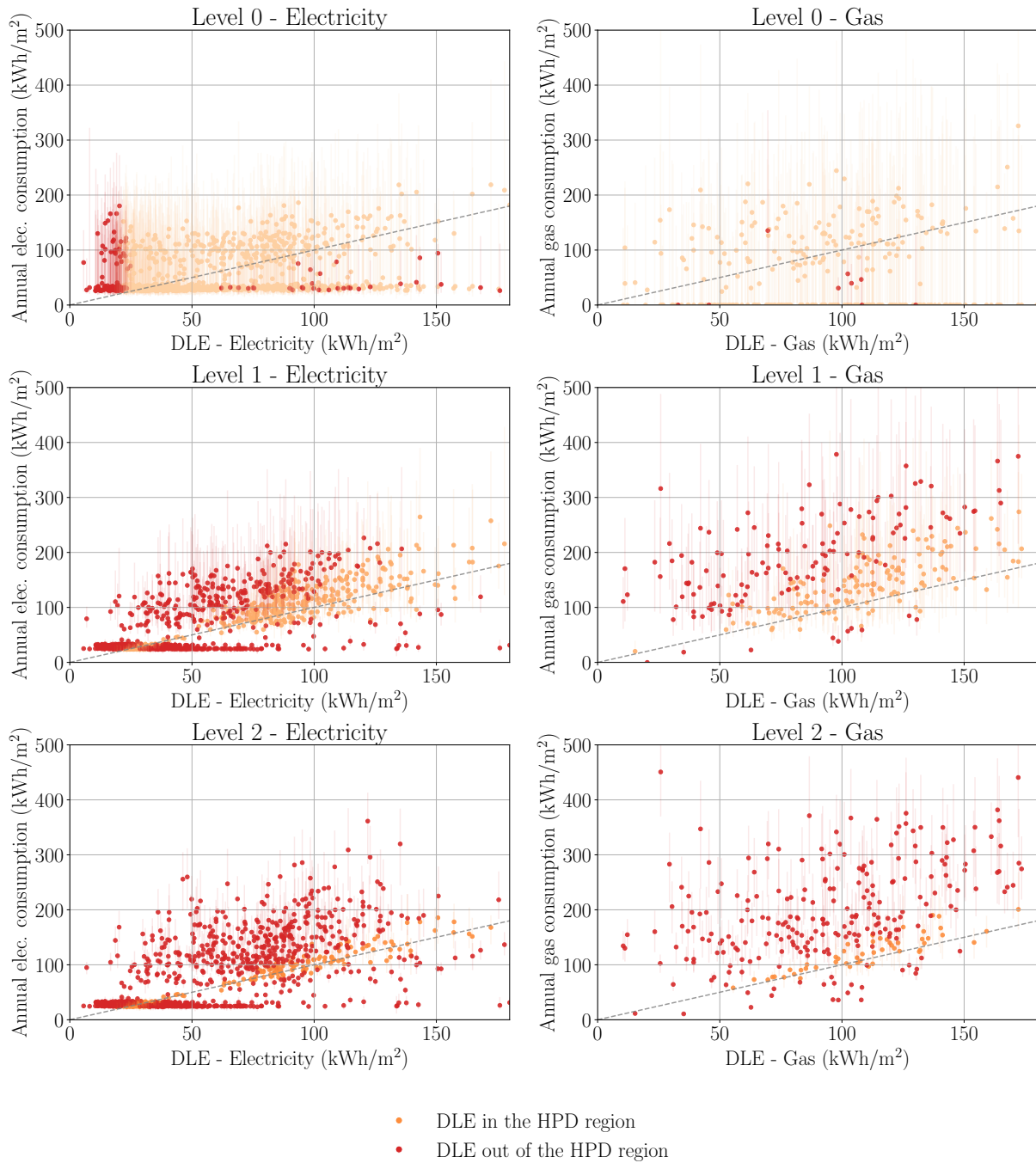


Figure 5.29: μ_{bldg} and the HPD region of each building obtained with Calomel against its DLE for three levels of data availability

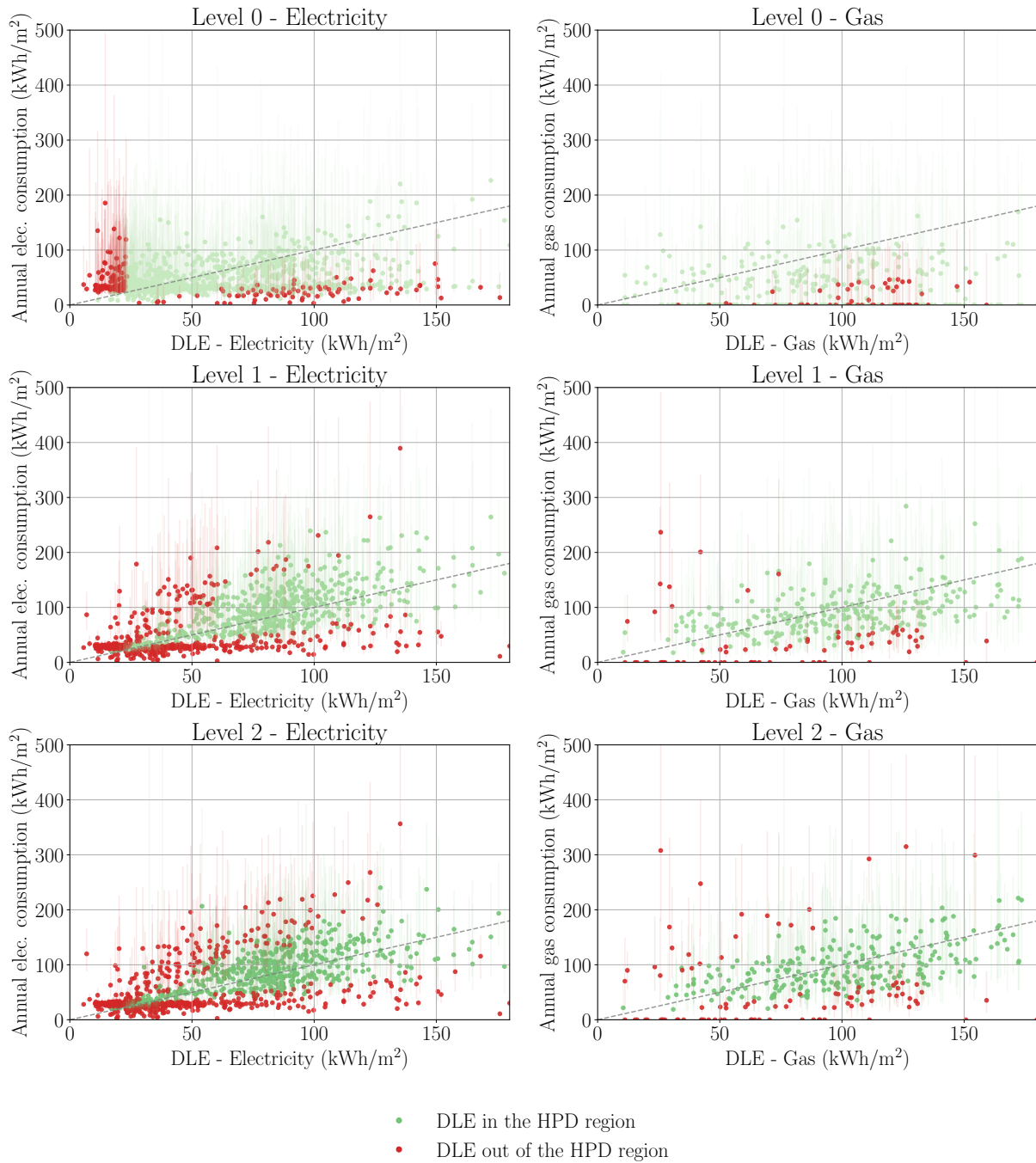


Figure 5.30: μ_{bldg} and the HPD region of each building obtained with Dimosim against its DLE for three levels of data availability



Figure 5.31: Mutual information between the residual obtained with Calomel and buildings characteristics for three levels of data availability

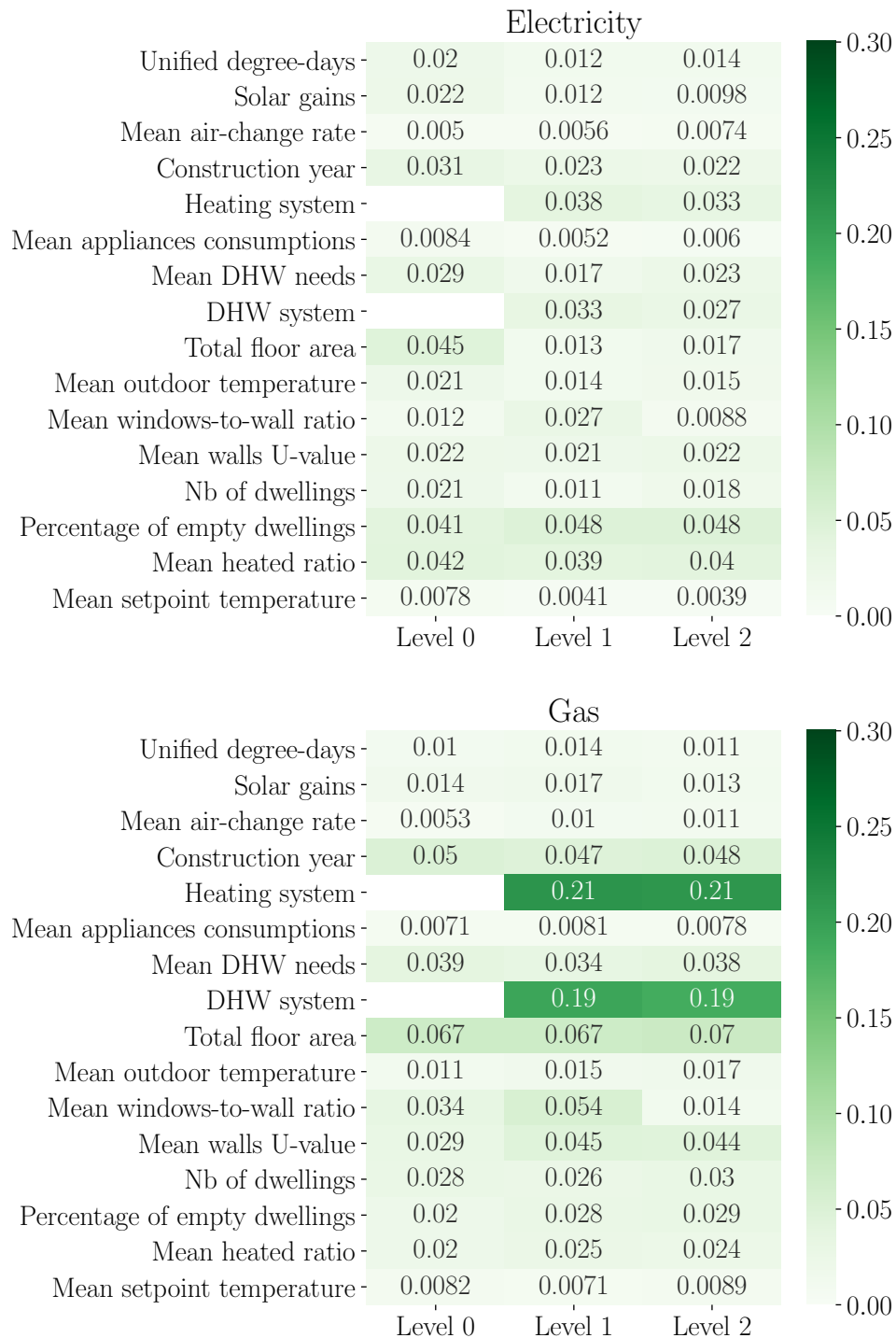


Figure 5.32: Mutual information between the residual obtained with Dimosim and buildings characteristics for three levels of data availability

It can be noticed for both tools that the heating and the DHW systems seem to influence greatly the percentage of error for the level 0 on both gas and electricity. However, focussing on electricity, it appears that the appliances consumptions also stands out for Calomel. A same trend is visible for Dimosim although with a lower intensity.

Looking at the vectors of mutual information obtained for the level 2, three parameters are noticeable as having an influence on the error on gas: the heated ratio, the air-change rate and the DHW needs. The two first were also noticeable for Simply. On the contrary, none of the three remaining uncertain parameters used in Calomel seem to impact the error. This would indicate that the bias highlighted on this tool originates from parameters that were kept outside this uncertainty analysis.

To push further the analysis, a PCA was run on the vectors of mutual information obtained for Calomel and Dimosim. The results are presented in Figures 5.37 and 5.38 and in Figures 5.39 and 5.40.

The same trend as the one observed on the scatter plots for Simply is also apparent here for Calomel and Dimosim: the points seem randomly spread in the plane at level 0 while two groups of buildings are clearly discernible at level 1 and level 2. The distribution of the points in the plane is also very similar between level 1 and level 2 for all tools.

Apart from that, the results obtained with Calomel for level 2 are not of much help, the PCA being even unnecessary in this case since the dimension of the vectors of mutual information is already small in this case.

In spite of that, the results obtained with Dimosim are similar to what was observed for Simply: once parameters with an influence on the error at the inferior level of data availability are set at a fixed value thanks to observations available on each building, other parameters influencing the error can be revealed. In the case of the error on gas consumptions with Dimosim, the heated ratio and the air-change rate are then highlighted in the second eigenvector for obtained for the level 2 of data availability once the thermal properties have been set to a fixed value.

The analysis conducted on Calomel and Dimosim enabled then to identify trends similar to the one that were noticed for Simply: a bias reveals itself when the level of data availability increases and the level of uncertainty decreases. This bias is explained by parameters for which data is unavailable such as the DHW needs and the appliances consumptions. It also shows that having thermal properties of buildings does not suffice to adequately simulate the annual electricity and gas consumptions at the building scale.

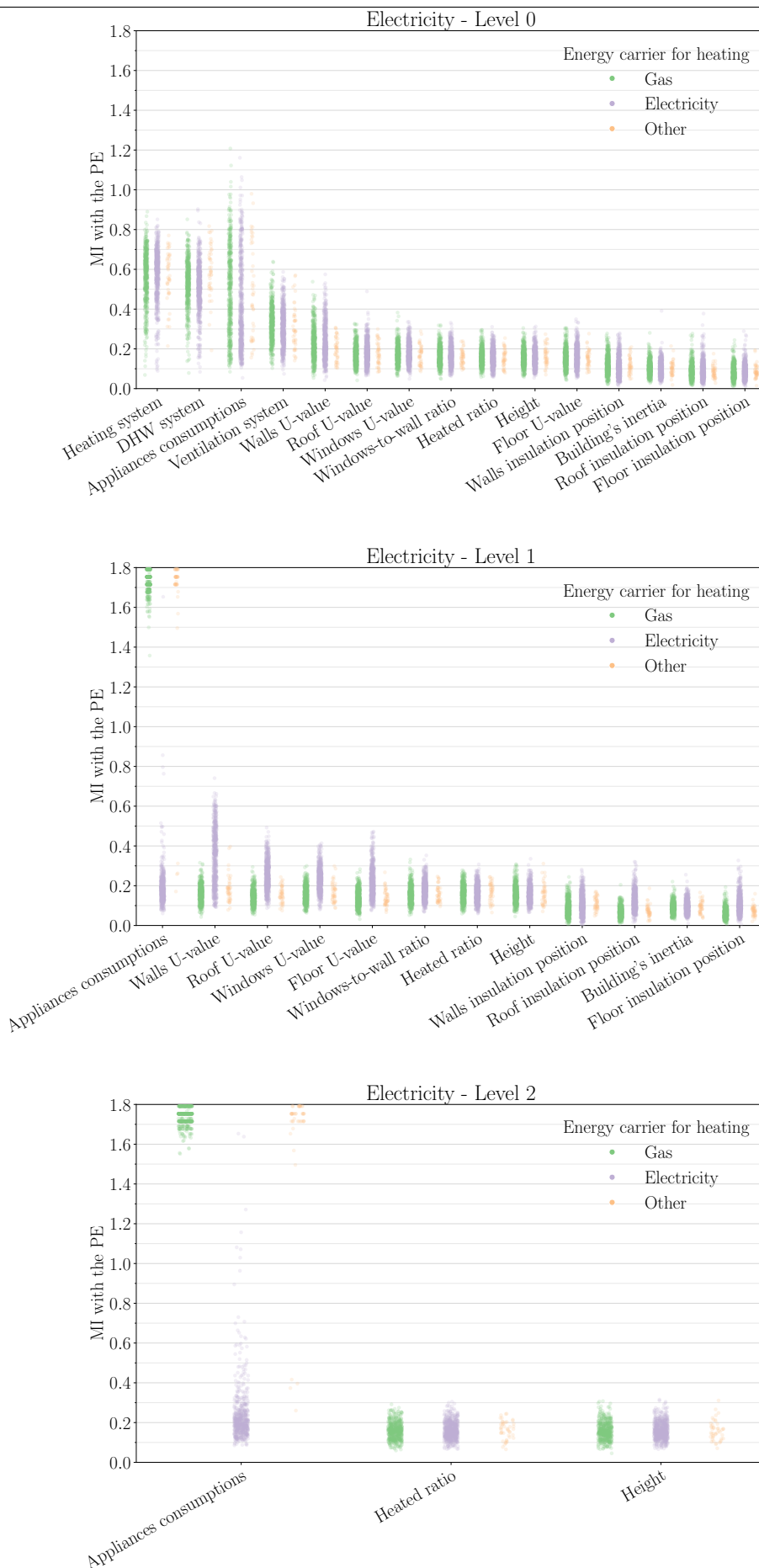


Figure 5.33: Vectors of mutual information between uncertain parameters and percentage of error on electricity consumptions obtained for the buildings of the validation dataset with Calomel and for three levels of data availability

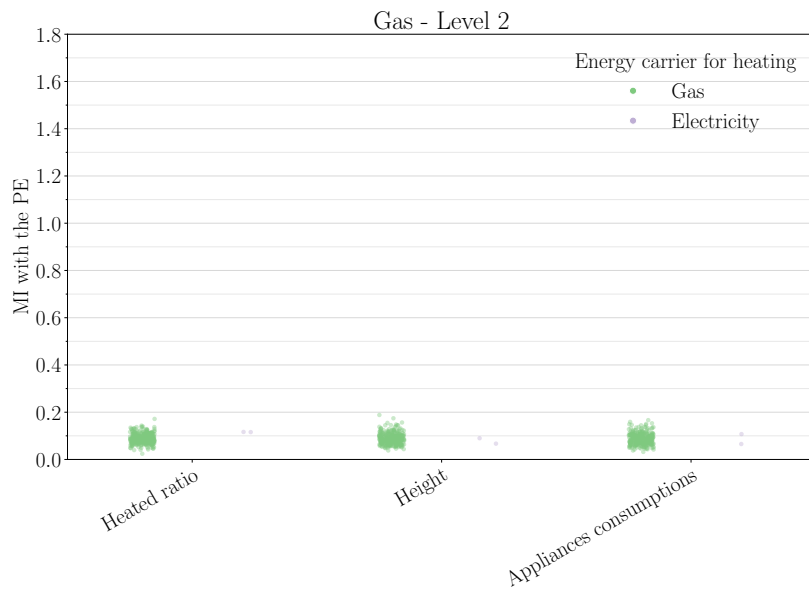
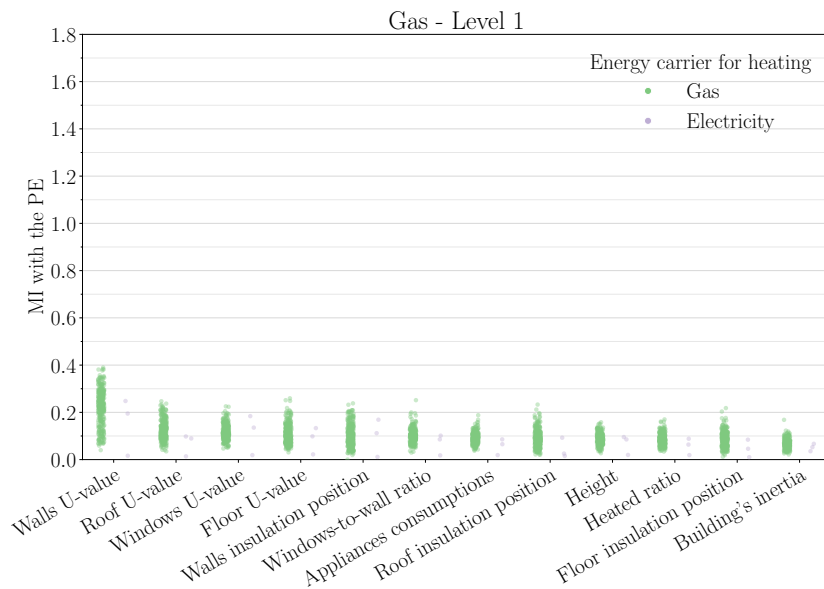
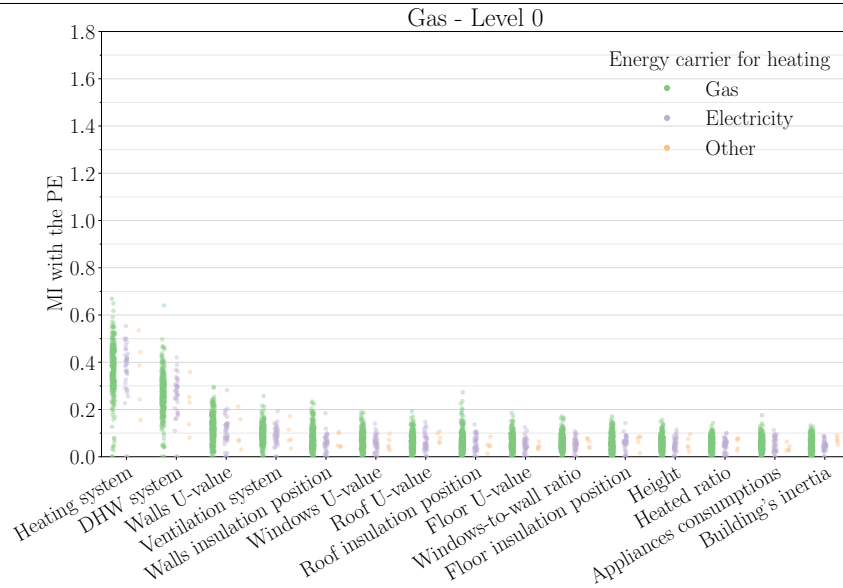


Figure 5.34: Vectors of mutual information between uncertain parameters and percentage of error on gas consumptions obtained for the buildings of the validation dataset with Calomel and for three levels of data availability

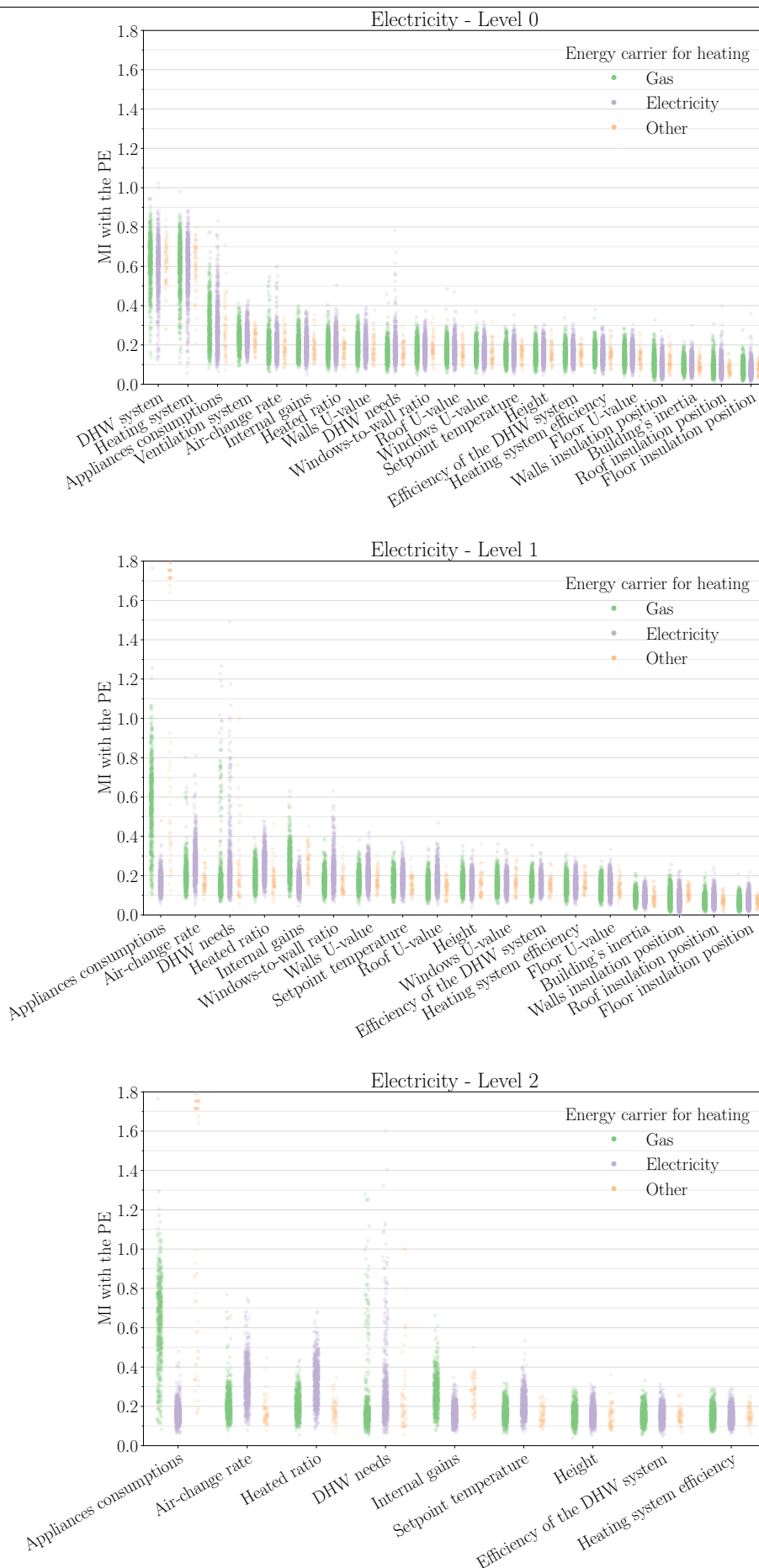


Figure 5.35: Vectors of mutual information between uncertain parameters and percentage of error on electricity consumptions obtained for the buildings of the validation dataset with Dimosim and for three levels of data availability

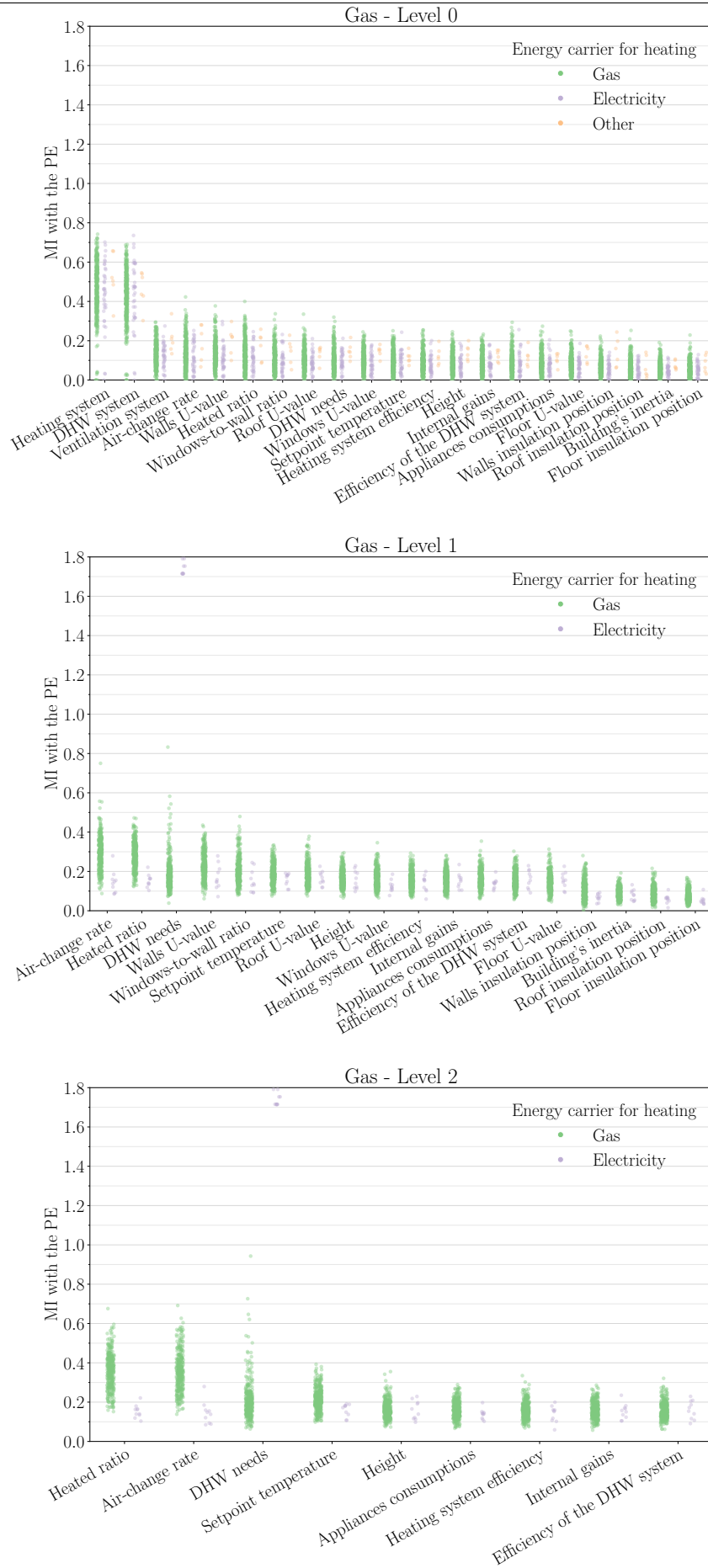


Figure 5.36: Vectors of mutual information between uncertain parameters and percentage of error on gas consumptions obtained for the buildings of the validation dataset with Dimosim and for three levels of data availability

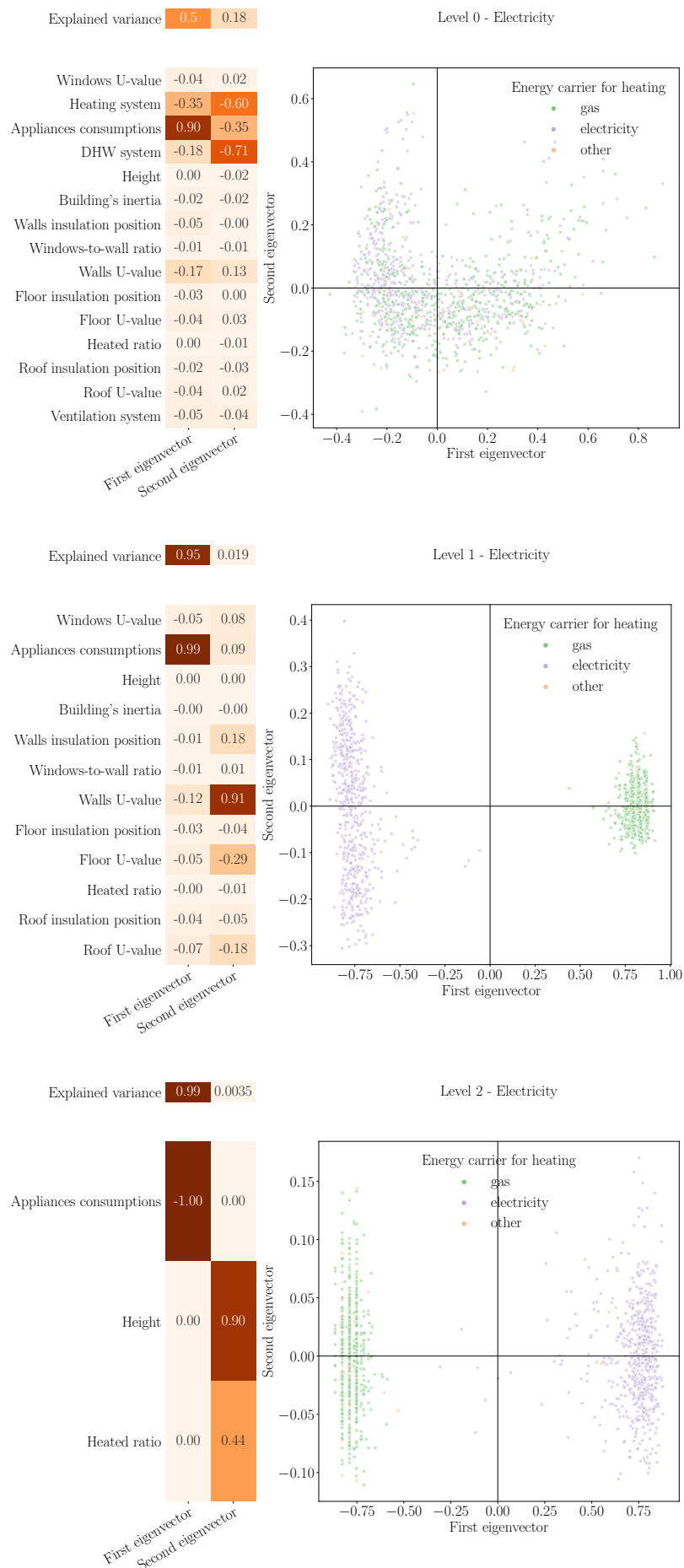


Figure 5.37: Results of principal components analysis applied to vectors of mutual information obtained for error on electricity consumptions with Calomel for three levels of data availability

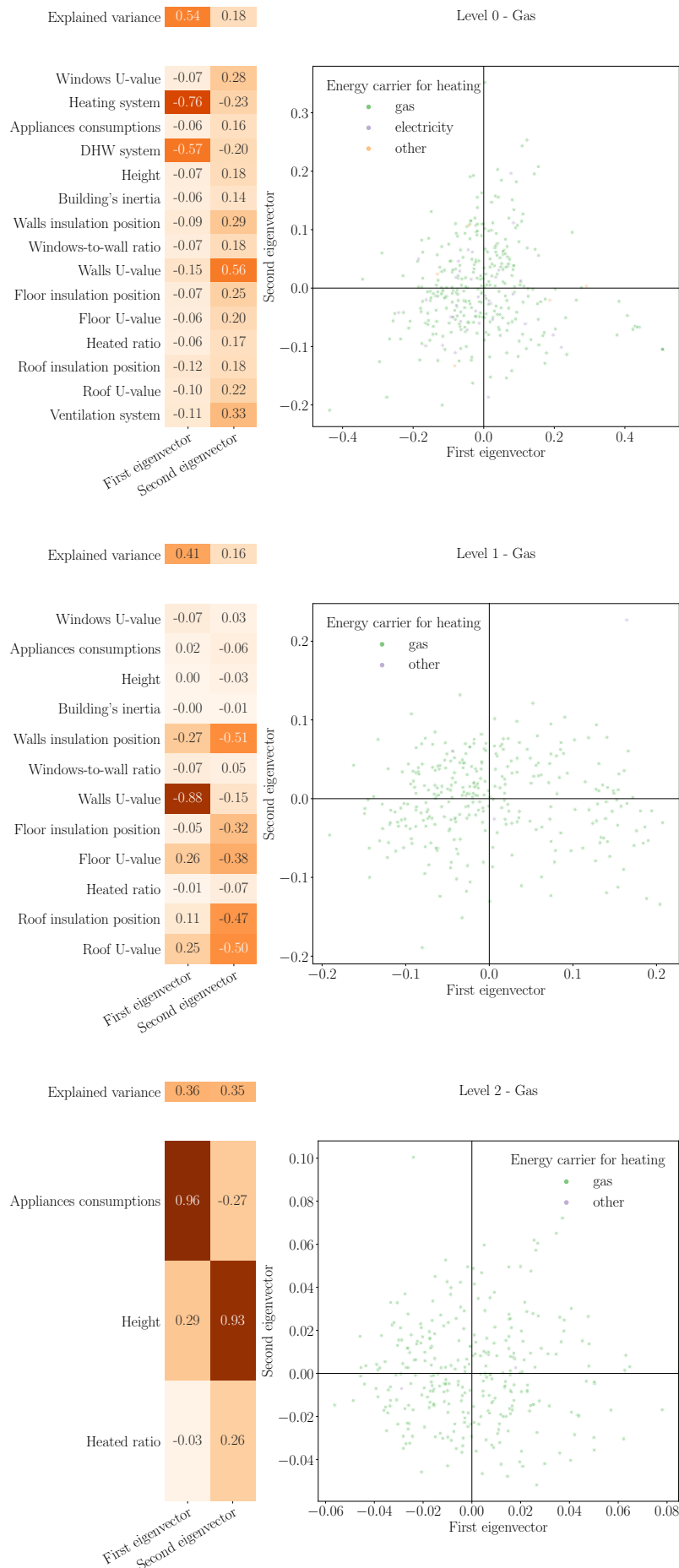


Figure 5.38: Results of principal components analysis applied to vectors of mutual information obtained for error on gas consumptions with Calomel for three levels of data availability

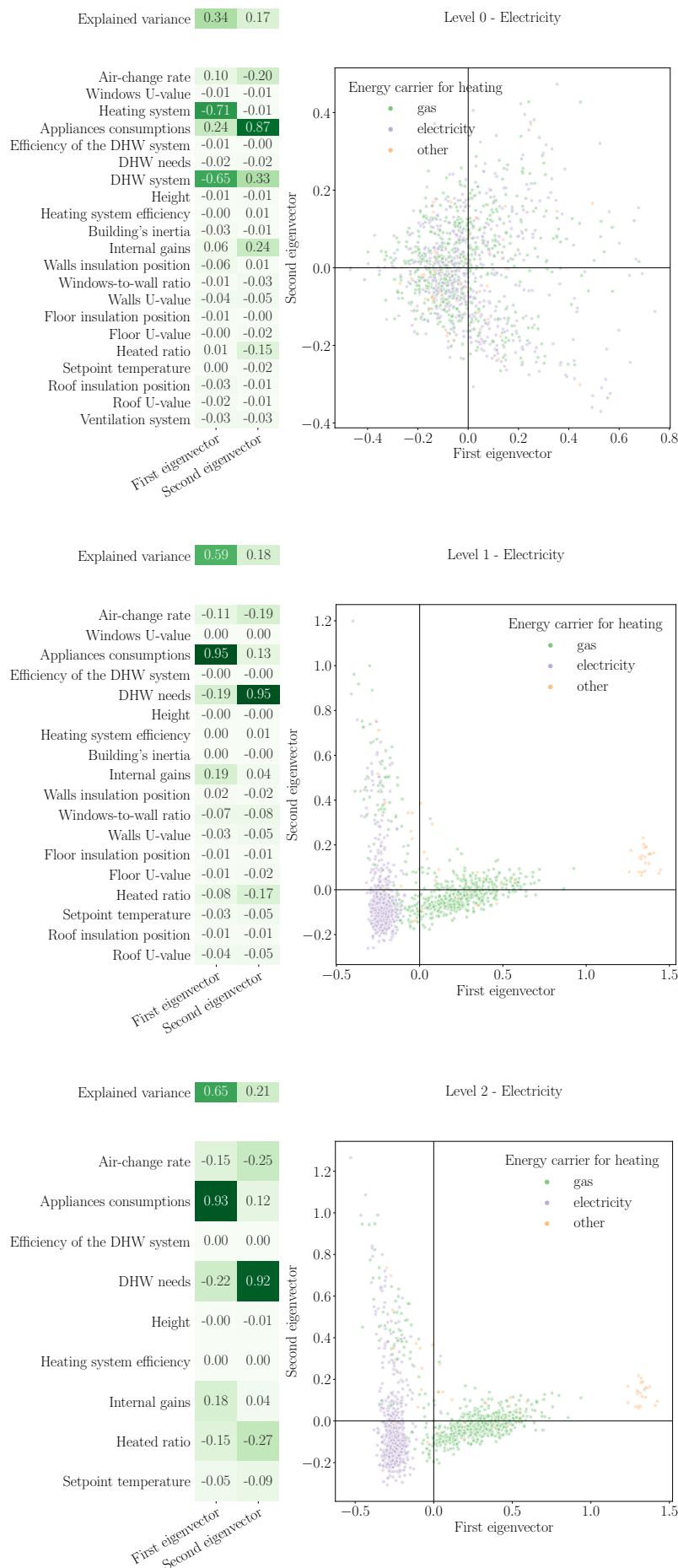


Figure 5.39: Results of principal components analysis applied to vectors of mutual information obtained for error on electricity consumptions with Dimosim for three levels of data availability

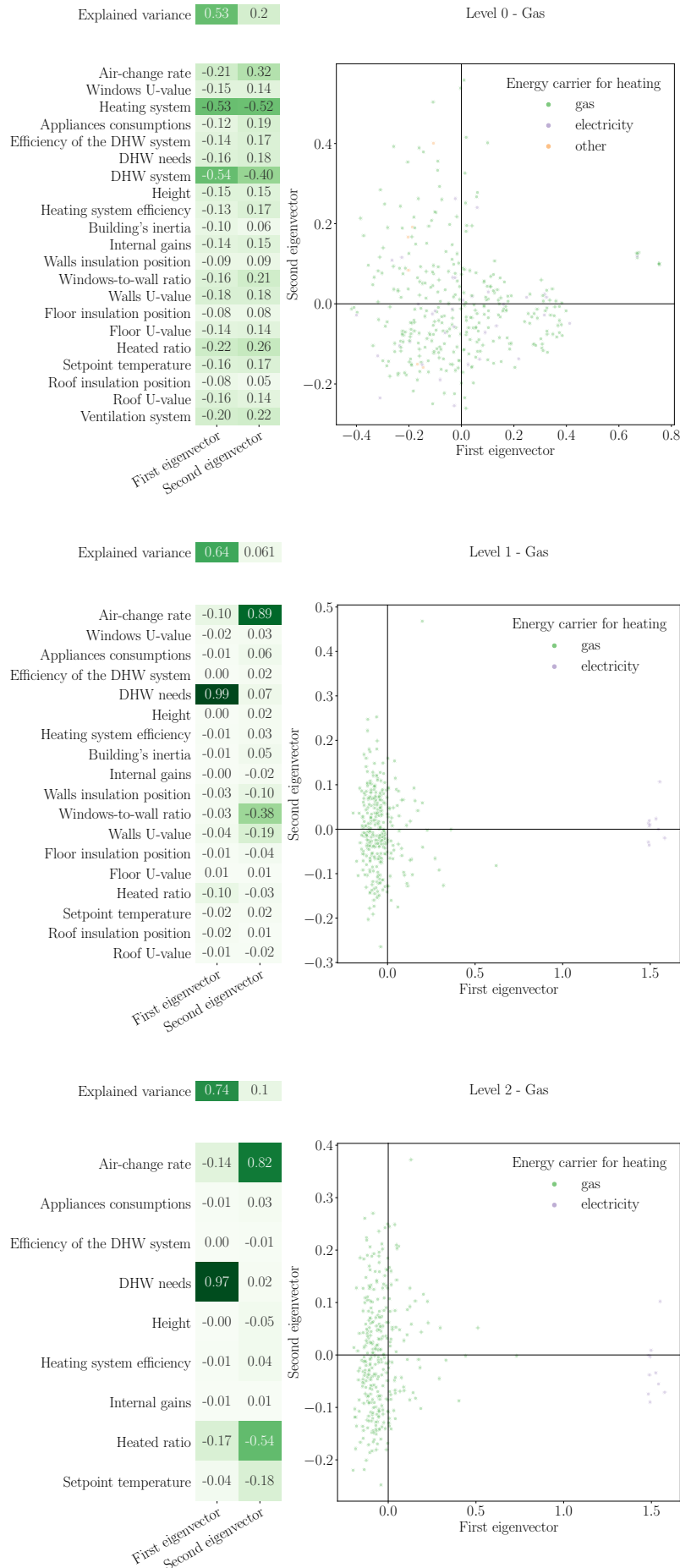


Figure 5.40: Results of principal components analysis applied to vectors of mutual information obtained for error on gas consumptions with Dimosim for three levels of data availability

Conclusion

A diagnostic procedure was introduced in this chapter to help understand the error displayed by UBEM tools at the assessment step of this validation framework.

The aim of this procedure was to distinguish between three main possible causes for the error:

- the data;
- the UBEM;
- the way the validation framework was applied.

To help understand which one of these three causes could be the most influential one, three aspects were studied:

1. The parameters with an impact on the error at the sample scale were retrieved;
2. The parameters explaining the variation of the error for each building were identified;
3. Groups of buildings with a similar behaviour towards the error were defined.

It was then identified that buildings with the same energy carrier for heating and for DHW have a similar behaviour towards the error. Parameters driving the error for each group of buildings were identified thanks to a PCA.

The same procedure was applied to three tools that all displayed a similar behaviour. This enabled to discard the physical model as one of the cause of error.

In order to identify between the three remaining sources of error (the data quality, the enrichment process and the uncertainty quantification) the most influent one, another survey was led with a varying level of data availability. Although this survey did not permit to fully assess the influence of data quality, it enabled to study the impact of the model parameterisation. Indeed, with more data available, the level of uncertainty that was accounted for decreases (once the value for a parameter is known, no uncertainty is accounted for in this case) along with the influence of the enrichment process (since the value of the parameter is known, the enrichment process is not required for this input anymore). Thus, if the error decreases when the level of data availability increases, it implies that the enrichment process is biased and generates a large part of the error. On the contrary, if the error is not impacted by the addition of information, it implies that the physical model used is not sensitive to the new information.

In this case, it appeared that the uncertainty quantification was completely driving the tool behaviour when almost no information was retrieved from the dataset. Then, when more information becomes available, a clear although biased relationship between the simulated and measured consumptions appears. This bias seems to be mostly driven by parameters for which no information was available: the DHW needs and the appliances consumptions. In this case, it implies that the uncertainty quantification on these parameters was not well adjusted. Having more information on these parameters should then help improve significantly the results obtained by the three tools.

Chapter 6

Conclusion and outlook

Conclusion and main contributions

As it was described at the beginning of this manuscript, Urban Building Energy Models are decision-support tools that are designed to assist cities in their energy transition. In order to ensure that these tools provide a real aid to the decision process, their reliability must be challenged. A validation framework dedicated to UBEM was then developed in this thesis to verify their fitness for purpose.

UBEM are defined as decision-support tool able to model energy consumptions at the building scale for a hundred to several thousands of buildings (**Chapter 1**). In most cases, UBEM suffer from data scarcity which makes impossible to know all the inputs required by the physical model for the simulation of the buildings' energy behaviour. To overcome this issue, several strategies are implemented and are identified as the enrichment process. At this stage, many hypotheses are made which risk to jeopardize the final results provided by the UBEM.

Regarding the physical models used in UBEM, several levels of model complexity can be found from one tool to another. The two most common approaches are the use of an existing Building Energy Model or the use of a Resistance-Capacitance model.

The tools found in the UBEM field are then very diverse and so are the decision-making processes they are involved in. A review of the most common use cases was provided in this chapter. The UBEM outputs used in the decision-making process for each use case were identified along with the weight they have on the final decision and the preciseness which is then expected from these outputs.

Since many UBEM use a BEM as their physical model, it was interesting to study the validation frameworks existing for BEM (**Chapter 2**). These tools have benefitted from several research projects on the topic such as the BESTEST or the PASSYS project. Although validation approaches such as the BESTEST are very interesting to identify parts of the models that could be faulty, they do not enable to test the tools under their real conditions of use.

Regarding the validation of UBEM, a project was launched to create a DESTEST with the same approach as the BESTEST. A synthetic district was then created and several UBEM were compared on this test case. Such approach is interesting to identify weaknesses in numerical implementations and to compare intrinsic performances of the physical models. However, it does not permit to test the whole UBEM since it leaves aside the enrichment process, all parameters being perfectly known. Examples of empirical validation applied to UBEM can also be found in the scientific literature, but they are rarely replicable, and there is a lack of a common framework enabling the comparison of results from one paper to another.

In order to overcome the identified limitations in UBEM validation, a validation framework was then proposed. The proposal is to validate UBEM per use case and to assess the output on real data while taking into account the uncertainty surrounding the simulation. As a reminder, this validation framework is described in Figure 6.1.

The method used for the ad hoc validation of the enrichment process and the physical model was then discussed in **Chapter 3**. The ad hoc validation of the enrichment process uses common practise from the machine learning field while the ad hoc validation of the physical model is based on the BESTEST and DESTEST.

A first set of use cases were then chosen for a first application of the validation framework. This choice was made based on the databases that were publicly available and that provided energy measurements for a large number of buildings. The UBEM outputs selected were then the annual electricity consumption and the annual gas consumption at the building scale.

The creation of a validation dataset to assess UBEM on the chosen use cases was then described.

Next, the evaluation stage of the validation framework was detailed in **Chapter 4**. It starts with the presentation of the uncertainty quantification. This step is of the utmost importance since it influences all the results of the validation process, and it must be adapted to every new validation dataset. Then, the whole evaluation procedure was introduced on a very simple tool, *Simply*. Indicators were proposed to assess the tool performances such as the residual and the empirical coverage. The evaluation stage was then applied to Calomel and Dimosim, and the results obtained by the three tools were compared.

In order to gain a better understanding of the error, a diagnostic stage was proposed in **Chapter 5**. The aim of this procedure was to distinguish between three main possible causes for the error: the data; the UBEM and the way the validation framework was applied.

To help understand which one of these three causes could be the most influential, three aspects were studied:

1. The parameters with an impact on the error at the sample scale were retrieved;
2. The parameters explaining the variation of the error for each building were identified;
3. Groups of buildings with a similar behaviour towards the error were defined.

It was then identified that buildings with the same energy carrier for heating and for DHW have a similar behaviour towards the error. Parameters driving the error for each group of buildings were identified thanks to a PCA.

The same procedure was applied to three tools that all displayed a similar behaviour. This enabled to discard the physical model as one of the causes of error.

In order to identify between the three remaining sources of error (the data quality, the enrichment process and the uncertainty quantification) the most influent one, another survey was led with a varying level of data availability. Although this survey did not permit to fully assess the influence of data quality, it enabled to study the impact of the model parameterisation.

It appeared that the uncertainty quantification was completely driving the tool behaviour when almost no information was retrieved from the dataset. Then, when more information becomes available, a clear although biased relationship between the simulated and measured consumptions appears. This bias seems to be mostly driven by parameters for which no information was available: the DHW needs and the appliances consumptions. In this case, it implies that the uncertainty quantification on these parameters was not well adjusted. Having more information on these parameters should then help improve significantly the results obtained by the three tools.

Outlook

As it was highlighted above, a validation framework dedicated to UBEM was proposed in this thesis and applied to a first set of use cases and to three different tools. However, several aspects of the work presented here would require more research.

First, as explained above, the uncertainty quantification influences greatly the results of the validation procedure, and, as such, this step should be challenged by further research works. The hypotheses selected to model uncertainty on each input variable could then be examined and their impact on the final results could be studied. For instance, no uncertainty was considered on the parameters related to the weather but the uncertainty on parameters such as the outdoor temperature is probably high, namely in urban area due to the urban heat island effect. Modelling

uncertainty on the outdoor temperature would then help assess the tools under more realistic conditions and would enable to assess the influence of this given parameter on the error.

Second, only a specific enrichment process, Qiriél, was studied in this manuscript. It would be interesting to test the same physical models with different approaches for the enrichment process. It would namely be interesting to study the tools performances when the enrichment process is based on deterministic archetypes.

Third, in order to understand how robust are the results obtained here by the three tools, it would be interesting to run the validation framework for the same use cases and for the same tools but with a completely different validation dataset. Similar databases as the one used to create the validation dataset used in this manuscript are available in most European countries and in the USA. It would be possible then to create similar validation datasets and to study the evolution of the tools performances from one validation dataset to another.

Lastly but more importantly, the validation framework applied here to assess tools performances on annual energy consumptions at the building scale should be adapted to the assessment of dynamic outputs such as electric load profiles or thermal load profiles. The general method as introduced Chapter 2 should be applicable as is, but the indicators provided here are well-fitted to the evaluation of static outputs only, and other indicators should be proposed in the case of dynamic outputs.

To summarize, this thesis started the development of a validation framework dedicated to UBEM. Many of the tools weaknesses were highlighted along with the influence of uncertainty on the value displayed by a UBEM output. Every modeller should always keep in mind that even though a tool has been tested through a validation framework, it should **never** be considered as entirely validated, especially in the case of UBEM where so much uncertainty can interfere. The results provided by a UBEM should then always be taken with caution and an indication on the uncertainty level should always be provided with the results themselves. Ultimately, understanding and addressing the weaknesses of UBEM would help these tools to provide more adequate and reliable support to decision-makers in the decision-making process for which they are used.

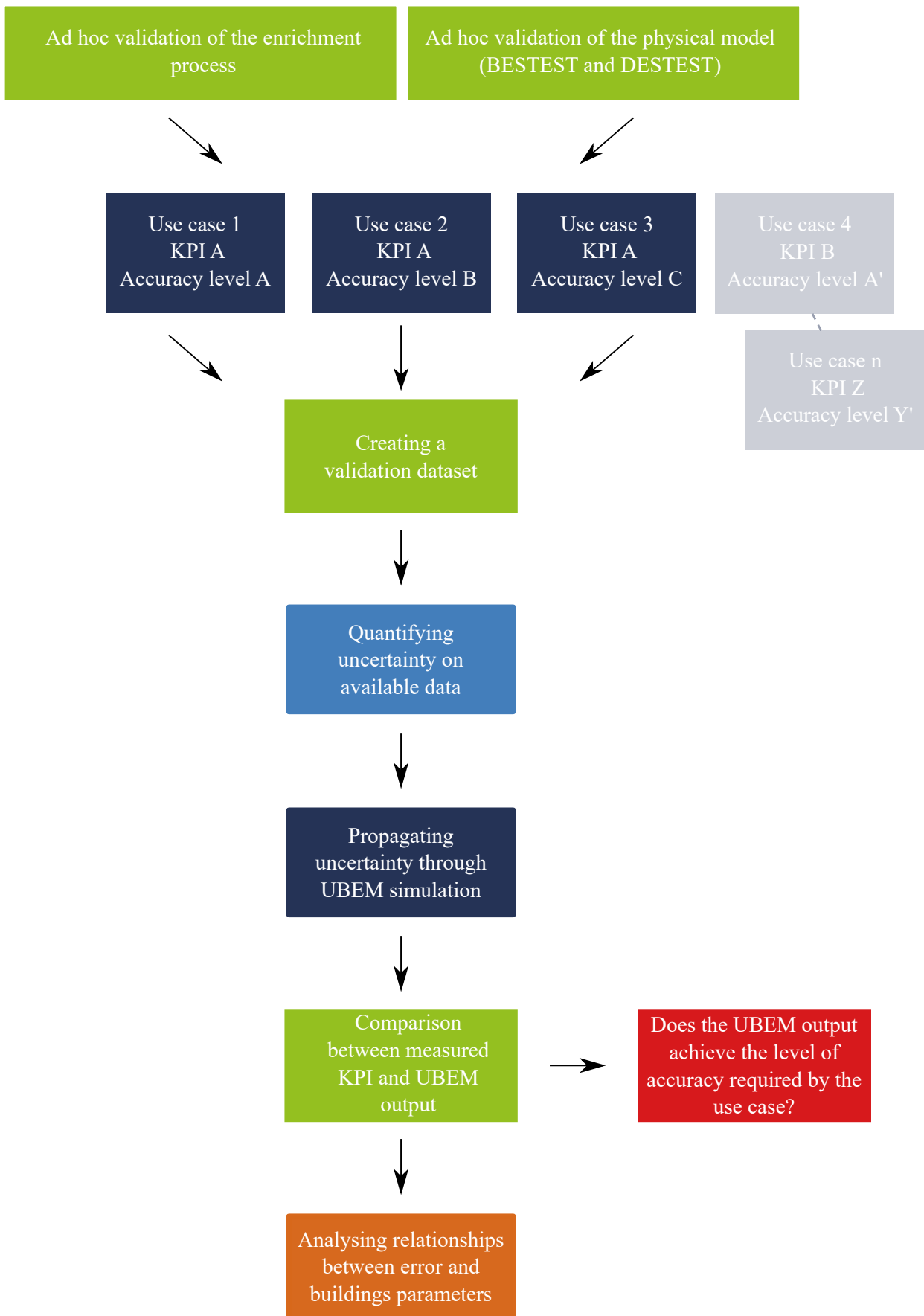


Figure 6.1: Validation framework for UBEM: a proposal

Bibliography

- [1] UN, “Revision of world urbanization prospects,” 2018. v, vii, 2
- [2] M. of London, “Zero carbon london: A 1.5oc compatible plan,” *Greater London Authority*, 2018. v, vii, 2
- [3] NYC, “Onenyc 2050: Building a strong and fair city,” vol. 1, 2019. v, vii, 2
- [4] ADEME, “Chiffres-clés climat, air et Énergie,” 2015. v, vii, 2
- [5] C. F. Reinhart and C. C. Davila, “Urban building energy modeling - a review of a nascent field,” *Building and Environment*, 2016. Version papier aux Mines. v, vii, 2
- [6] S. Østergaard Jensen, “Validation of building energy simulation programs: a methodology,” *Energy and Buildings*, vol. 22, pp. 133–144, 1995. Version papier aux Mines. v, vii, xi, 2, 18, 19, 162
- [7] R. Judkoff and J. Neymark, “Twenty years on!: Updating the IEA bestest building thermal fabric test cases for ASHRAE standard 140,” *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association*, pp. 63–70, 2013. v, vii, 2, 18, 19, 37, 162
- [8] D. Saelens, I. D. Jaeger, F. Büning, M. Mans, A. Vandermeulen, B. V. D. Heijde, E. Garreau, A. Maccarini, Øystein Rønneseth, I. Sartori, and L. Helsen, “Towards a destest: a district energy simulation test developed in IBPSA Project 1,” 2019. Version papier aux Mines. v, vii, xi, 22, 27, 30, 37, 40, 162
- [9] J. Sokol, C. C. Davila, and C. F. Reinhart, “Validation of a Bayesian-based method for defining residential archetypes in urban building energy models,” *Energy and Buildings*, 2017. v, vii, 9, 23
- [10] J. Widén, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegård, and E. Wäckelgård, “Constructing load profiles for household electricity and hot water from time-use data-modelling approach and validation,” *Energy and Buildings*, vol. 41, pp. 753–768, 2009. Presentation of a model for electricity and hot water load profiles calibrated with time-use data and validated against measured electricity consumption. v, vii, 24
- [11] E. Garreau, Y. Abdelouadoud, E. Herrera, W. Keilholz, G. E. Kyriakodis, V. Partenay, and P. Riederer, “District modeller and simulator (DIMOSIM) – a dynamic simulation platform based on a bottom-up approach for district and territory energetic assessment,” *Energy and Buildings*, vol. 251, p. 111354, 2021. xi, 5, 36, 37, 39, 40
- [12] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2017. xi, 61, 62
- [13] K. P. Murphy, *Machine learning - A probabilistic perspective*. 2012. xii, 48, 62, 76, 77
- [14] U. IPCC, “Climate change 2022 - impacts, adaptation and vulnerability - summary for policymakers,” 2022. 2
- [15] UN-Habitat, “Energy.” 2, 159
- [16] V. de Paris, “Plan climat de Paris : vers une ville neutre en carbone et 1002
- [17] U. D. of Energy, “Energyplus.” 2, 5

- [18] C. E. Kontokosta and C. Tull, “A data-driven predictive model of city-scale energy use in buildings,” *Applied Energy*, vol. 197, pp. 303–317, 2017. 2, 5, 31
- [19] G. Happle, J. A. Fonseca, and A. Schlueter, “A review on occupant behavior in urban building energy models,” *Energy and Buildings*, vol. 174, pp. 276–292, 9 2018. Version papier aux Mines. 4, 10, 12
- [20] E. Vorger, “Etude de l’influence du comportement des occupants sur la performance énergétique des bâtiments,” 2014. 4
- [21] D. Robinson, F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, and U. Wilke, “Citysim: Comprehensive micro-simulation of resource flows for sustainable urban planning,” 2009. Presentation of CitySim: RC model for thermal simulation, multi-zone Radiation model from SunTool Views to link it with a transportation model. 4, 5, 10, 11, 12, 13
- [22] J. Allegrini, K. Orehoung, G. Mavromatidis, F. Ruesch, V. Dorer, and R. Evins, “A review of modelling approaches and tools for the simulation of district-scale energy systems,” *Renewable and Sustainable Energy Reviews*, 2015. A review from 2015 on the different models and tools existing for the simulation of: district energy systems renewable energies urban micro-climates. 4
- [23] J. A. Fonseca, T. A. Nguyen, A. Schlueter, and F. Marechal, “City energy analyst (cea): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts,” *Energy and Buildings*, 2016. Presentation of CEA : modelling of heating/cooling sources optimisation regarding costs and CO2 emissions. 4, 5, 10, 11
- [24] L. Frayssinet, L. Merlier, F. Kuznik, J. L. Hubert, M. Milliez, and J. J. Roux, “Modeling the heating and cooling energy demand of urban buildings at city scale,” *Renewable and Sustainable Energy Reviews*, 2018. 4
- [25] C. F. Reinhart, T. Dogan, A. Jakubiec, T. Rakha, and A. Sang, “Umi-an urban simulation environment for building energy use, daylighting and walkability 2 3,” 2013. Presentation of UMI : Energy simulation of buildings based on EnergyPlus and integrated as a Rhino plug-in Daylight simulation Walkability modeling. 5
- [26] T. Hong, Y. Chen, S. H. Lee, and M. A. Piette, “Citybes: A web-based platform to support city-scale building energy efficiency,” 2016. Version papier aux Mines. 5, 12, 13
- [27] B. Polly, C. Kutscher, D. Macumber, and M. Schott, “From zero energy buildings to zero energy districts,” pp. 1–16, 2016. 5
- [28] T. Hong, Y. Chen, X. Luo, N. Luo, and S. H. Lee, “Ten questions on urban building energy modeling,” *Building and Environment*, vol. 168, 1 2020. Version papier aux Mines. 5, 10, 11, 12
- [29] T. Berthou, B. Duplessis, P. Rivière, P. Stabat, D. Casetta, and D. Marchio, “Smart-e: A tool for energy demand simulation and optimization at the city scale,” pp. 1782–1789, 2015. Présentation de Smart-E Présentation des bases de données utilisées pour générer les occupants et les consommations. 5, 51
- [30] C. C. Davila, C. F. Reinhart, and J. L. Bemis, “Modeling boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets,” *Energy*, vol. 117, pp. 237–250, 12 2016. 7, 10, 12

- [31] J. A. Fonseca and A. Schlueter, “Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts,” *Applied Energy*, 2015. Presentation of a UBEM, examples of application and presentation of validation process (comparison with collected measures and the output of a peer model).
-> 1 to 197, 10, 11, 12, 13, 24
- [32] Y. Chen, T. Hong, X. Luo, and B. Hooper, “Development of city buildings dataset for urban building energy modeling,” *Energy and Buildings*, vol. 183, pp. 252–265, 1 2019. Version papier aux Mines.
- [33] P. Remmen, M. Lauster, M. Mans, M. Fuchs, T. Osterhage, and D. Müller, “Teaser: an open tool for urban energy modelling of building stocks,” *Journal of Building Performance Simulation*, vol. 11, pp. 84–98, 1 2018. Version papier aux Mines. 7, 10, 12
- [34] I. D. Jaeger, J. Lago, and D. Saelens, “A probabilistic approach to allocate building parameters within district energy simulations,” 2018. 7
- [35] ADEME, “Diagnostics de performance énergétique pour les logements.” 6, 43
- [36] S. Goy, V. Coors, and D. Finn, “Grouping techniques for building stock analysis: A comparative case study,” *Energy and Buildings*, vol. 236, p. 110754, 2021. 8
- [37] J. F. Rosser, G. Long, S. Zakhary, D. S. Boyd, Y. Mao, and D. Robinson, “Modelling urban housing stocks for building energy simulation using citygml energyade,” *ISPRS International Journal of Geo-Information*, vol. 8, 3 2019. 8, 27
- [38] C. Cerezo, J. Sokol, S. AlKhaled, C. Reinhart, A. Al-Mumin, and A. Hajiah, “Comparison of four building archetype characterization methods in urban building energy modeling (ubem): A residential case study in kuwait city,” *Energy and Buildings*, vol. 154, pp. 321–334, 11 2017. 8, 9, 62
- [39] C. kai Wang, S. Tindemans, C. Miller, G. Agugiaro, C. kai Wang, S. Tindemans, C. Miller, and G. Agugiaro, “Bayesian calibration at the urban scale : a case study on a large residential heating demand application in amsterdam demand application in amsterdam,” *Journal of Building Performance Simulation*, vol. 0, pp. 1–15, 2020. Version papier aux Mines. 8, 22, 31
- [40] O. Pasichnyi, J. Wallin, and O. Kordas, “Data-driven building archetypes for urban building energy modelling,” *Energy*, vol. 181, pp. 360–377, 2019. Version papier à Efficacity. 8
- [41] E. Garrison and J. New, “Quality control methods for advanced metering infrastructure data,” *Smart Cities*, pp. 195–203, 2021. 8, 23, 26, 28
- [42] I. D. Jaeger, G. Reynders, C. Callebaut, and D. Saelens, “A building clustering approach for urban energy simulations,” *Energy & Buildings*, vol. 208, p. 109671, 12 2020. Version papier aux Mines. 8
- [43] G. Ledesma, O. Pons-Valladares, and J. Nikolic, “Real reference buildings for urban energy modelling a multistage validation and diversification approach,” *Building and Environment*, p. 135907, 2021. 8
- [44] S. Nagpal and C. F. Reinhart, “A comparison of two modeling approaches for establishing and implementing energy use reduction targets for a university campus,” *Energy and Buildings*, vol. 173, pp. 103–116, 8 2018. Version papier aux Mines. 9, 10, 12, 13, 22
- [45] I. D. Jaeger, “On the impact of input data uncertainty on the reliability of urban building energy models,” 2021. 9, 27, 29, 32

- [46] P. Schetelat, L. Lefort, and N. Delgado, “Urban data imputation using multi-output multi-class classification,” 2020. 9, 32, 44, 165
- [47] M. Gholami, D. Torreggiani, P. Tassinari, and A. Barbaresi, “Narrowing uncertainties in forecasting urban building energy demand through an optimal archotyping method,” *Renewable and Sustainable Energy Reviews*, vol. 148, p. 111312, 2021. 9
- [48] J. Chambers, K. Narula, M. Sulzer, and M. K. Patel, “Mapping district heating potential under evolving thermal demand scenarios and technologies: A case study for switzerland,” *Energy*, vol. 176, pp. 682–692, 2019. 10
- [49] B. Despres, “Etude de faisabilité et de dimensionnement d’un réseau de chaleur avec chaufferie biomasse,” 2012. 10
- [50] V. M. Jourjon, T. Berthou, and P. Stabat, “Rapport 2a : Modélisation adaptée à l ’évaluation des performances énergétiques des quartiers,” pp. 1–82, 2019. 11
- [51] RTE, “Évaluation des économies d’énergie et des effets de bord associés aux effacements de consommation,” 2016. 12
- [52] ADEME, *Trajectoires d ’ évolution du mix électrique 2020-2060*. 2018. 12
- [53] Y. Chen, T. Hong, and M. A. Piette, “Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis,” *Applied Energy*, vol. 205, pp. 323–335, 2017. 12, 13
- [54] G. Sousa, B. M. Jones, P. A. Mirzaei, and D. Robinson, “An open-source simulation platform to support the formulation of housing stock decarbonisation strategies,” *Energy and Buildings*, vol. 172, pp. 459–477, 8 2018. Version papier à Efficacity. 12, 13
- [55] T. Hong, J. Langevin, and K. Sun, “Building simulation: Ten challenges,” *Building Simulation*, vol. 11, pp. 871–898, 2018. Version papier aux Mines. 18, 20
- [56] R. Judkoff, “Validation of building energy analysis simulation programs at the solar energy research institute,” *Energy and Buildings*, 1988. 18, 21, 30, 162
- [57] E. P. D. Barrio and G. Guyon, “Theoretical basis for empirical model validation using parameters space analysis tools,” *Energy and Buildings*, vol. 35, pp. 985–996, 2003. 19, 20, 21, 102, 162
- [58] E. M. Ryan and T. F. Sanquist, “Validation of building energy modeling tools under idealized and realistic conditions,” *Energy and Buildings*, vol. 47, pp. 375–382, 2012. Version papier aux Mines. 20, 27
- [59] S. Bontemps, A. Kaemmerlen, G. Blatnian, and L. Mora, “Reliability of dynamic simulation models for building energy in the context of low-energy buildings,” pp. 1952–1959, 2013. 20
- [60] C. Spitz, “Analyse de la fiabilité des outils de simulation et des incertitudes de métrologie appliquée à l’efficacité énergétique des bâtiments,” 2012. 20, 21
- [61] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte, and J. Glazer, “Energyplus: Creating a new-generation building energy simulation program,” *Energy and Buildings*, vol. 33, pp. 319–331, 2001. 20
- [62] E. P. del Barrio and G. Guyon, “Application of parameters space analysis tools for empirical model validation,” *Energy and Buildings*, vol. 36, pp. 23–33, 2004. 20

- [63] R. D. Moser and T. A. Oliver, "Validation of physical models in the presence of uncertainty," 2016. Version papier aux Mines. 20, 27, 76
- [64] W. Tian, Y. Heo, P. de Wilde, Z. Li, D. Yan, C. S. Park, X. Feng, and G. Augenbroe, "A review of uncertainty analysis in building energy assessment," *Renewable and Sustainable Energy Reviews*, vol. 93, pp. 285–301, 2018. Version papier aux Mines. 20, 60
- [65] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis: The Primer*. Wiley and Sons, 2008. 21, 32, 33, 62, 98
- [66] J. Goffart, M. Rabouille, and N. Mendes, "Uncertainty and sensitivity analysis applied to hygrothermal simulation of a brick building in a hot and humid climate," *Journal of Building Performance Simulation*, vol. 10, pp. 37–57, 1 2015. 21
- [67] S. Bontemps, "Validation expérimentale de modèles : application aux bâtiments basse consommation," 2015. 21
- [68] P. Nageler, G. Zahrer, R. Heimrath, T. Mach, F. Mauthner, I. Leusbrock, H. Schranzhofer, and C. Hochenauer, "Novel validated method for gis based automated dynamic urban building energy simulations," *Energy*, vol. 139, pp. 142–154, 2017. Version papier aux Mines. 22
- [69] I. Dochev, P. Gorzalka, V. Weiler, J. E. Schmiedt, M. Linkiewicz, U. Eicker, B. Hoffschmidt, I. Peters, and B. Schroter, "Calculating urban heat demands: An analysis of two modelling approaches and remote sensing for input data and validation," *Energy & Buildings*, p. 110378, 2020. Version papier aux Mines. 22
- [70] P. Nageler, A. Koch, F. Mauthner, I. Leusbrock, T. Mach, C. Hochenauer, and R. Heimrath, "Comparison of dynamic urban building energy models (ubem): Sigmoid energy signature and physical modelling approach," *Energy and Buildings*, vol. 179, pp. 333–343, 2018. Version papier aux Mines. 23
- [71] J. N. Fonseca and M. J. O. Panao, "Monte carlo housing stock model to predict the energy performance indicators," *Energy and Buildings*, vol. 152, pp. 503–515, 2017. Version papier aux Mines. 23
- [72] E. Walter and J. H. Kämpf, "A verification of citysim results using the bestest and monitored consumption values," 2015. 23, 26
- [73] X. Yang, M. Hu, N. Heeren, C. Zhang, T. Verhagen, A. Tukker, and B. Steubing, "A combined gis-archetype approach to model residential space heating energy: A case study for the netherlands including validation," *Applied Energy*, vol. 280, p. 115953, 2020. 24, 31
- [74] A. Taniguchi, T. Inoue, M. Otsuki, Y. Yamaguchi, Y. Shimoda, A. Takami, and K. Hanaoka, "Estimation of the contribution of the residential sector to summer peak demand reduction in japan using an energy end-use simulation model," *Energy and Buildings*, 2016. Model for energy consumption in Japan based on 912 archetypes and with occupant behavior schedule model based on time of use survey. Study of several energy saving policies to reduce summer peak consumption.
Validation by comparison with data measured on 1237 households during summer 2012. 24
- [75] D. Fischer, A. Härtl, and B. Wille-Haussmann, "Model for electric load profiles with high time resolution for german households," *Energy and Buildings*, 2015. Presentation and validation of a model simulating electric load profile per household, no UEM, but validation of the model against electric load profiles measured for a year with a timestep of 1 hour. 24

- [76] D. Fischer, T. Wolf, J. Scherer, and B. Wille-Hausmann, “A stochastic bottom-up model for space heating and domestic hot water load profiles for german households,” *Energy and Buildings*, vol. 124, pp. 120–128, 2016. Model for space-heating and domestic hot water based on synPRO. Validation with 100 of households whose heat pumps operations were recorded over nearly two year. Stochastic behavior for occupancy modelling. Correlation coefficient obtained around 0.92. 24
- [77] F. Bünning, A. Bollinger, P. Heer, R. S. Smith, and J. Lygeros, “Empirical validation of a data-driven heating demand simulation with error correction methods,” 2019. 24
- [78] R. Nouvel, M. Zirak, V. Coors, and U. Eicker, “The influence of data quality on urban heating demand modeling using 3d city models,” *Computers, Environment and Urban Systems*, vol. 64, pp. 68–80, 2017. Nouvel - SimStadt

Impact of level of details of the GIS, of year of refurbishment, window-to-wall ratio, meteorologic file. 24, 28
- [79] V. Todeschi, R. Boghetti, J. H. Kämpf, and G. Mutani, “Evaluation of urban-scale building energy-use models and tools — application for the city of fribourg , switzerland,” *Sustainability*, 2021. 24, 28
- [80] A. Malhotra, J. Bischof, A. Nichersu, K.-H. Häfele, J. Exenberger, D. Sood, J. Allan, J. Frisch, C. van Treeck, J. O’Donnell, and G. Schweiger, *Information modelling for urban building energy simulation—A taxonomic review*. Elsevier Ltd., 2021. Version papier aux Mines. 27
- [81] E. Prataiviera, J. Vivian, G. Lombardo, and A. Zarrella, “Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis,” *Applied Energy*, vol. 311, p. 118691, 2022. 27, 29
- [82] G. Happle, J. A. Fonseca, and A. Schlueter, “Effects of air infiltration modeling approaches in urban building energy demand forecasts,” vol. 122, pp. 283–288, Elsevier Ltd, 2017. 27
- [83] J. Chapman, P. O. Siebers, and D. Robinson, “On the multi-agent stochastic simulation of occupants in buildings,” *Journal of Building Performance Simulation*, vol. 11, pp. 604–621, 9 2018. Version papier aux Mines. 27
- [84] E. Garreau, “Développement d’une méthodologie d’analyse de la parcimonie pour la simulation énergétique urbaine,” 2021. 27, 37
- [85] P. J. Fennell, P. A. Ruysevelt, Érika Mata, and M. Jakob, “A review of the status of uncertainty and sensitivity analysis in building-stock energy models,” 2019. 28, 29, 162
- [86] A. T. Booth, R. Choudhary, and D. J. Spiegelhalter, “Handling uncertainty in housing stock models,” *Building and Environment*, vol. 48, pp. 35–47, 2012. 28
- [87] J. F. Rosser, D. S. Boyd, G. Long, S. Zakhary, Y. Mao, and D. Robinson, “Predicting residential building age from map data,” *Computers, Environment and Urban Systems*, vol. 73, pp. 56–67, 1 2019. Version papier aux Mines. 28, 30, 44, 163, 165
- [88] SDES, “Données locales de consommation d’énergie,” 2021. 31
- [89] A. Neale, M. Kummert, and M. Bernier, “Development of a stochastic virtual smart meter data set for a residential building stock—methodology and sample data,” *Journal of Building Performance Simulation*, vol. 13, pp. 583–605, 2020. 31
- [90] W. Walker, P. Harremoës, J. Rotmans, J. van der Sluijs, M. van Asselt, P. Janssen, and M. K. von Krauss, “Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support,” *Integrated Assessment*, vol. 4, pp. 5–17, 2003. 32

-
- [91] W. L. Oberkampf, J. C. Helton, and K. Sentz, “Mathematical representation of uncertainty,” *19th AIAA Applied Aerodynamics Conference*, 2001. 32
- [92] G. Geenens, “Copula modeling for discrete random vectors,” *Dependence Modeling*, vol. 8, pp. 417–440, 2020. 32
- [93] J. C. for Guides in Metrology, “Evaluation of measurement data — supplement 1 to the “guide to the expression of uncertainty in measurement” — propagation of distributions using a monte carlo method,” *Evaluation*, vol. JCGM 101:2, p. 90, 2008. 32, 61
- [94] C. Soize, *Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering*, vol. 47. 2017. 32, 61, 80
- [95] G. Ansanay-Alex, Y. Abdelouadoud, and P. Schetelat, “Statistical and stochastic modelling of french households and their energy consuming activities,” 2016. 36, 41
- [96] V. Kuleshov and S. Ermon, “Probabilistic graphical models - introductory course,” 2022. 41, 43
- [97] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. 2009. 43
- [98] INSEE, “Logements ordinaires en 2016 - recensement de la population - fichiers détail,” 2016. 43, 68
- [99] C. N. de l’Information Statistique, “Enquête performance de l’habitat, Équipements, besoins et usages de l’énergie (phébus),” 2013. 43
- [100] L. Lefort, R. B. D. Rougé, P. Schetelat, T. Berthou, P. Riederer, B. Duplessis, and E. Peirano, “Development of a methodology of validation for urban building energy models and application to french residential consumption,” 2020. 51
- [101] IGN, “Bd topo version 3.0 - descriptif de contenu,” 2021. 53, 63
- [102] G. Gröger, T. H. Kolbe, C. Nagel, and K.-H. Häfele, “Opengis city geography markup language (citygml) encoding standard, version 2.0.0,” *OGC Document No. 12-019*, p. 344, 2012. 53
- [103] CEREMA, “Les fichiers fonciers,” 2020. 53
- [104] A. H. T. Nguyen, “Modélisation du parc de bâtiments du secteur tertiaire et simulation énergétique,” 2021. 55
- [105] S. bâtiments et logements and G. et Direction fédérale des mensurations cadastrales, “Directive sur la saisie des bâtiments dans la mensuration officielle et le registre fédéral des bâtiments et des logements,” 2021. 55
- [106] CSTB, “Base de données nationale des bâtiments,” 2022. 55
- [107] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948. 61
- [108] L. Rivalin, “Vers une démarche de garantie des consommations énergétiques dans les bâtiments neufs : Méthodes d ’ évaluation des incertitudes associées à la simulation thermique dynamique dans le processus de conception et de réalisation .,” 2016. 62
- [109] X. Fu, W. Tian, Y. Sun, C. Zhu, and B. Yin, *Uncertainty Analysis of Urban Building Energy Based on Two-Dimensional Monte Carlo Method*. Springer Singapore, 2020. Version papier aux Mines. 62

- [110] INSEE, “Enquête emploi du temps,” 2010. 66
- [111] W. W. Nazaroff, “Residential air-change rates: A critical review,” *Indoor Air*, vol. 31, pp. 282–313, 2021. 67
- [112] A. Breitwiller, R. Lahrech, T. Rieser, and B. Rozel, “Projet pecoic - prise en compte du comportement des occupants et des incertitudes associées en phase conception de bâtiment,” 2019. 67, 68
- [113] M. Dupret, J.-P. Zimmermann, N. Andreau, M. Guernevel, and Enertech, “Panel usages électrodomestiques - consommations électrodomestiques françaises basées sur des mesures collectées en continu dans 100 logements,” 2021. 68, 78
- [114] M. G. Cox, P. M. Harris, and I. M. Smith, “Software specifications for uncertainty evaluation,” 2010. 80
- [115] M. D. L. T. E. Logement, “Arrêté du 8 octobre 2021 modifiant la méthode de calcul et les modalités d’établissement du diagnostic de performance énergétique,” 2021. 84
- [116] V. Lefieux, *Statistiques numériques*. 2019. 102

Résumé long en français

Chapitre 1 : Panorama des outils de simulation énergétique urbaine

Introduction

Avec 75% des consommations mondiales d'énergie primaire concentré dans les zones urbaines [15], les villes ont rapidement été identifiées comme des acteurs incontournables de la transition écologique. Elles disposent notamment d'un pouvoir politique suffisant pour implémenter de nombreuses mesures en faveur de l'environnement dans de nombreux domaines tels que la gestion des déchets, la gestion de l'eau, la biodiversité ou encore les consommations d'énergie. Il semble d'ailleurs que les bâtiments soient un angle d'attaque particulièrement intéressant pour réduire les consommations d'énergie.

Pour aider les villes à réduire les consommations d'énergie liées aux bâtiments, des outils d'aide à la décision ont été développés par la communauté scientifique parmi lesquels se trouvent les outils de simulation énergétique urbaine aussi appelés UBEM pour Urban Building Energy Models. Les UBEM sont des outils d'aide à la décision capables de modéliser les consommations d'énergie à l'échelle bâtiment sur des échelles spatiales allant généralement du quartier à la ville. L'objectif est donc d'aider les décideurs à arbitrer entre plusieurs scénarios énergétiques en modélisant l'impact de chaque scénario sur les consommations d'énergie des bâtiments.

Afin de modéliser le comportement énergétique de chaque bâtiment, les outils de simulation énergétique urbaine utilisent pour beaucoup des modèles de simulation thermique dynamique du bâtiment, mais peuvent aussi avoir recours à des modèles de machine learning ou à des modèles physiques simplifiés. Quel que soit le type de modèle choisi, l'obtention des données nécessaires à l'instanciation des outils reste la plupart du temps problématique. Par exemple, s'il est généralement possible de connaître précisément pour un bâtiment donné le type d'isolant installé ainsi que la qualité des vitrages, il est rare de disposer de toutes ces informations pour tout un quartier ou pour toute une ville. Pour palier le manque de données précises décrivant chaque bâtiment, des stratégies ont été développées afin de compléter les informations manquantes pour la simulation à l'aide d'enquêtes nationales. Ce processus est qualifié ici de processus d'enrichissement des données.

Si l'enrichissement des données est nécessaire pour permettre de modéliser plusieurs centaines de bâtiments à la fois, les hypothèses faites à cette étape risquent de nuire à la qualité des résultats obtenus en sortie de simulation par les outils de simulation énergétique urbaine. Il est donc important de vérifier que malgré toutes les hypothèses faites, les outils de simulation énergétique urbaine fournissent des résultats suffisamment fiables au regard des processus de décision dans lesquels ils sont impliqués.

L'objectif de cette thèse est donc de développer un processus de validation permettant de vérifier l'adéquation des UBEM aux processus de décision dans lesquels ils

sont utilisés.

Pour cela, il est nécessaire de dresser tout d'abord un panorama complet des outils de simulation énergétique urbaine puis de se pencher sur les différents processus de décision dans lesquels ces outils sont impliqués.

Panorama des outils de simulation énergétique urbaine

Les outils de simulation énergétique urbaine ont donc été définis comme des outils d'aide à la décision capable de modéliser le comportement énergétique de plusieurs bâtiments en même temps. La procédure de simulation implémentée dans les outils de simulation énergétique urbaine est présentée Figure 1.



Figure 1: Principales étapes de simulation pour les outils de simulation énergétique urbaine

Elle se compose de trois grandes étapes :

1. Collecte des données disponibles pour décrire chaque bâtiment à modéliser;
2. Enrichissement des données manquantes à l'aide d'enquêtes nationales;
3. Modélisation du comportement physique de chaque bâtiment.

Les données disponibles à l'échelle bâtiment sont généralement assez similaires d'un cas à l'autre et contiennent souvent la géométrie du bâtiment et son année de construction. Pour ce qui est de la modélisation du comportement physique du bâtiment, cela fait généralement intervenir plusieurs modèles interconnectés modélisant quatre grands phénomènes : l'occupation, la météo, les systèmes et l'enveloppe du bâtiment. Pour chacune de ces catégories, différents niveaux de détail de modèles peuvent être utilisés dans les UBEM.

De même, différentes approches existent pour le processus d'enrichissement des données permettant de représenter les bâtiments avec une diversité plus ou moins importante suivant l'approche choisie.

Processus de décisions associés aux outils de simulation énergétique urbaine

Une revue des processus de décisions associés aux outils de simulation énergétique urbaine est présentée en détail dans ce chapitre et est synthétisée Figure 2.

La plupart des cas d'usages des outils de simulation énergétique urbaine, ainsi que les processus de décision associés, découlent de la volonté des villes d'agir face au changement climatique. Pour chaque cas d'usage identifié Figure 2, la sortie de l'outil de simulation énergétique urbaine utilisée dans le processus de décision a été identifiée. En outre, la fiabilité de la sortie attendue au regard du processus de décision associé a également été évaluée. En effet, suivant le cas d'usage, l'impact de la sortie d'un UBEM sur la décision finale peut beaucoup varier et le niveau d'attente quant à la fiabilité de la sortie n'est donc pas le même.

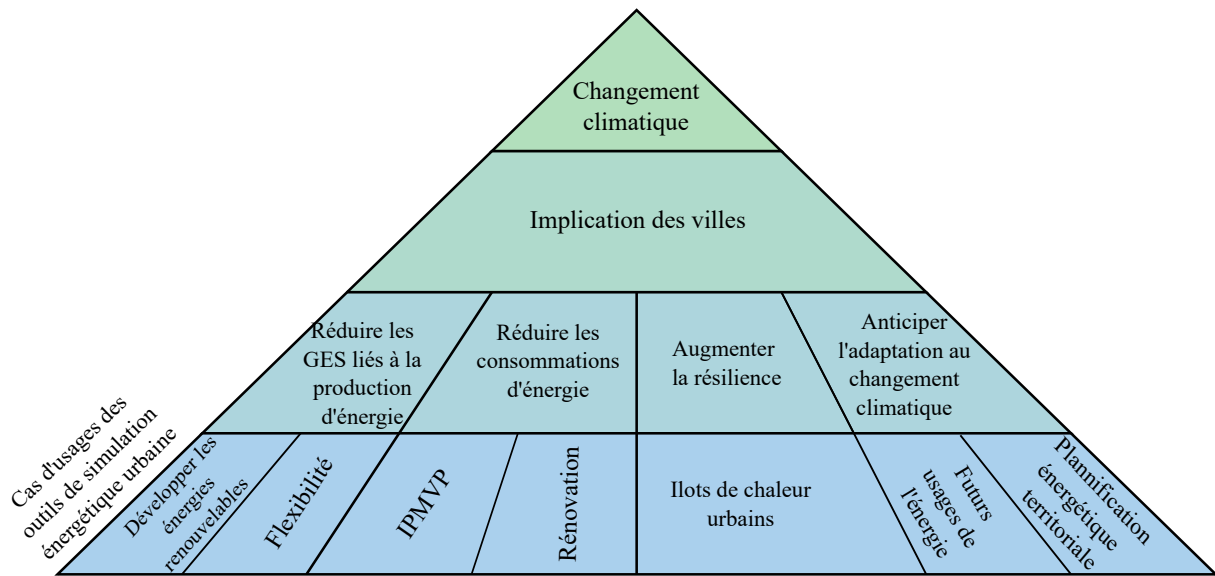


Figure 2: Cas d'usages associés aux UBEM

Chapitre 2 : Les outils de simulation énergétique urbaine et la validation

Etat de l'art de la validation appliquée aux outils de simulation thermique dynamique du bâtiment

Puisque de nombreux outils de simulation énergétique urbaine s'appuient sur des outils de simulation thermique dynamique (STD) du bâtiment, il semblait pertinent de commencer par réaliser un état de l'art des procédures de validation existant pour les outils de STD du bâtiment.

Parmi les projets de recherche les plus célèbres dans le domaine, la procédure BESTEST [7, 56] permet d'évaluer les outils sur un cas test synthétique parfaitement décrit. Plusieurs variantes de ce cas-test existent, permettant ainsi de tester différentes parties des outils et d'isoler de potentielles faiblesses.

Un autre projet similaire, le projet PASSYS [6], propose de découper la procédure de validation en deux grandes parties :

- La validation de chaque process pris indépendamment ;
- La validation du modèle dans son ensemble.

Cette approche permet ainsi de procéder à la fois à l'évaluation du modèle mais également au diagnostic des faiblesses potentielles de l'outil.

Cependant, aucun de ces deux projets ne présente une validation empirique sur des bâtiments réels. Cet aspect est étudié par Del Barrio et Guyon [57] qui proposent notamment de prendre en compte l'incertitude entourant les outils de STD du bâtiment dans la validation.

Si ces approches de validation utilisées à l'échelle bâtiment sont très intéressantes pour construire une procédure de validation des outils à l'échelle urbaine, elles ne sont pas suffisantes pour évaluer les UBEM puisqu'elles ne prennent pas en compte le processus d'enrichissement des données.

Etat de l'art de la validation appliquée aux outils de simulation énergétique urbaine

Concernant la validation appliquée aux UBEM, de nombreux exemples de validation empirique peuvent être trouvés dans la littérature scientifique. Néanmoins, on observe pour l'évaluation d'une même sortie beaucoup de métriques différentes ce qui rend impossible la comparaison des résultats d'un article à l'autre. Par ailleurs, les exemples de validation empirique présents dans la littérature sont souvent dépourvus de cadre méthodologique et les jeux de données utilisés dans l'exercice de validation ne sont généralement pas mis à disposition, empêchant ainsi de reproduire les résultats présentés.

Si l'on s'intéresse à la comparaison inter-modèles, un groupe de travail a entrepris de créer un DESTEST [8] à l'image des BESTEST. Ce DESTEST est composé d'un quartier virtuel de seize bâtiments strictement identiques reliés par un réseau de chaleur. Si ce cas-test permet d'étudier les performances de composantes propres aux UBEM telle que la modélisation d'un réseau de chaleur, cette approche ne prend pas en compte le processus d'enrichissement des données et ne permet donc pas d'évaluer les outils dans leurs conditions réelles d'utilisation.

Par ailleurs, on trouve très peu d'analyses de sensibilité appliquées aux UBEM comme le soulignent Fennell et al. [85]. Cela signifie que l'étape de diagnostic liée à tout processus de validation est encore très peu développée pour les outils de simulation énergétique urbaine.

Proposition d'un cadre méthodologique pour la validation des outils de simulation énergétique urbaine

Afin de dépasser les limitations constatées lors de l'état de l'art, un cadre méthodologique pour la validation des UBEM est proposé dans cette thèse et est décrit Figure 3.

La méthode proposée se décompose en plusieurs étapes:

1. La validation ad hoc du processus d'enrichissement et du modèle physique : il s'agit d'évaluer a priori les deux principaux composants d'un UBEM. L'évaluation a priori du modèle physique se fait à l'aide des BESTEST et des DESTEST tandis que l'évaluation a priori du processus d'enrichissement peut se faire à l'aide de techniques classiques de machine learning comme proposé par Rosser et al. [87].
2. La revue des cas d'usages des UBEM et l'identification des sorties à évaluer pour vérifier la fiabilité des outils sur chaque cas d'usage.
3. La création d'un jeu de validation : une fois qu'il a été décidé d'appliquer la méthode de validation pour évaluer une sortie précise d'un UBEM, un jeu de données devrait être construit afin de permettre d'évaluer la sortie choisie (aussi appelée KPI pour Key Priority Indicator). Ce jeu de données de validation doit fournir des données mesurées de la sortie à évaluer ainsi qu'une description aussi précise que possible des bâtiments sur lesquels le KPI a été mesuré.
4. La quantification de l'incertitude : afin d'évaluer les UBEM dans leur contexte d'utilisation, l'incertitude qui les entoure doit être prise en compte. La façon de modéliser l'incertitude sur les différents paramètres est décidée à l'étape de quantification de l'incertitude et peut avoir un grand impact sur les résultats obtenus après propagation de cette incertitude.
5. La propagation de l'incertitude : à l'étape précédente, plusieurs jeux de paramètres ont été obtenus pour chaque bâtiment afin de prendre en compte l'incertitude qui existe sur les paramètres d'entrée. Chacune de ces jeux de paramètres est ensuite utilisés pour simuler le comportement énergétique du bâtiment ce qui permet d'obtenir une distribution des valeurs possibles du KPI étudié.
6. L'évaluation du KPI : l'erreur faite par l'outil sur la sortie que l'on cherche à évaluer est quantifiée à l'aide d'une métrique définie a priori.
7. Le diagnostic des faiblesses de l'outil : Le but de cette étape est d'identifier les paramètres responsables de l'erreur observée à l'étape précédente.



Figure 3: Proposition d'un cadre méthodologique pour la validation des outils de simulation énergétique urbaine

Chapitre 3 : Application du cadre méthodologique de validation à un premier ensemble de cas d’usages

Validation ad hoc du modèle physique et du processus d’enrichissement

La validation ad hoc est présentée ici pour Dimosim qui est un UBEM déjà largement utilisé et pour Qiriel, un outil dédié au processus d’enrichissement des données basé sur des réseaux bayésiens. La validation ad hoc de Dimosim se fait en évaluant ses résultats sur les BESTEST ainsi que sur le DESTEST. Dans les deux cas, Dimosim présente des performances comparables aux outils de sa catégorie. Pour la validation ad hoc de Qiriel, les métriques proposées par Schetelat et al. [46] ont été appliquée dans un contexte de validation croisée comme proposé par Rosser et al. [87].

Choix de cas d’usages pour appliquer la méthode proposée

Afin de tester la méthode proposée, un ensemble de cas d’usages basés sur le même indicateur ont été sélectionnés. Cette sélection s’est faite à partir des données disponibles afin de pouvoir utiliser autant que possible des données publiques pour la création du jeu de données de validation. Les KPI retenus ont donc été la consommation annuelle de gaz et d’électricité à l’échelle bâtiment. Des mesures de ces deux grandeurs sont fournies par les Données Locales de l’Énergie pour tous les bâtiments résidentiels de plus de 10 logements ou consommant plus de 200 MWh par an. Les cas d’usages correspondant à ces deux indicateurs sont les suivants :

- Identification de bâtiments à rénover en priorité sur un quartier ;
- Évaluation des économies d’énergie réalisées grâce à la rénovation d’un groupe de bâtiments ;
- Évaluation du potentiel d’extension d’un réseau de chaleur.

Création d’un jeu de données de validation

Afin de créer un jeu de données de validation sur lequel tester les deux KPI sélectionnés, plusieurs bases de données ont été utilisées :

- La BD TOPO[®] : une cartographie en 2.5D de tous les bâtiments du territoire français. Cette base de données permet ainsi d’obtenir une géométrie grossière de chaque bâtiment (emprise au sol et hauteur) et de déterminer les murs mitoyens de chaque bâtiment ainsi que les masques solaires.
- Les fichiers fonciers : il s’agit d’informations collectées par le ministère des Finances afin de faire payer la taxe foncière. De nombreuses informations sont fournies à la parcelle cadastrale contenant le bâtiment telles que le nombre de locaux du bâtiment, l’usage spécifique de chaque local ou encore l’année de construction du bâtiment.
- Les diagnostics de performance énergétique (DPE) : cette base de données mise à disposition par l’ADEME réunit tous les DPE réalisés depuis 2012. Un DPE est un audit d’un logement réalisé par un diagnostiqueur certifié et obligatoire lors de la vente ou de la location d’un bien immobilier. Les DPE fournissent des renseignements sur les systèmes de chauffage et d’ECS installés ainsi que sur les propriétés thermiques du bâti. Malheureusement tous les bâtiments en France ne dispose pas à l’heure actuelle d’un DPE et ceux collectés depuis 2012 montrent certains biais.

- Les Données Locales de l'Énergie : Cette base de données contient les consommations annuelles de gaz et d'électricité localisées à l'adresse pour les bâtiments résidentiels de plus de dix logements ou pour les bâtiments tertiaires et industriels consommant plus de 50MWh par an. Elle est mise à jour chaque année par les gestionnaires de réseaux de distribution.

Un travail de croisement de ces bases de données a été réalisé par le CSTB afin de créer la BDNB, la Base de Données Nationale des Bâtiments. De cette base, un jeu de validation a été extrait contenant 1200 bâtiments résidentiels.

Une présentation synthétique de ce jeu de données de validation est présenté Figure 4.

Geographical distribution across France of the buildings selected in the validation dataset

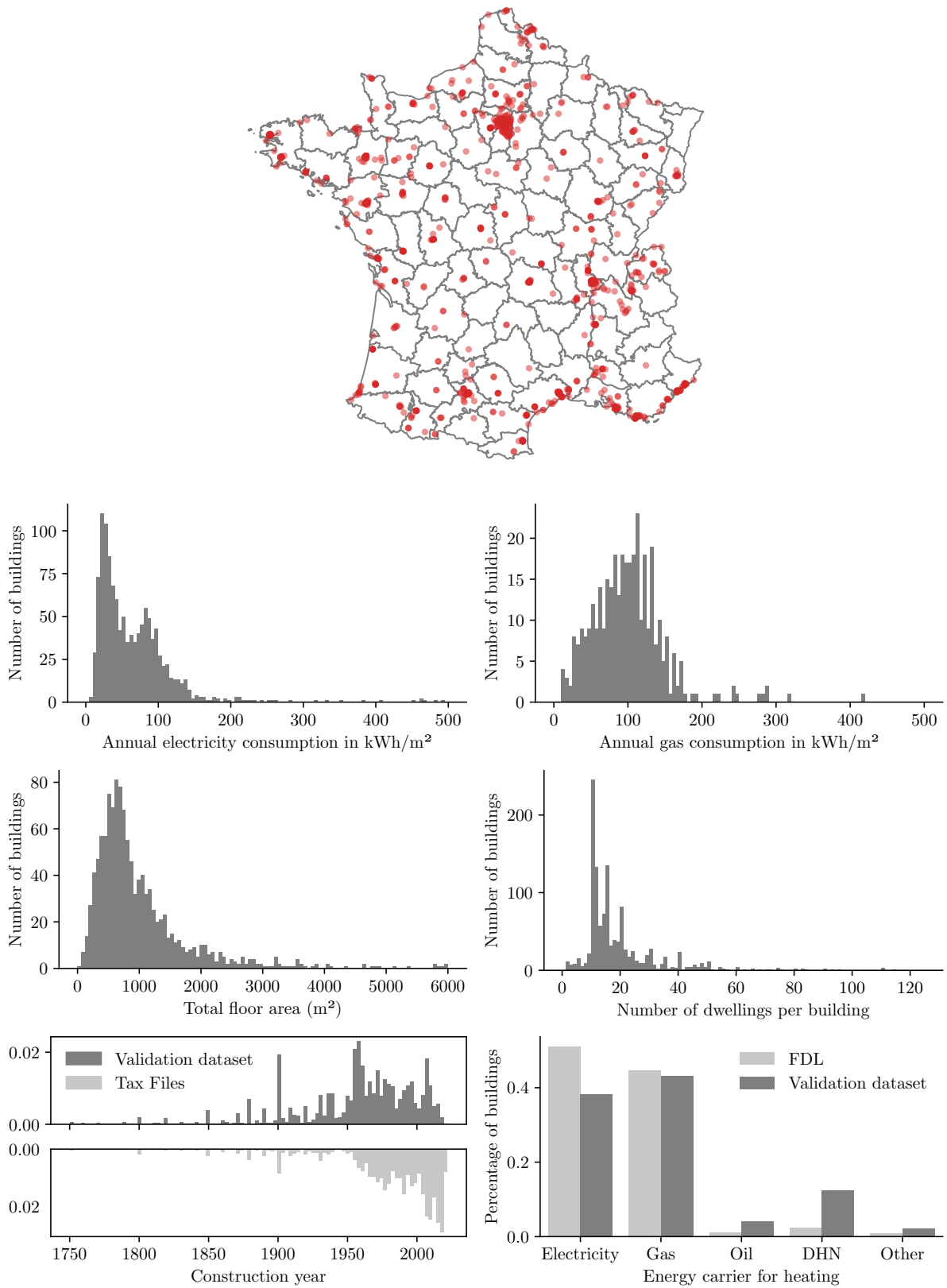


Figure 4: Présentation du jeu de données de validation retenu pour la suite du manuscrit

Chapitre 4 : Évaluation des outils de simulation énergétique urbaine

Ce chapitre présente l'étape d'évaluation proposée dans la procédure de validation. Dans l'approche proposée ici, les outils sont évalués en prenant en compte les incertitudes entourant les données disponibles pour leur paramétrage.

La première étape consiste donc à quantifier l'incertitude sur les données d'entrée.

Quantification de l'incertitude

L'approche choisie dans ce manuscrit pour la quantification de l'incertitude est l'approche probabiliste dans laquelle l'incertitude sur une variable est représentée à l'aide d'une distribution de probabilité. La quantification de l'incertitude se décompose alors en deux étapes :

1. La définition des distributions de probabilité associées à chaque variable incertaine ;
2. L'échantillonnage des distributions de probabilités pour générer des jeux de variables probables.

Pour la définition des distributions de probabilité, les variables indépendantes sont séparées des variables corrélées. En effet, dans les données d'entrée d'un outil de simulation énergétique urbaine, certaines variables sont liées et ne devraient pas être prédites indépendamment. Par exemple, la valeur du coefficient de transmission thermique du toit est souvent lié au coefficient de transmission thermique des murs puisque pour les bâtiments récents ces deux valeurs devraient être assez basses. De même, il serait peu réaliste de prédire beaucoup de bâtiments avec une valeur U_{mur} plus faible que U_{toit} .

Les distributions de probabilité associées à ces variables corrélées sont alors représentées par des distributions multivariées qui peuvent engendrer des problèmes de grande dimension. Pour modéliser ces distributions multivariées, il a été choisi dans ce manuscrit d'utiliser des réseaux bayésiens. Les réseaux bayésiens permettent en effet de simplifier ces distributions multivariées en faisant certaines hypothèses d'indépendances.

Pour échantillonner le réseau bayésien et prédire des jeux de variables probables, une méthode de chaîne de Markov par Monte Carlo, l'échantillonneur de Gibbs, est utilisée.

La dépendance ou non de deux variables est quant à elle étudiée à l'aide de l'information mutuelle, un outil statistique similaire au coefficient de corrélation mais étudiant l'ensemble des relations entre deux variables et non seulement les relations linéaires.

L'incertitude sur l'ensemble des données d'entrée habituellement requises par un UBEM est ainsi définie en s'appuyant sur des enquêtes nationales ou sur des règles tirées de la littérature.

Application de la procédure d'évaluation à un outil simplifié, Simply

Afin de mettre l'accent sur la méthode proposée, celle-ci est appliquée dans un premier temps à un UBEM créé pour l'occasion, Simply, et dont le modèle physique est extrêmement simple.

La méthode d'évaluation proposée permet d'évaluer les outils à l'échelle de chaque bâtiment mais également à l'échelle de tout l'échantillon de validation.

Évaluation à l'échelle bâtiment A l'échelle bâtiment, on évalue l'écart entre la donnée mesurée et les valeurs calculées par l'outil à l'aide du pourcentage d'erreur. Comme chaque

bâtiment a été simulé avec une centaine de configurations différentes obtenues lors de la quantification de l'incertitude, on obtient une distribution de valeurs pour le pourcentage d'erreur. Cette distribution du pourcentage d'erreur est caractérisée par le résidu qui est la valeur médiane, μ_{bldg} , et par l'écart type.

Évaluation à l'échelle de tout l'échantillon de validation Le processus décrit au-dessus est appliqué à tous les bâtiments de l'échantillon. On obtient ainsi entre autre une distribution des résidus qui caractérise le comportement de l'outil sur tout l'échantillon de validation.

En outre, il est possible de définir pour chaque bâtiment l'équivalent d'un intervalle de confiance : la HPD région. On peut alors évaluer pour chaque bâtiment si la valeur mesurée se trouve dans l'intervalle de confiance (HPD région) des valeurs calculées avec l'outil. En comptant à l'échelle de tout l'échantillon de validation le pourcentage de bâtiments pour lesquels la donnée mesurée est dans la HPD région, on obtient un indicateur synthétique caractérisant la performance de l'outil : la couverture empirique.

Etude de convergence Comme les résultats sont obtenus suite à une analyse d'incertitude, il est important de vérifier que la convergence des calculs a bien été atteinte. Une analyse de convergence a donc été menée avec Simply pour s'assurer de la fiabilité des résultats obtenus.

Application de la procédure d'évaluation à trois outils

La procédure d'évaluation présentée ci-dessus est appliquée à deux autres outils, Calomel et Dimosim, qui présentent un niveau de détail croissant de leur modèle physique.

Dimosim obtient de très bonnes performances sur la modélisation des consommations d'électricité tandis que Simply semble être plus ajusté que les deux autres outils sur les consommations de gaz.

Chapitre 5 : Diagnostic, comprendre les faiblesses des UBEM

Dans le chapitre précédent, une procédure d'évaluation des outils UBEM basée sur l'incertitude a été introduite ainsi qu'un ensemble de métriques permettant de quantifier correctement les performances des outils. Ces métriques ont permis d'attribuer un score à chaque outil testé. Ce score est rarement aussi élevé que ce que le modélisateur pourrait attendre. Une étape de diagnostic est alors nécessaire suite à l'étape d'évaluation pour comprendre où se situent les faiblesses des outils et comment améliorer le score de l'outil. Plusieurs raisons peuvent expliquer un mauvais score, dont notamment :

- Les données d'entrée sont de mauvaise qualité ("garbage in, garbage out") ;
- Le processus d'enrichissement est défectueux ;
- Le modèle physique est défectueux ;
- Les hypothèses faites pour la quantification de l'incertitude sont inappropriées.

Ces raisons peuvent être regroupées en trois sources d'erreur principales :

- les données ;

- l'UBEM lui-même ;
- l'application du cadre de validation.

Il est difficile de faire la distinction entre ces trois sources, mais pour aider à clarifier le sujet, plusieurs sous-questions peuvent être étudiées :

1. Est-il possible d'identifier les paramètres qui ont une influence sur l'erreur à l'échelle de l'échantillon ?
2. Est-il possible d'identifier les paramètres influençant l'erreur pour chaque bâtiment ?
3. Est-il possible d'identifier des groupes de bâtiments ayant un comportement similaire vis-à-vis de l'erreur ?

Aucune de ces questions ne permet de distinguer strictement les trois sources d'erreur en soi, mais les résultats obtenus pour chacune d'elles lorsqu'elles sont comparées aux autres devraient permettre d'identifier des tendances et de comprendre si l'une des sources (les données, l'UBEM ou l'application du cadre de validation) est davantage responsable de l'erreur que les deux autres. Les résultats obtenus par Simply, Calomel et Dimosim sont étudiés à la lumière de ces trois sous-questions. Etant donné que la même procédure de validation a été appliquée aux trois outils avec le même processus d'enrichissement, la comparaison du comportement affiché par chacun de ces outils sur chacune des sous-questions devrait permettre de comprendre plus précisément l'influence du modèle physique sur l'erreur.

Les utilisateurs d'UBEM sont souvent confrontés à différents niveaux de disponibilité des données d'une étude de cas à l'autre, ce qui modifie la façon dont le modèle physique de l'UBEM est paramétré. Pour comprendre comment ce changement dans le paramétrage du modèle peut influencer les résultats d'un UBEM, une étude est menée avec différents niveaux de disponibilité des données. L'approche choisie ici a été de ne pas modéliser d'incertitude sur les données jugées disponibles.

Identification des paramètres qui ont une influence sur l'erreur à l'échelle de l'échantillon

A cette étape, l'information mutuelle entre le résidu obtenu pour chaque bâtiment et les caractéristiques principales du bâtiment est calculée. Il ressort que l'énergie de chauffage ainsi que l'énergie d'ECS influencent l'erreur observée.

Identification des paramètres qui ont une influence sur l'erreur à l'échelle du bâtiment Pour chaque bâtiment, un vecteur d'information mutuel entre le pourcentage d'erreur et les paramètres incertains est calculé.

Création de groupes de bâtiments avec un comportement similaire face à l'erreur

Une analyse en composantes principales (ACP) est réalisée sur les vecteurs d'information mutuelle calculés précédemment. Cela permet d'identifier des groupes de bâtiments qui se distinguent par leur énergie de chauffage et leur énergie d'ECS. Pour chacun de ces groupes, les paramètres avec le plus d'influence sur l'erreur sont identifiés.

Application du diagnostic aux autres outils

L'étape de diagnostic appliquée aux trois outils a permis d'identifier des groupes de bâtiments ayant un comportement similaire vis-à-vis de l'erreur et d'identifier les paramètres d'entrée influents pour chaque groupe. Parmi les raisons indiquées en introduction de ce chapitre qui

pourraient expliquer l'écart des outils par rapport à la réalité, cette étape de diagnostic a permis d'en écarter au moins une. En effet, les tendances identifiées étant assez similaires d'un outil à l'autre, cela indique que l'erreur principale est susceptible de provenir d'une autre source que les modèles physiques. Les causes restantes de l'erreur sont alors :

- La qualité des données ;
- Le processus d'enrichissement ;
- La façon dont l'incertitude a été prise en compte.

Il est assez difficile d'exclure la qualité des données puisqu'il est difficile d'obtenir une référence qui pourrait permettre de caractériser le niveau de qualité des données. Cependant, il est possible d'étudier comment le processus d'enrichissement et l'incertitude modifient le comportement des outils. L'impact de ces deux processus porte sur le paramétrage du modèle. Afin d'étudier l'impact du paramétrage du modèle, les procédures d'évaluation et de diagnostic ont été appliquées aux trois outils pour deux autres niveaux de données d'entrée disponibles. Ces travaux sont présentés ci-après.

Étude de l'influence du niveau de données disponibles sur les performances des UBEM

Des simulations ont été réalisées avec les trois outils pour trois niveaux de données disponibles et les résultats comparés. Bien que cette enquête n'ait pas permis d'évaluer pleinement l'influence de la qualité des données, elle a permis d'étudier l'impact du paramétrage du modèle. En effet, avec plus de données disponibles, le niveau d'incertitude pris en compte diminue (une fois que la valeur d'un paramètre est connue, aucune incertitude n'est prise en compte dans ce cas) ainsi que l'influence du processus d'enrichissement (puisque la valeur du paramètre est connue, le processus d'enrichissement n'est plus nécessaire pour cette entrée). Ainsi, si l'erreur diminue lorsque le niveau de disponibilité des données augmente, cela implique que le processus d'enrichissement est biaisé et génère une grande partie de l'erreur. Au contraire, si l'erreur n'est pas impactée par l'ajout d'informations, cela implique que le modèle physique utilisé n'est pas sensible aux nouvelles informations.

Il est apparu que la quantification de l'incertitude pilotait complètement le comportement de l'outil lorsque presque aucune information n'était extraite du jeu de données. Ensuite, lorsque plus d'informations sont disponibles, une relation claire, bien que biaisée, apparaît entre les consommations simulées et mesurées. Ce biais semble être principalement dû à des paramètres pour lesquels aucune information n'était disponible : la consommation d'eau chaude sanitaire (ECS) et les consommations des appareils électroménagers. Dans ce cas, cela implique que la quantification de l'incertitude sur ces paramètres n'est pas bien ajustée. Disposer d'informations sur ces paramètres devrait alors permettre d'améliorer significativement les résultats obtenus par les trois outils.

Chapitre 6 : Conclusions et perspectives

Comme il a été souligné ci-dessus, un cadre de validation dédié aux UBEMs a été proposé dans cette thèse et appliqué à un premier ensemble de cas d'usages et à trois outils différents. Cependant, plusieurs aspects du travail présenté ici nécessiteraient des recherches supplémentaires.

Premièrement, comme expliqué ci-dessus, la quantification de l'incertitude influence grandement les résultats de la procédure de validation et, en tant que telle, cette étape devrait être remise en

question par d'autres chercheurs. Les hypothèses retenues pour modéliser l'incertitude sur chaque variable d'entrée pourraient alors être examinées et leur impact sur les résultats finaux pourrait être évalué. Par exemple, aucune incertitude n'a été considérée sur les paramètres liés à la météo, mais l'incertitude sur des paramètres tels que la température extérieure est probablement élevée, notamment en zone urbaine en raison de l'effet d'îlot de chaleur urbain. La modélisation de l'incertitude sur la température extérieure permettrait d'évaluer les outils dans des conditions plus réalistes et ainsi d'évaluer l'influence de ce paramètre sur l'erreur.

Deuxièmement, seul un processus d'enrichissement spécifique, Qiriel, a été étudié dans ce manuscrit. Il serait intéressant de tester les mêmes modèles physiques avec différentes approches pour le processus d'enrichissement. Il serait notamment intéressant d'étudier les performances des outils lorsque le processus d'enrichissement est basé sur des archétypes déterministes.

Troisièmement, afin de comprendre la robustesse des résultats obtenus ici par les trois outils, il serait intéressant d'exécuter le processus de validation pour les mêmes cas d'usages et pour les mêmes outils, mais avec un ensemble de données de validation complètement différent. Des bases de données similaires à celle utilisée pour créer l'ensemble de données de validation utilisé dans ce manuscrit sont disponibles dans la plupart des pays européens et aux Etats-Unis. Il serait donc possible de créer des ensembles de données de validation similaires et d'étudier l'évolution des performances des outils d'un ensemble de données de validation à l'autre. Enfin, et c'est le plus important, le cadre de validation appliqué ici pour évaluer les performances des outils sur les consommations annuelles d'énergie à l'échelle du bâtiment doit être adapté à l'évaluation de sorties dynamiques telles que les courbes de charge électrique ou les courbes de charge de chauffage. La méthode générale telle qu'introduite au chapitre 2 devrait être applicable telle quelle, mais les indicateurs fournis ici sont bien adaptés à l'évaluation des sorties statiques uniquement, et d'autres indicateurs doivent être proposés dans le cas de sorties dynamiques.

En résumé, cette thèse a initié le développement d'un cadre de validation dédié aux UBEM. De nombreuses faiblesses des outils ont été mises en évidence ainsi que l'influence de l'incertitude sur la valeur affichée par une sortie UBEM. Chaque modélisateur devrait toujours garder à l'esprit que même si un outil a été testé à travers un cadre de validation, il ne doit jamais être considéré comme entièrement validé, en particulier dans le cas des UBEM où tant d'incertitudes peuvent interférer. Les résultats fournis par un UBEM doivent donc toujours être pris avec précaution et une indication sur le niveau d'incertitude doit toujours être fournie avec les résultats eux-mêmes. En fin de compte, comprendre et traiter les faiblesses des UBEM aiderait ces outils à fournir un soutien plus adéquat et plus fiable aux décideurs dans le processus de prise de décision pour lequel ils sont utilisés.

Appendices

Appendices to chapter 3

Ad hoc validation of the Bayesian network trained on PHEBUS

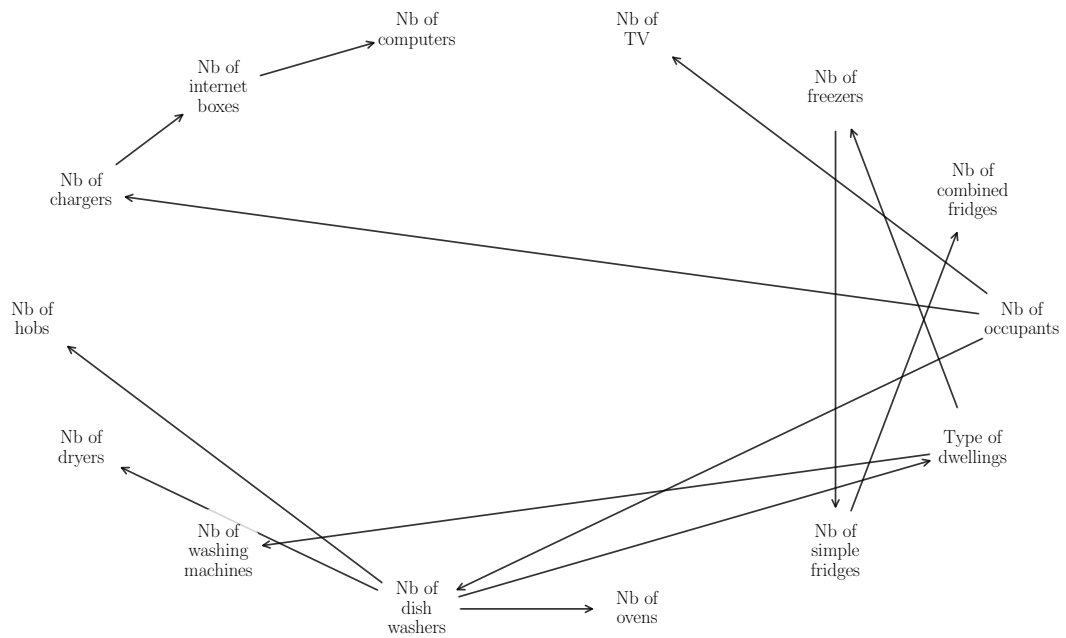


Figure 1: Bayesian network obtained by training on PHEBUS.

	Log-loss	RIG
Nb of combined fridges	1.08 (± 0.02)	0.01 (± 0.02)
Nb of freezers	1.12 (± 0.03)	0.10 (± 0.03)
Nb of TV	1.49 (± 0.02)	0.02 (± 0.01)
Nb of computers	1.23 (± 0.01)	0.04 (± 0.01)
Nb of internet boxes	0.76 (± 0.03)	0.13 (± 0.04)
Nb of chargers	2.28 (± 0.04)	0.11 (± 0.02)
Nb of hobs	1.06 (± 0.02)	-0.00 (± 0.02)
Nb of dryers	0.90 (± 0.01)	0.06 (± 0.02)
Nb of washing machines	0.34 (± 0.03)	0.06 (± 0.09)
Nb of dish washers	0.88 (± 0.04)	0.13 (± 0.04)
Nb of ovens	0.72 (± 0.03)	0.02 (± 0.05)
Nb of simple fridges	1.09 (± 0.01)	0.03 (± 0.01)

Table 1: Single feature performances assessment for the appliances network.

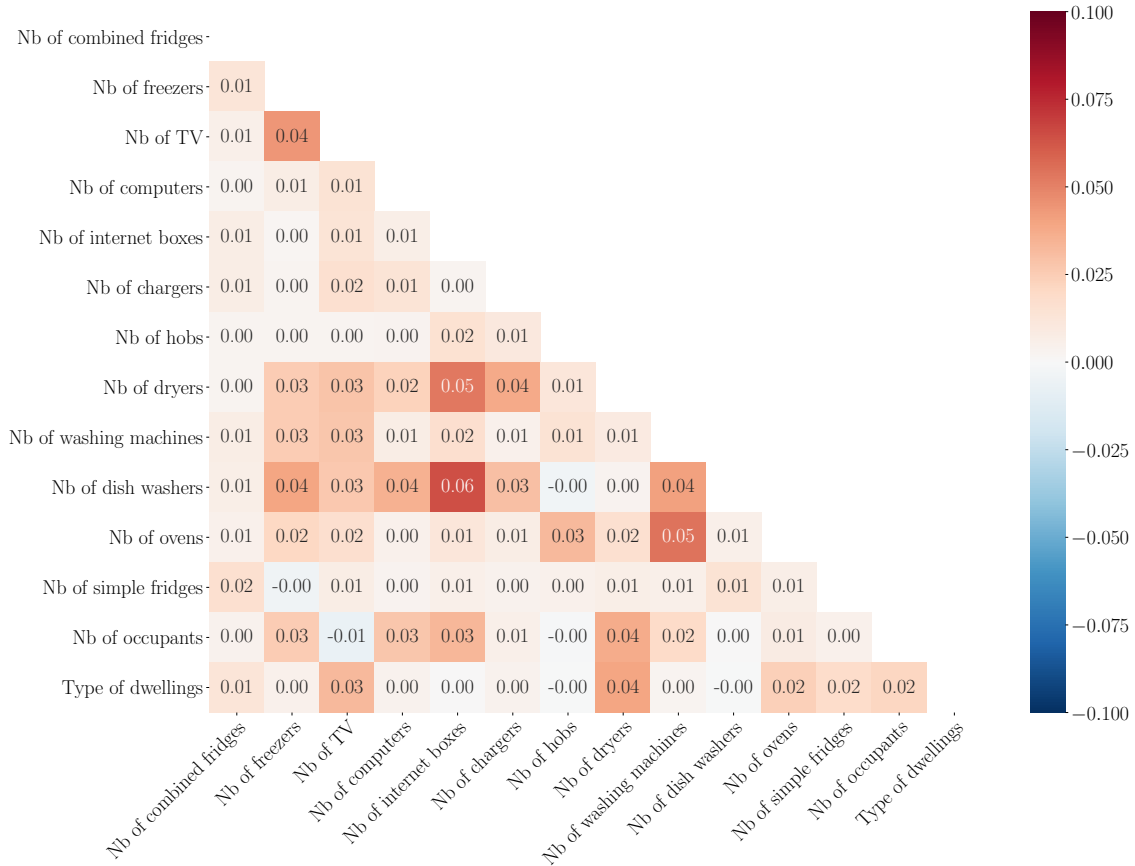


Figure 2: Mutual Information loss obtained for the appliances network.

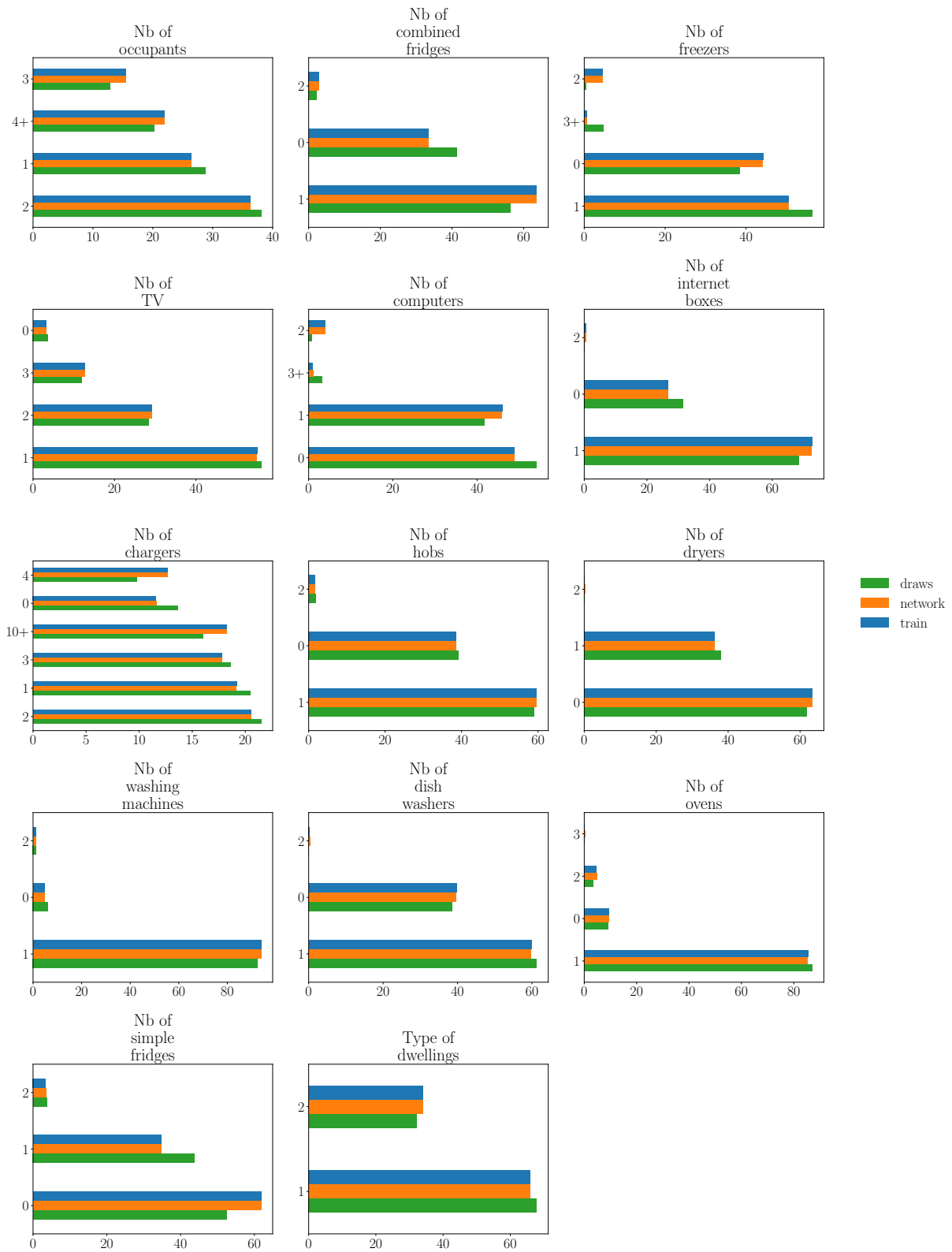


Figure 3: Marginal distributions of the training set, the trained network and of a sample drawn from the appliances network.

Ad hoc validation of the Bayesian network trained on the Fichiers Détail Logement

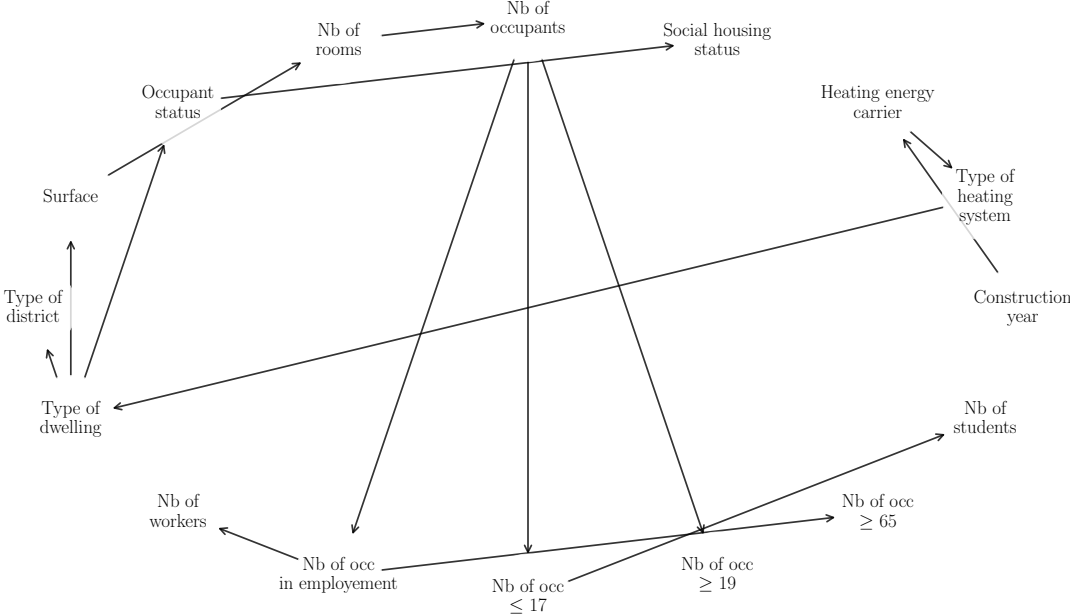


Figure 4: Bayesian network obtained by training on the census.

	Log-loss	RIG
Nb of workers	1.743858	0.02
Nb of occ in employment	1.793517	0.02
Nb of occ ≤ 17 years old	1.387179	0.03
Nb of occ ≤ 19 years old	1.283674	0.07
Nb of occ ≤ 65 years old	1.092516	0.03
Nb of students	1.366321	0.03
Nb of occupants	1.797292	0.07

Table 2: Single feature performances assessment for the census network.

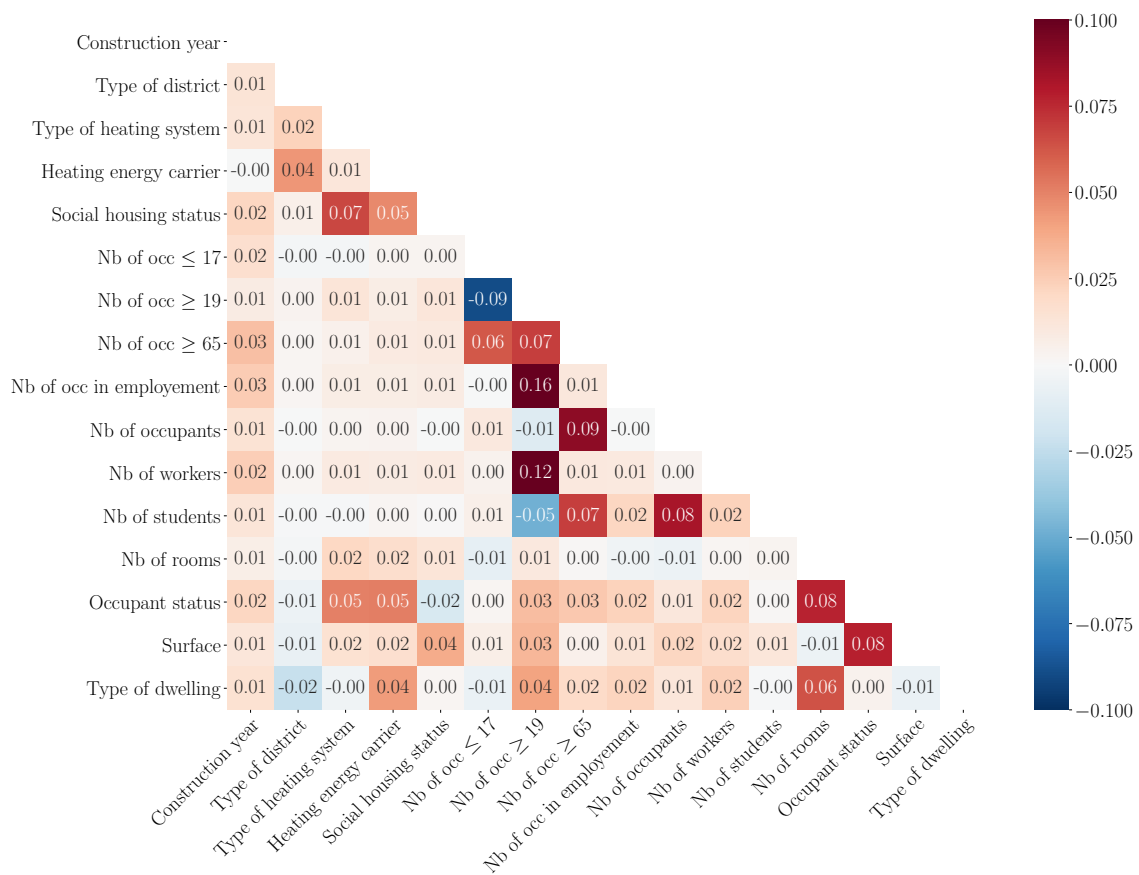


Figure 5: Mutual Information loss obtained for the census network.

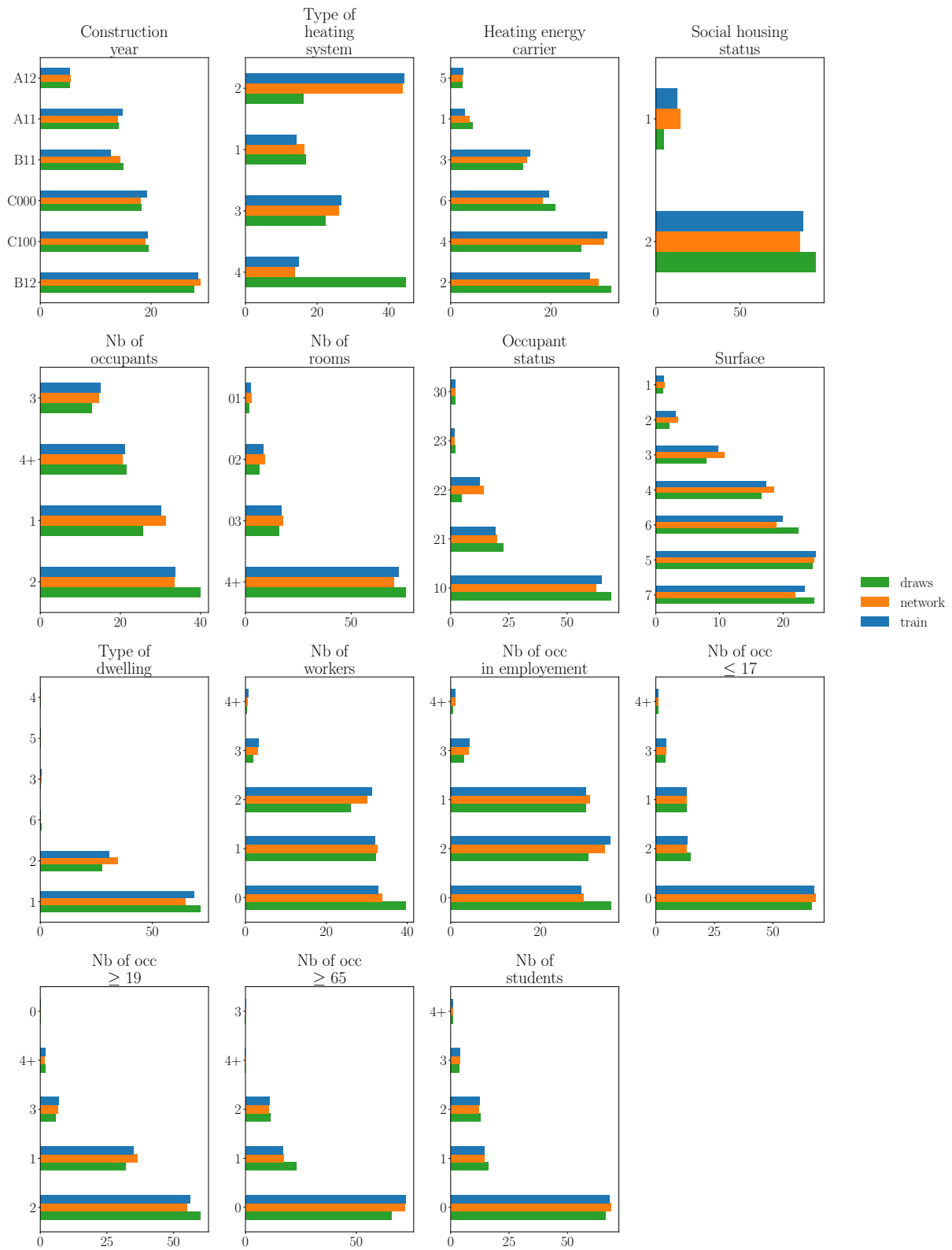


Figure 6: Marginals distributions of the training set, the trained network and of a sample drawn from the census network.

Paper submitted at uSIM2020

Development of a methodology of validation for urban building energy models and application to French residential consumption

Lucie Lefort^{1,2,3}, Romain Bonabe de Rougé¹, Pascal Schetelat², Thomas Berthou³,
Peter Riederer², Bruno Duplessis³, Eric Peirano^{1,3}

¹Efficacity, Champs-sur-Marne, France

²CSTB, Paris, France

³Mines ParisTech, PSL Research University, Paris, France

Abstract

Cities need to drastically reduce their energy consumption in order to fight climate change. To help them implement efficient energy saving measures, urban building energy models (UBEM) have been developed. A large amount of input data is required to model buildings at an urban scale and that information is often missing. To overcome this lack of model input data, data enrichment approaches including archetypes and calibration processes have been used, resulting in errors in the generated data which are not always well quantified. The aim of this paper is to develop a methodology of validation for UBEM that takes into account a data enrichment process. The proposed approach is to split the validation process into several parts based on UBEM use cases. UBEM outputs and enriched data to be validated are selected for each use case and the fitness for purpose of the model is then verified against measured data. This validation method has been tested on a specific use case: the prediction of yearly electricity and gas consumptions of dwellings aggregated at the district scale. It has been applied to two different UBEM tools: Smart-E and Dimosim. Smart-E uses its own hypothesis for data completion while a specific tool, Qiriël, is used to infer missing data for Dimosim. The simulation of one hundred districts (90 000 dwellings) located in the French department of Seine-et-Marne and the comparison to measured data enabled to study the error distribution of both UBEM tools. Further work should be led at smaller spatial and time scales.

Introduction

With an urgent need to act against climate change, more and more cities set ambitious environmental targets. In order to achieve such goals, reducing energy consumption in buildings seems to be a keystone. In France, for example, tertiary and residential sectors represented 45% of the national final energy consumption in 2015 (ADEME, 2015). To assist in lowering these consumptions at the city scale, urban building energy models (UBEM) have been developed over the last fifteen years (Reinhart and Cerezo Davila,

2016). UBEM compute the energy consumption of every building in a district and take into account interactions among them, like, for example, solar shading. The aim of UBEM is to provide cities with guidance to reduce final energy consumption not only for some specific buildings but also at the scale of a whole district.

However, the benefits brought by UBEM are balanced with certain challenges that may jeopardize their reliability.

A first challenge is the data enrichment process. It is much more difficult to gather data required for the simulation of one hundred buildings than for a single one. Information is often missing and assumptions must then be made. This practice of data enrichment creates an error that is not yet well-quantified. A second challenge is to take into account new phenomena that appear at the district scale, like the phenomenon of urban heat islands.

Finally, although a building energy model (BEM) has to be precise for each building it simulates, this is not necessarily the case for a UBEM. It may be acceptable for a UBEM to be statistically correct at the scale of a district but to be rather inaccurate at the scale of a single building. The challenge is then to identify the spatial and time scales on which the UBEM is reliable.

A thorough validation process would enable to verify the reliability of UBEM and would give them more credits as decision-support tools for cities. However, such validation process is not yet implemented within the UBEM research community and is therefore needed.

The aim of this paper is to present a validation methodology designed specifically for UBEM. This methodology is described in the next section before being applied to two UBEM tools in a second section. Results obtained by the two UBEM tools throughout the validation process are then being discussed.

Development of a validation methodology for urban building energy models

Most UBEM stem from building energy models (BEM). However, at the difference with the BEM,

there is no consensus yet on their validation. In this section, we review the existing validation attempts that can be found in the literature and we draw a list of requirements for a validation methodology. Then, we describe a first proposal of validation methodology based on UBE M use cases.

Existing attempts of validation

BEM validation

Building energy model can be seen as a cornerstone of UBE M. Therefore it seems natural to look at the existing validation process implemented for BEM.

A lot of research has been carried on BEM validation among which were the BESTEST and the PASSYS project.

In the BESTEST (Judkoff and Neymark, 2013), several sets of perfectly described buildings are provided and inter-comparison of models are computed on these well-known buildings. In addition, some empirical validation has been proposed with measurements on test cells.

Jensen (1995) gives a precise description of the methodology developed within the PASSYS project. The proposal is to split up the validation process into two main parts: the validation of single processes and the validation of the whole model. Jensen encourages namely to use analytical verification for single process validation and to use empirical validation for whole model validation.

The validation of thermal models used within UBE M through one of these two validation methods would not suffice to consider a UBE M tool as properly validated for several reasons:

- Additional physics and additional energy systems are modelled at the urban scale;
- An enrichment process is often used as a prior to UBE M and needs to be validated as well;
- It may be acceptable for UBE M to be accurate only at certain spatial scales, larger than the building scale.

However, some principles from these methodologies should be used as guidance for the development of a UBE M validation methodology. For example, separating single process validation and the whole model validation as advised by Jensen (1995) is of great interest for UBE M as well. Furthermore, in the same idea as the PASSYS project, the aim of the present work is to develop a validation methodology rather than validating a single tool.

Regarding validation attempts on UBE M, two approaches have been observed so far: comparative validation and empirical validation.

Comparative validation

Inter-model comparison has started within two international research projects, the IEA Annex 60 (Wetter, 2017) and the IBPSA Project 1 (Saelens et al., 2019) which is still ongoing. In both projects, the

taken approach is quite similar to the BESTEST (Judkoff and Neymark, 2013). A synthetic district has been designed to serve as a District Energy Simulation Test (DESTEST). Several UBE M tools are being tested on this DESTEST and their results are compared. This setup enables to control all the simulation parameters and thus proceed with the single process validation as advocated by Jensen (1995). However, since the district used is not real, it is not possible to compare the obtained results with measured data nor to test the enrichment process used by each UBE M tool.

Empirical validation

The aim of empirical validation is to ensure that UBE M tools represent the real-life urban projects accurately. However, such process is made difficult due to a lack of data at district and urban scales. Some examples can be found in the literature like in Sokol et al. (2017) or in Fonseca and Schlueter (2015). These attempts give a first idea of UBE M performances on real projects. However, datasets used in these validation processes are rarely publicly available thus preventing other research teams from going through the same validation process. Furthermore, there is no agreement so far on the metrics to be used to measure the error on UBE M outputs. This prevent the reader from comparing results obtained in different papers.

Nonetheless, some requirements can be drawn from these existing attempts of UBE M validation.

Requirements for a UBE M validation methodology

Some lessons can be learned from this brief literature review and it can be summarized as follows. A validation methodology dedicated to UBE M should comply with the following requirements:

- Specifying the spatial and time scales on which a given UBE M tool is reliable;
- Defining metrics to properly measure the UBE Ms accuracy throughout the validation process;
- Identifying the error generated by the model setting;
- Ensuring applicability and replicability of the methodology in different contexts (country, building sector etc.);
- Proposing test cases to apply the developed validation method.

A first proposal of methodology is presented hereafter based on these requisites.

A validation method based on UBE M's use cases

The methodology developed here offers to inventory UBE Ms use cases and to identify, for each use case, critical UBE M outputs. The aim is to verify that the outputs of interest for each use case are accu-

rate enough to provide a real aid to the decision-maker. Such outputs of interest are called key performance indicators (KPI) in the rest of the paper. Once use cases and KPI have been identified, validation datasets are used in order to study, for each use case, the error distribution on the KPI. This methodology is described in Figure 1.

In an ideal case, several batches of measured data would be available, enabling to obtain an error distribution for each KPI of each use case and UBEM tool. The main idea behind this search for an error distribution is to provide the UBEM user with a confidence interval on the KPI, depending on the use case. UBEM use cases and corresponding KPI for validation were identified through a literature review that is not presented in this paper. Although, the main use cases are listed below:

- Development of renewable energies at the district scale;
- Energy flexibility;
- Performance measurements and verification;
- Buildings retrofitting;
- Study of the urban heat island effects and its impact on indoor comfort;
- Future energy use;
- Urban energy planning.

Future works should properly define metrics for each use case and for each KPI in order to measure the error on the KPI accurately.

A drawback brought by the proposed validation methodology is the need for large validation datasets when such datasets are often hard to find. For example, if the aim of the study is an error distribution at the district scale at least hundred districts with measured data would be needed to obtain this distribution. To tackle this issue, it is proposed to start the validation process with use cases for which additional data are available. Then, we could study in what extend such use cases could give some bounds for the error distribution of use cases with less validation data. For instance, if the annual energy consumption of a building is accurate, it does say something about its annual load profile. Furthermore, more and more energy disclosure laws are being implemented in western countries, giving the opportunity to access large buildings energy consumptions databases. Examples of these energy disclosure laws can be found in France but also in other countries like in the USA (Kontokosta and Tull, 2017).

Application of the proposed methodology to two UBEM tools

Validation framework

A first use case was chosen to apply and test the proposed validation methodology. This use case can be described as territorial energy diagnosis. It consists in being able to give an accurate outlook of a district

energy consumption. In this study, the annual electricity and gas consumptions of residential buildings were studied at the district scale. This use case may be seen as a preliminary study for other use cases since a first step for UBEM is to provide accurate representation of existing consumptions before evaluating different scenarios like the retrofitting of buildings or the development of renewable energies. The choice of this use case was also driven by the publicly available datasets.

This use case was applied on the French department of Seine-et-Marne at the scale of the IRIS. The IRIS is an administrative division of the French territory. Each IRIS gathers around 2000 inhabitants. All residential buildings located in about a hundred of IRIS were simulated over a year and their annual electricity and gas consumptions aggregated at the IRIS scale were compared to measured data. Simulations were completed with two different UBEM tools, Dimosim (Riederer et al., 2015) and Smart-E (Berthou et al., 2015), both relying on automated data enrichment processes, and results were compared. This case study deals only with the residential sector since less databases are available for the simulation of the tertiary sector in France. Datasets used for the simulation and for the validation are presented hereafter.

Simulation datasets

Three databases were used for simulation:

- BD TOPO® (IGN, 2019) This database provides a map of the whole French territory where each building is described by a footprint and a height. Additional information is also given per building such as:
 - Building function (residential, commercial, etc.)
 - Construction year
 - Number of floors
 - Number of dwellings

Though the footprint and the height are given for every building, the additional information is not always available. This additional information is obtained through a joint between tax record files and the map. Information is missing when the joint has not been successful. The BD TOPO® is not publicly available yet but it can be granted for free for any research project. It should become completely public by 2022.

- PHEBUS (Conseil National de l'Information Statistique, 2013): This database is built upon a survey completed in 2013 for 5000 households and for each household, their appliances and their energy bills were reported. Following this first step, energy performance audits were completed in 2000 out of these 5000 households.
- Fichiers Détails logements (FDL) (INSEE, 2016): This database originates from the population census. Each row corresponds to 1 to

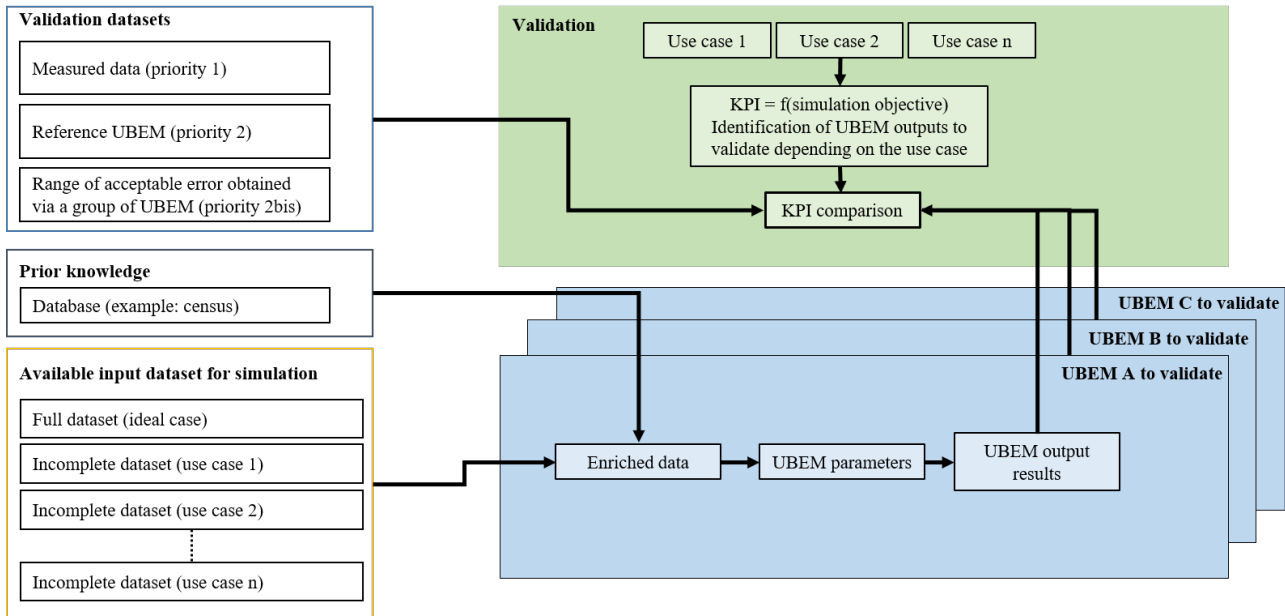


Figure 1: Proposed validation methodology

5 dwellings with identical properties. The precise location of these dwellings is unknown but the zipcode of the IRIS where they are located is provided on each row. Information contained in this database includes:

- Dwelling’s number of occupants;
- Socio-professional status of the household;
- Surface category of the dwelling;
- Type of dwelling (individual housing or collective housing);
- Construction year of the building;
- Main heating energy;
- Main heating system type (centralized or decentralized).

Validation dataset

The validation dataset employed for this use case is the Données Locales de l’Energie database (SDES, 2020). This database originates from a French law for the ecological transition that makes compulsory for energy suppliers to disclose annual energy consumptions data aggregated at the IRIS scale and detailed by sector. Annual electricity and gas consumption of the residential sector are thus available at the IRIS scale. These data are publicly-available and are updated every year. The related disclosure law was recently amended compelling energy suppliers to publish energy consumption data at the building scale, given some privacy threshold for the residential sector. These datasets requires some preprocessing but are of great interest for the validation of UBEM.

Two different UBEMs, one validation process

The proposed validation method aims to be applicable to a large range of UBEM. Two first UBEM tools were used here to test this method: Dimosim

(Riederer et al., 2015) and Smart-E (Berthou et al., 2015). These tools are very different from each other. A first difference comes from the data enrichment process. Smart-E uses internal hypotheses to complete missing data while Dimosim is coupled with a tool designed for the enrichment process: Qiriél (Ansanay-Alex et al., 2016).

A second difference comes from the datasets used as inputs data and as parameters. Dimosim takes as input data the geometry of the buildings retrieved from BD TOPO[®] and then tries to assess thermal properties and household characteristics thanks to PHEBUS and FDL. Conversely, Smart-E takes as input data the description of households provided by FDL that includes the surface of the dwellings but not the geometry of the buildings and then tries to allocate façade areas and roof areas thanks to BD TOPO[®]. BD TOPO[®] and FDL both give information on dwellings but since FDL comes from the census, it is anonymized to prevent identifying the households and the buildings. Therefore, the link between these two databases is not straightforward and assumptions must be made. Furthermore, information provided by these databases is not always consistent. For example, the number of dwellings recorded in BD TOPO[®] and FDL is different.

A third difference is the number of thermal zones defined for the simulation. Several thermal zones per building can be defined in Dimosim but for the sake of simplicity only one thermal zone per building was used in this case. Smart-E uses two thermal zones per dwelling by default, one for the heated space and one for the non heated space.

Booth et al. (2012) identified the most influential parameters on housing stock modelling uncertainties. The mean value of these parameters distributions

used for each UBEM tool in this study are summarized in Table 1. These values are the average on all the simulated dwellings weighted by the dwellings surfaces. These mean values have been displayed to illustrate the difference of parameterization between the two tools. For example, the share of dwellings heated with gas and the share of dwellings heated with electricity vary from an IRIS to another and are often slightly different between Dimosim and Smart-E due to different enrichment processes.

Table 1: Influential parameters on results uncertainty

Parameters	Dimosim	Smart-E
Fraction of space heated	100 %	77 %
Heating temperature set-point (C)	19.6	19.6
Share of dwellings with gas as heating energy	48.3 %	50.6%
Share of dwellings with electricity as heating energy	38.2 %	35.7 %
Efficiency of electric heaters	1.0	1.0
Efficiency of gas boilers	0.78	0.69
Window U-value (W/m ² K)	3.21	2.69
Wall U-value (W/m ² K)	0.96	1.38
Window-to-wall ratio	0.25	0.15
Air-change rate (including infiltrations and ventilation) (vol/h)	0.6	0.52
Number of simulated dwellings	89876	93717
Average surface per dwelling (m ²)	105	92

Results & Discussion

KPI and metrics

The proposed methodology was tested on 93 IRIS (around 90000 dwellings), all located in the French department of Seine-et-Marne. The studied KPI in this case were the annual electricity and gas consumptions at the IRIS scale but also the annual thermosensitivity for gas and electricity at the IRIS scale. The thermosensitivity is defined as the increase of electricity or gas consumption due to the decrease of the outdoor temperature. Its estimation for each tool was done following the methodology provided with the DLE (ENEDIS et al., 2018). Measured data were available in the DLE for each of these KPI and for each IRIS. The error on annual consumption and thermosensitivity was measured through a percentage of error (PE) defined in Equation 1.

$$PE_X = \frac{X_{Simulated} - X_{measured}}{X_{measured}} \cdot 100 \quad (1)$$

The results are presented in Figures 2 and 4.

Analysing the error distribution with respect to input data

The error distribution obtained with Dimosim on annual consumptions is shifted toward the positive values, meaning that Dimosim mainly overestimates the electricity and gas consumptions. Conversely, Smart-E tends to underestimate these consumptions. In absolute values, Dimosim seems to generate a greater error than Smart-E on this particular use case. This performance gap could come from the fact that Smart-E is calibrated in order to fit national electricity consumptions while there is no similar process in Dimosim.

Regarding the error on the thermosensitivity, it seems that Dimosim also overestimates this phenomenon for both gas and electricity. An hypothesis is then that Dimosim overestimates the heating needs in general. When comparing inputs parameters between Smart-E and Dimosim, the fraction of space heated looks like a good candidate to help reducing this error. In order to confirm this hypothesis, simulations should be run again with a different space-heated ratio and results should then be compared.

When taking a closer look at the error generated by Smart-E, the error on the electrical thermosensitivity seems surprisingly high compare to the error on the annual electricity consumption. In order to get a better understanding of all these error distributions and to try to explain phenomenon like the one observed on the thermosensitivity for Smart-E, an analysis of the correlation between the error and a large set of inputs parameters have been completed.

Correlations were estimated through the Pearson correlation coefficient. This coefficient is defined in equation 2 for two variable X and Y, where $cov(X,Y)$ is the covariance of X and Y, σ_X is the standard deviation of X and σ_Y is the standard deviation of Y. A coefficient of 1 indicates a linear relationship between variables while a coefficient close to 0 indicates that there is no linear relationship (but a non-linear relationship is still possible between variables).

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y} \quad (2)$$

Coefficient correlations between the percentage of error on annual gas and electricity consumptions, gas thermosensitivity and electrical thermosensitivity and inputs parameters are presented in Figure 3.

In general, the obtained coefficients are quite low, most of them being below 0.5 in absolute value. It is then difficult to draw influential parameters from this study. However, some trends can be identified. Thus, the two UBEM tools seem to behave quite differently in terms of error. Smart-E presents a coefficient of 0.64 between the percentage of error on the gas consumption and the one on the electricity consumption but such correlation does not appear for Dimosim.

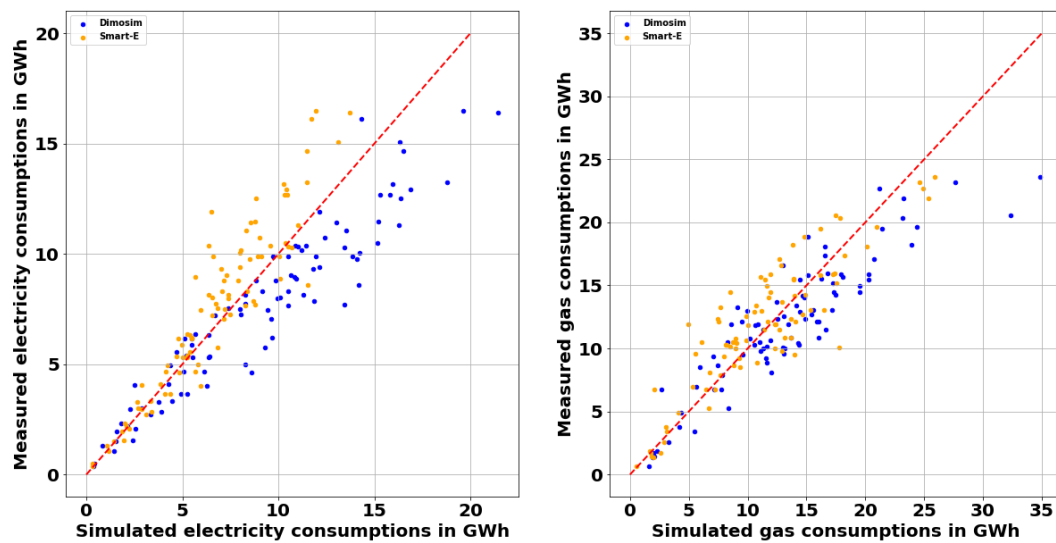


Figure 2: Annual measured and simulated consumptions of gas and electricity

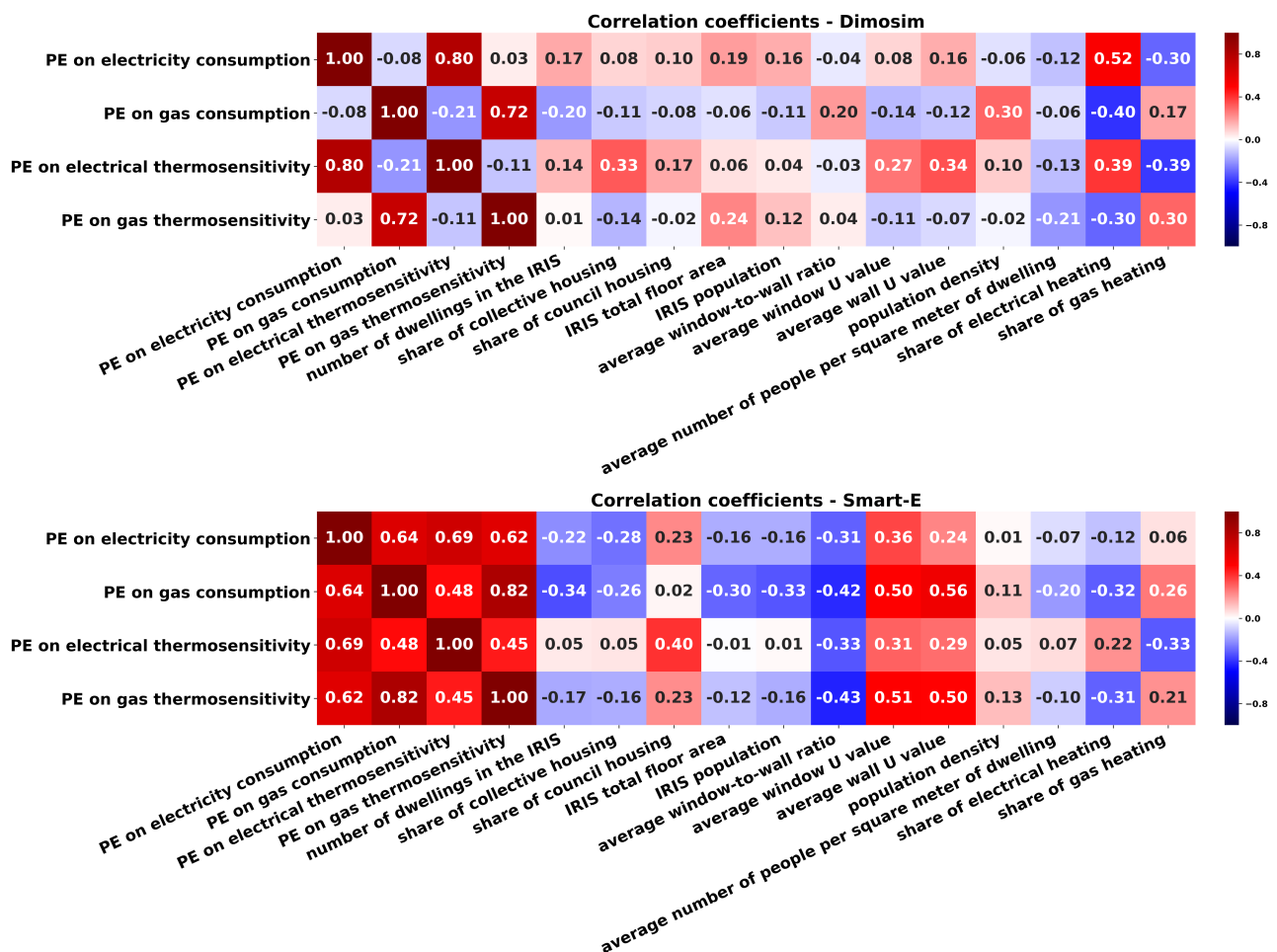


Figure 3: Correlation coefficients between the error and several input parameters

Furthermore, the higher coefficients can be found for Smart-E between the percentages of error (gas and

electricity) and the window and wall U-values. In Dimosim, a relatively high correlation is observed be-

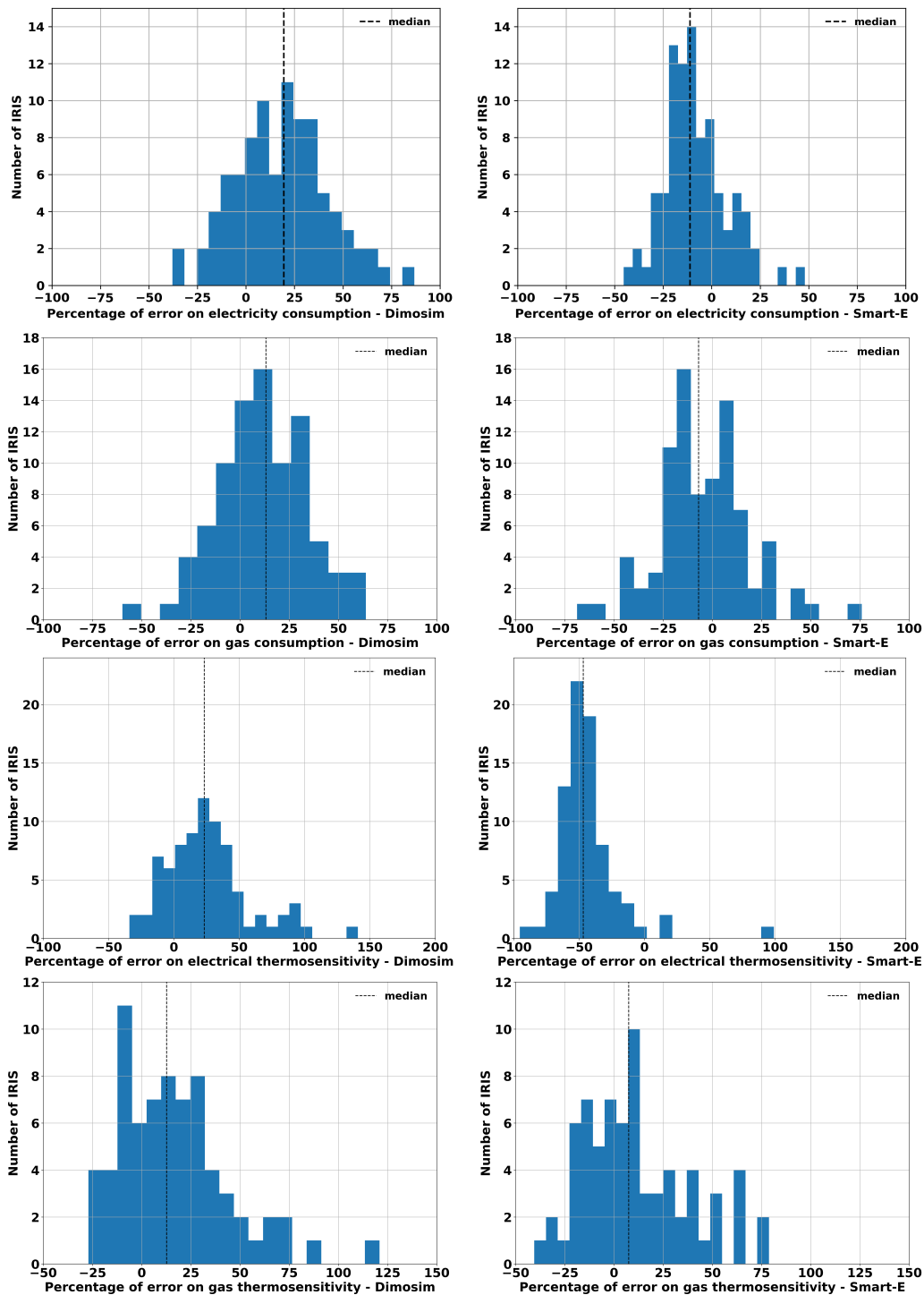


Figure 4: Percentage of error on annual energy consumptions and on thermosensitivity

tween the percentage of error on electricity and the share of dwellings using electricity as heating energy. Finally, a mild correlation between Dimosim and Smart-E errors was identified as well, questioning the quality of databases used for simulation and validation and the error they might bring in. Thus, this analysis enabled to identify levers to reduce the overall error of both tools.

Conclusion

A methodology of validation designed specifically for UBEM was presented in this paper. This method aims to verify that UBEM are accurate enough to provide a real aid in the decision-making process. To do so, the validation procedure is broken down based on UBEM use cases and error distributions are studied for KPI of each use cases.

This method was tested with two first UBEM tools on a French test case. The analysis of the obtained error distributions enabled to identify weaknesses of both UBEM tools and hence improve their performances. The development of this methodology is still ongoing and future work should be carried on UBEM outputs at smaller spatial and time scales. The error analysis and the identification of critical inputs parameters in the output error generation should also be strengthened. Lastly, an important step would be to verify the applicability of the methods to a large range of UBEM tools. Since the available datasets for simulation influence greatly the enrichment process and UBEM behaviour in general, it would be of great interest to apply this methodology in other countries than France and with other UBEM tools as well. The authors would then be very interested in collaborating with other research teams on this topic.

References

- ADEME (2015). Chiffres-clés Climat, Air et Énergie. Technical report.
- Ansanay-Alex, G., Y. Abdelouadoud, and P. Schetelat (2016). Statistical and Stochastic Modelling of French Households and their Energy Consuming Activities. In *12th REHVA World Congress (CLIMA)*.
- Berthou, T., B. Duplessis, P. Rivière, P. Stabat, D. Casetta, and D. Marchio (2015). SMART-E: A TOOL FOR ENERGY DEMAND SIMULATION AND OPTIMIZATION AT THE CITY SCALE. In *Proceedings of BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India, Dec. 7-9, 2015*, pp. 1782–1789.
- Booth, A. T., R. Choudhary, and D. J. Spiegelhalter (2012). Handling uncertainty in housing stock models. *Building and Environment* 48(1), 35–47.
- Conseil National de l'Information Statistique (2013). Enquête Performance de l'Habitat, Équipements, Besoins et Usages de l'énergie (Phébus).
- ENEDIS, GRDF, and Agence ORE (2018). Estimation des données de thermosensibilité et de part thermosensible. Technical report.
- Fonseca, J. A. and A. Schlueter (2015). Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. *Applied Energy*.
- IGN (2019). BD TOPO Version 30. - Descriptif de contenu.
- INSEE (2016). Logements ordinaires en 2016 - Recensement de la population - Fichiers détail.
- Jensen, S. Ø. (1995). Validation of building energy simulation programs: a methodology. *Energy and Buildings* 22(2), 133–144.
- Judkoff, R. and J. Neymark (2013). Twenty years on!: Updating the IEA bestest building thermal fabric test cases for ASHRAE standard 140. *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association* (January 2013), 63–70.
- Kontokosta, C. E. and C. Tull (2017). A data-driven predictive model of city-scale energy use in buildings. *Applied Energy* 197, 303–317.
- Reinhart, C. F. and C. Cerezo Davila (2016). Urban building energy modeling - A review of a nascent field. *Building and Environment*.
- Riederer, P., V. Partenay, N. Perez, C. Nocito, R. Triggance, and T. Guiot (2015). Development of a simulation platform for the evaluation of district energy system performances. In *Proceedings of BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India, Dec. 7-9, 2015*, pp. 2499–2506.
- Saelens, D., I. De Jaeger, F. Bünning, M. Mans, A. Vandermeulen, B. Van Der Heijde, E. Garreau, A. Maccarini, Ø. Rønneseth, I. Sartori, and L. Helsen (2019). Towards a DESTEST: a District Energy Simulation Test Developed in IBPSA Project 1. In *Proceedings of BS2019: 16th Conference of International Building Performance Simulation Association, Rome, Italy, Sept. 2-4, 2019*.
- SDES (2020). Données locales de consommation d'énergie.
- Sokol, J., C. Cerezo Davila, and C. F. Reinhart (2017). Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. *Energy and Buildings*.
- Wetter, M. (2017). New Generation Computational Tools for Building & Community Energy Systems Annex 60 Final Report. Technical report, International Energy Agency.

Appendices to chapter 5

Results obtained with Calomel for three different levels of data availability

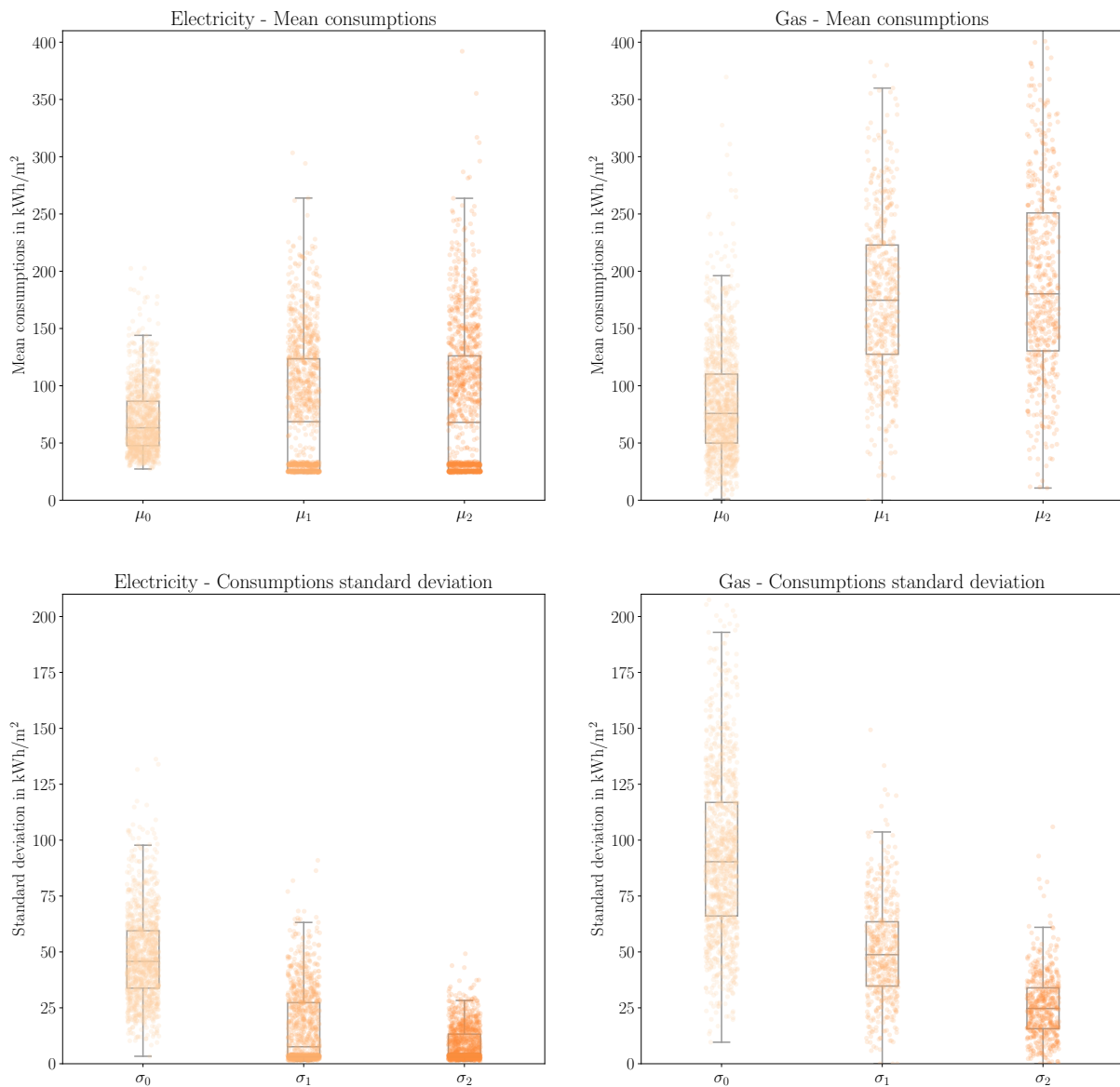


Figure 7: Evolution of the mean and standard deviation of the consumptions obtained with Calomel for each building between the three levels of data availability

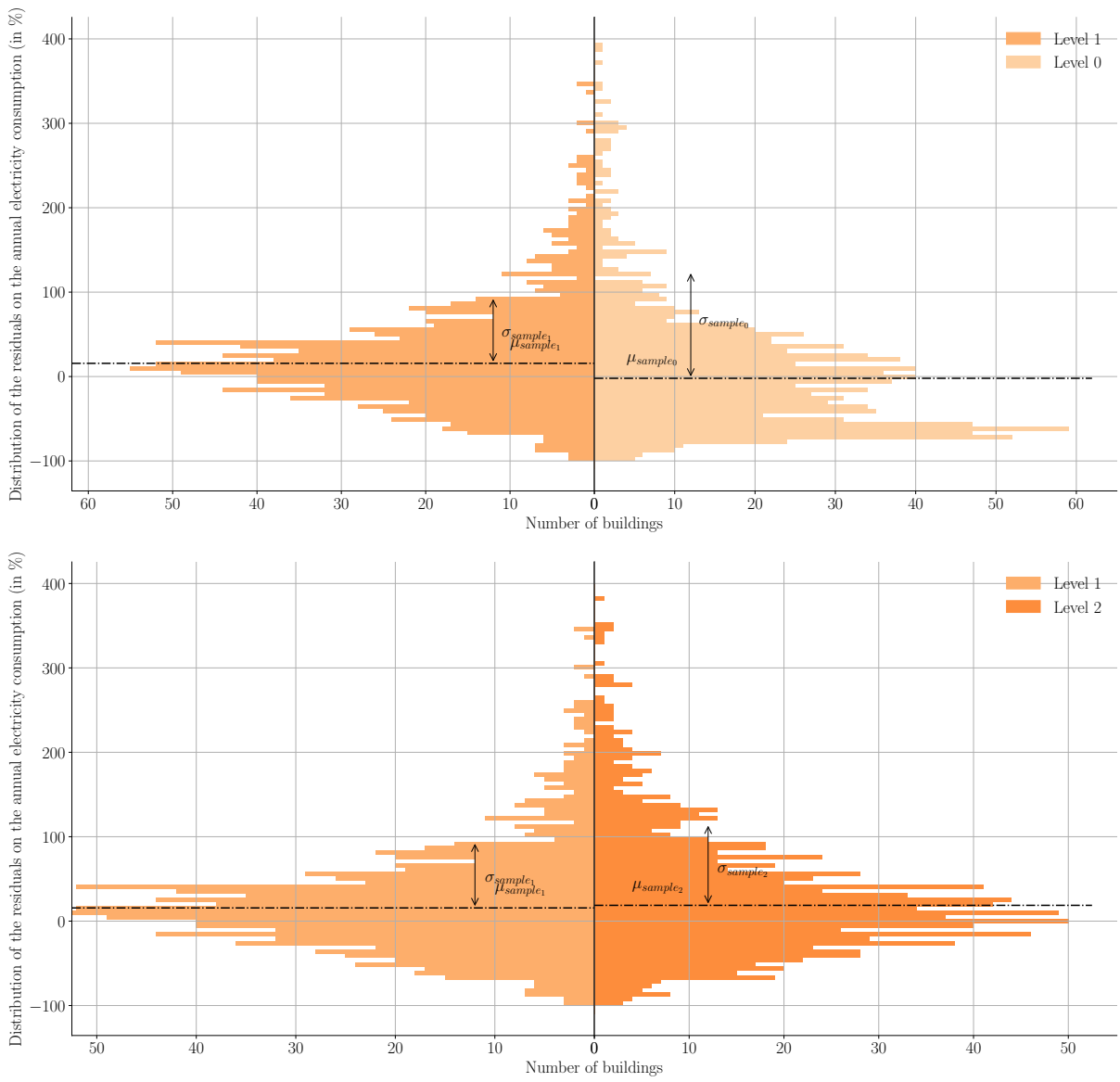


Figure 8: Distributions of residuals obtained with Calomel for the three levels of data availability on electricity

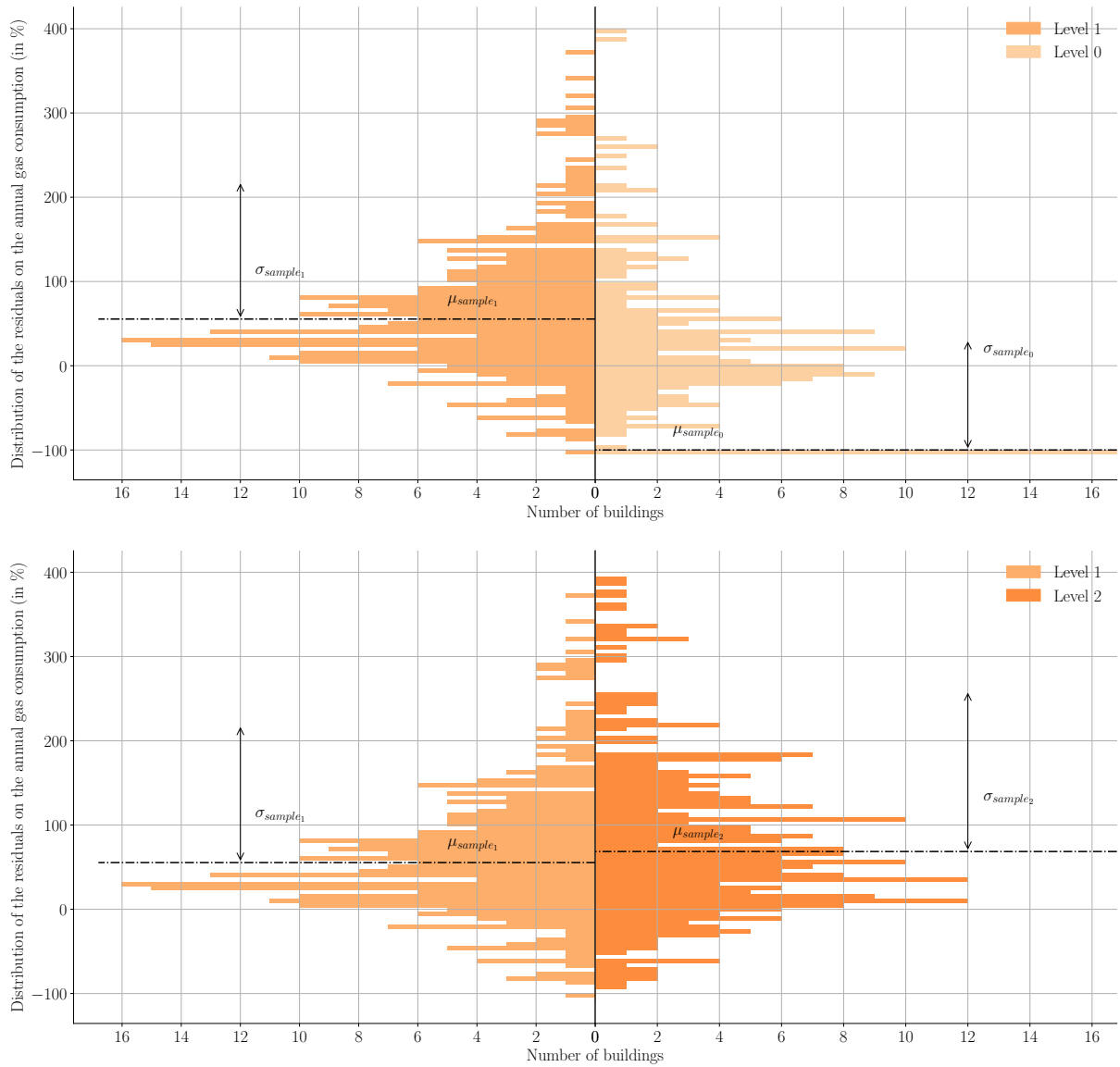


Figure 9: Distributions of residuals obtained with Calomel for the three levels of data availability on gas

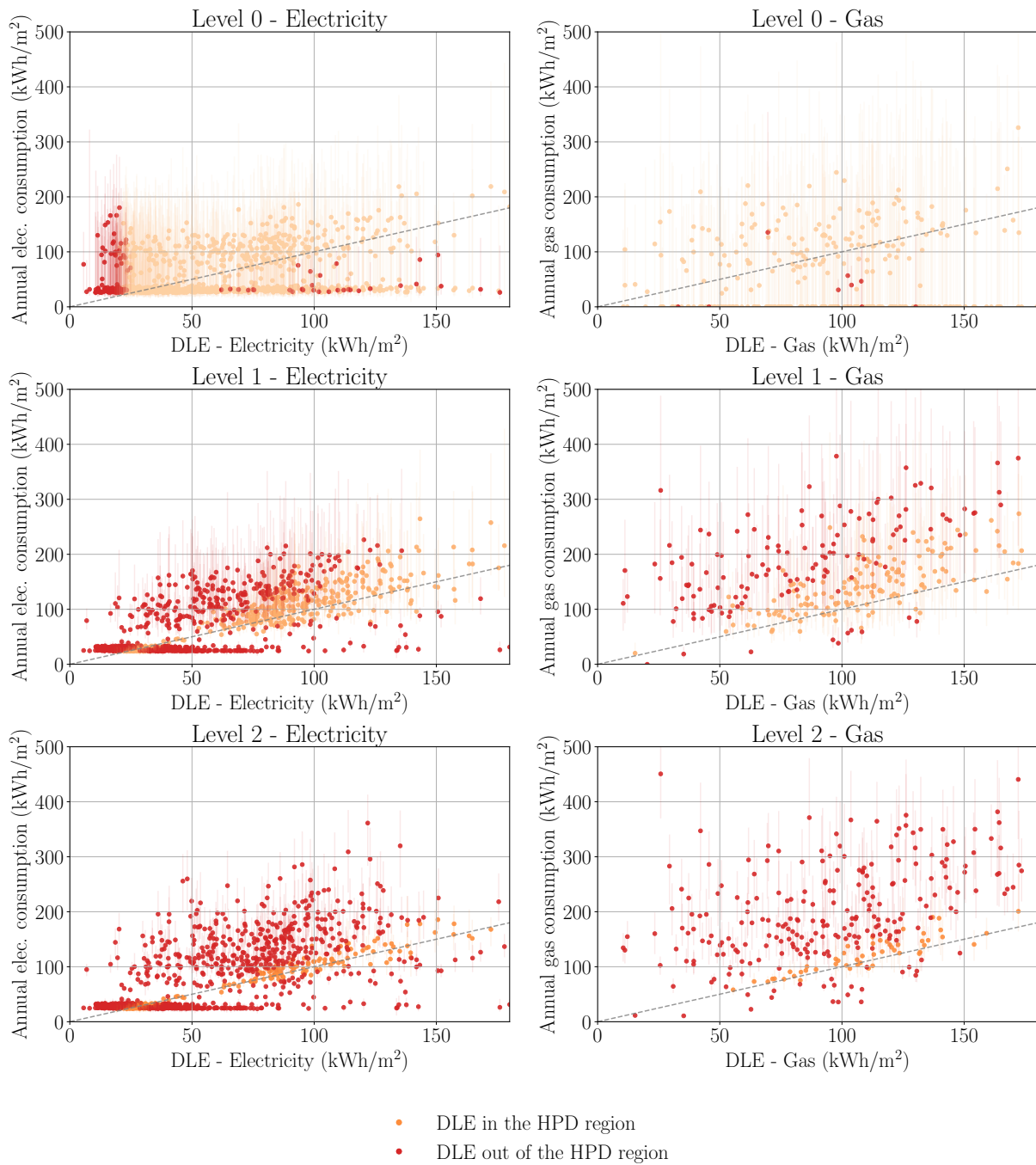


Figure 10: μ_{bldg} and the HPD region of each building against its DLE for three levels of data availability for Calomel

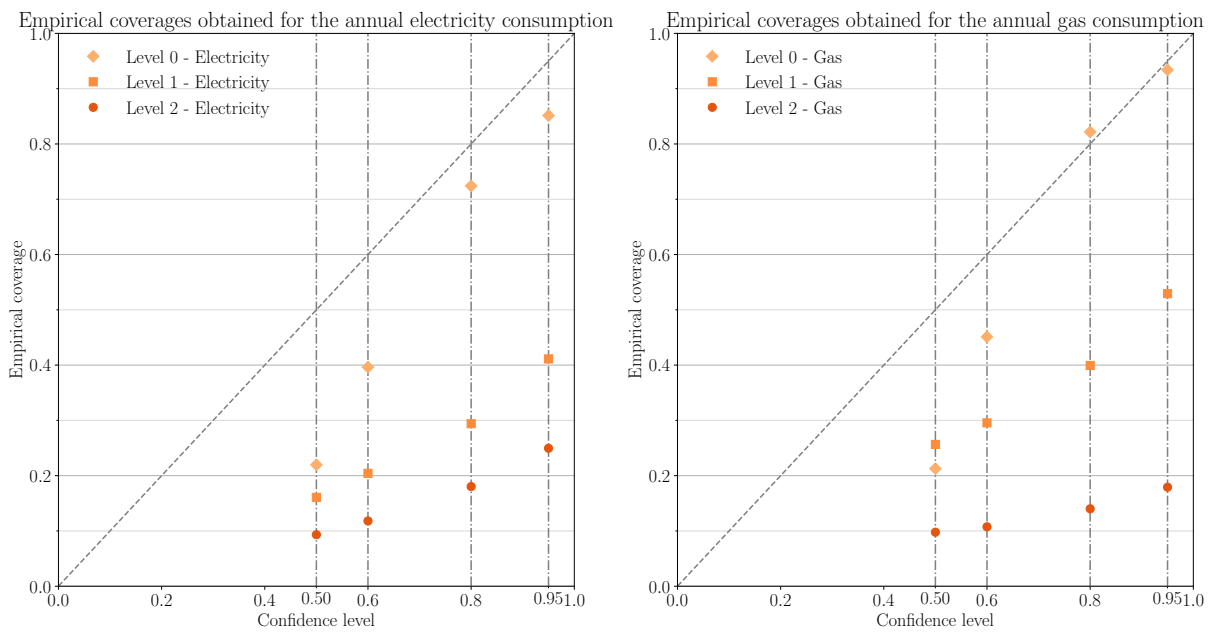


Figure 11: Evolution of the empirical coverage depending on the data availability for Calomel

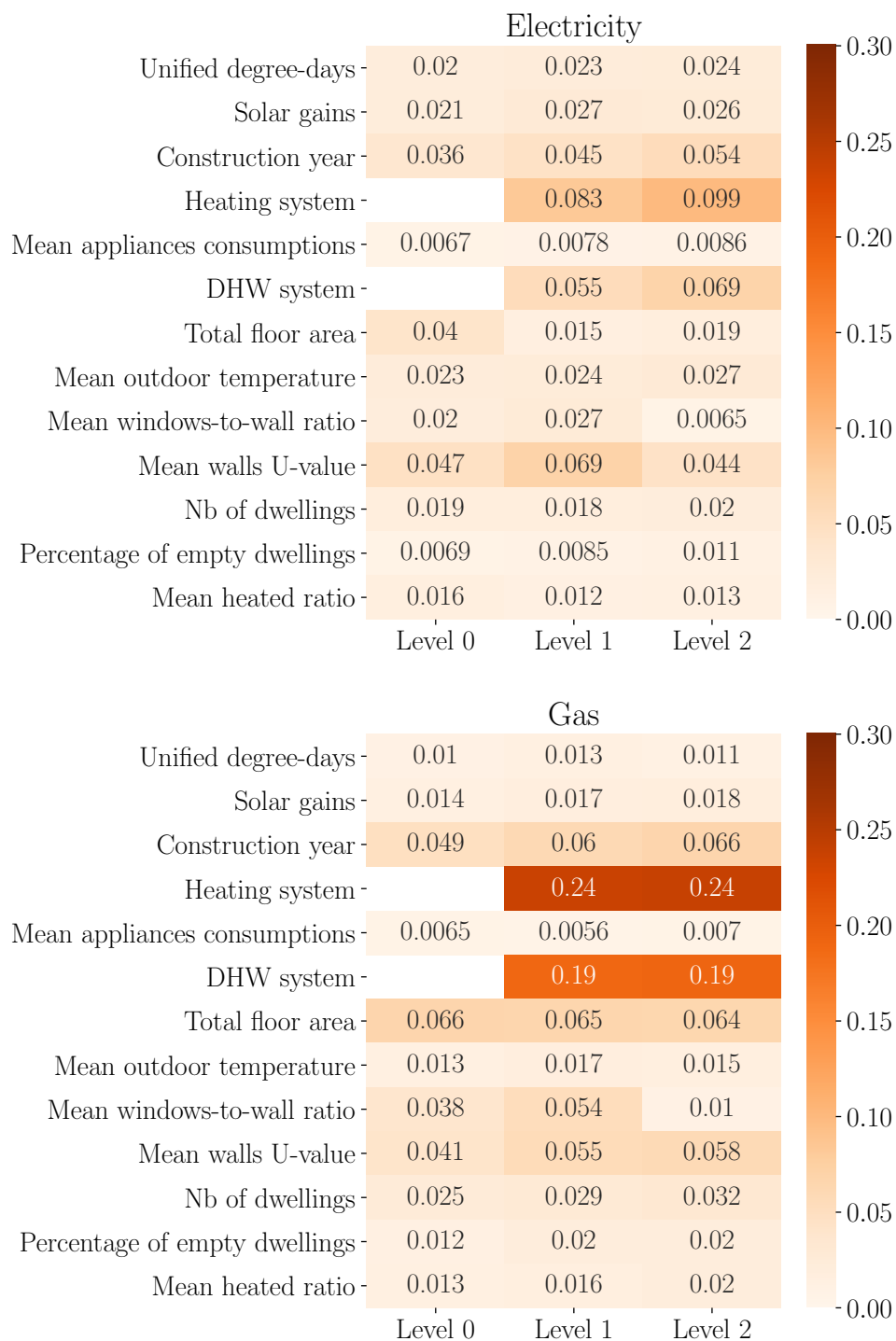


Figure 12: Mutual information between the residual and building's characteristics for three levels of data availability for Calomel

Results obtained with Dimosim for three different levels of data availability

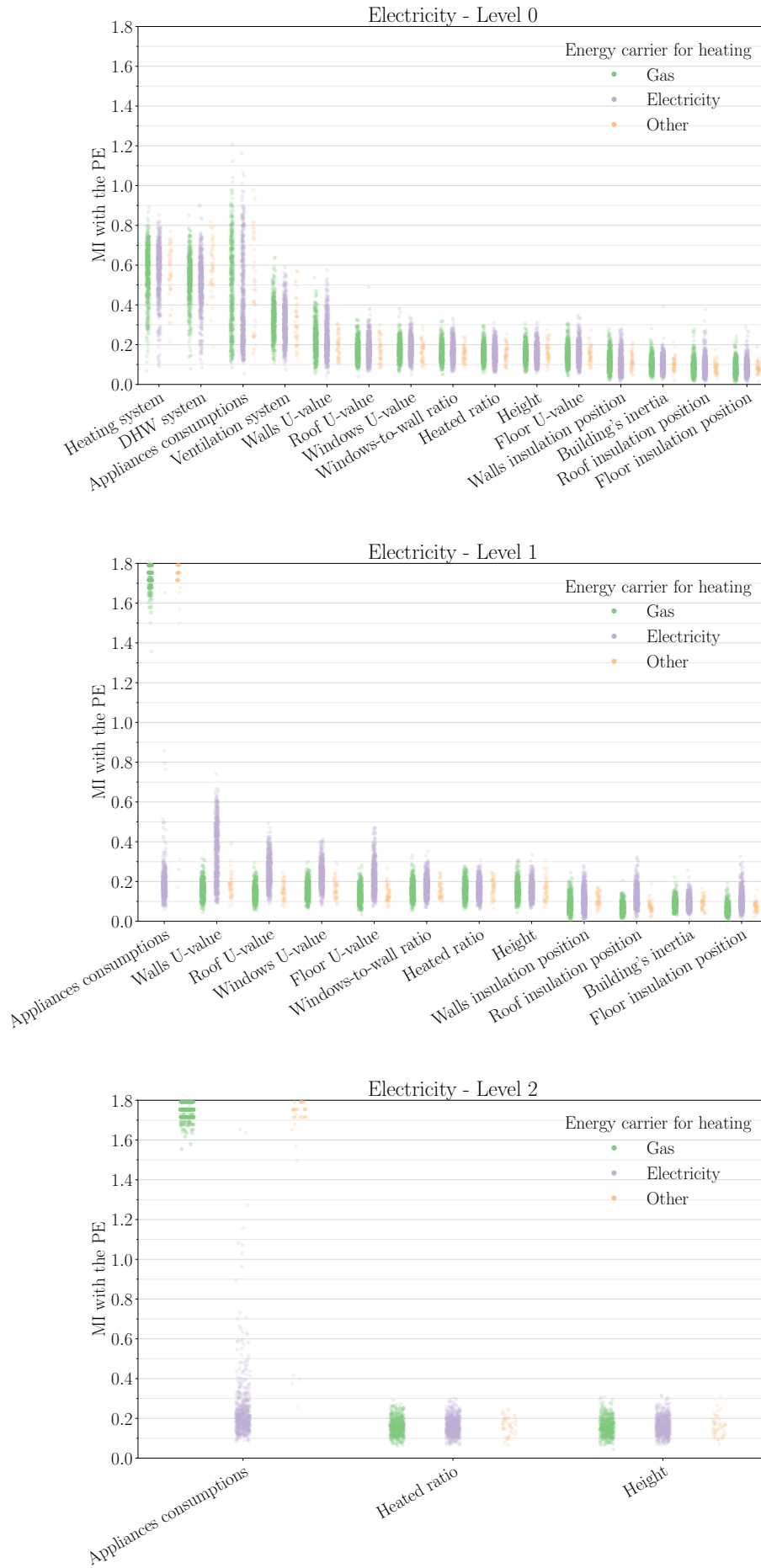


Figure 13: Vectors of mutual information between uncertain parameters and percentage of error on electricity consumption obtained for the buildings of the validation dataset and for three levels of data availability with Calomel

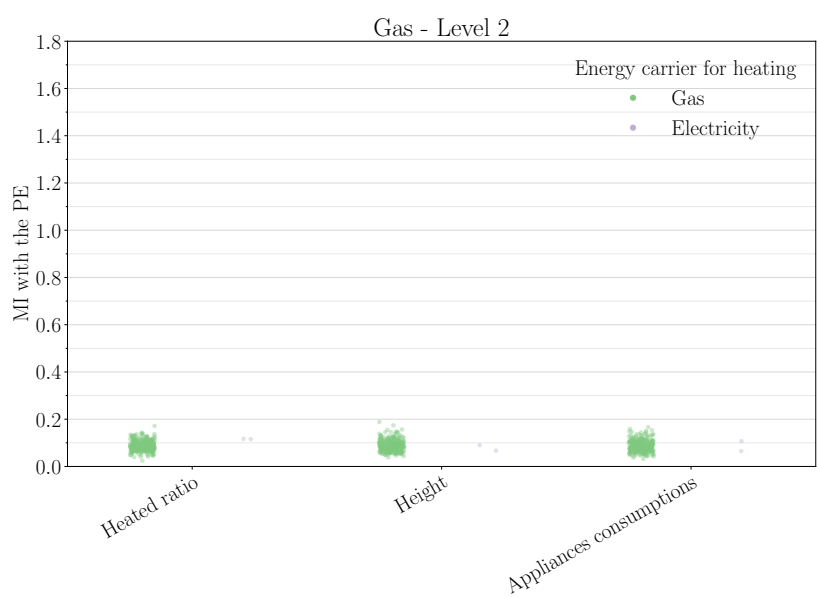
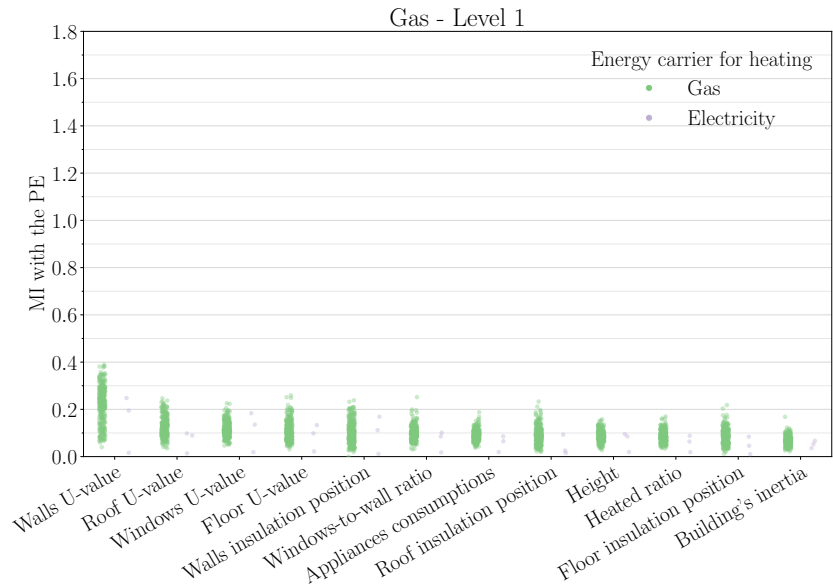
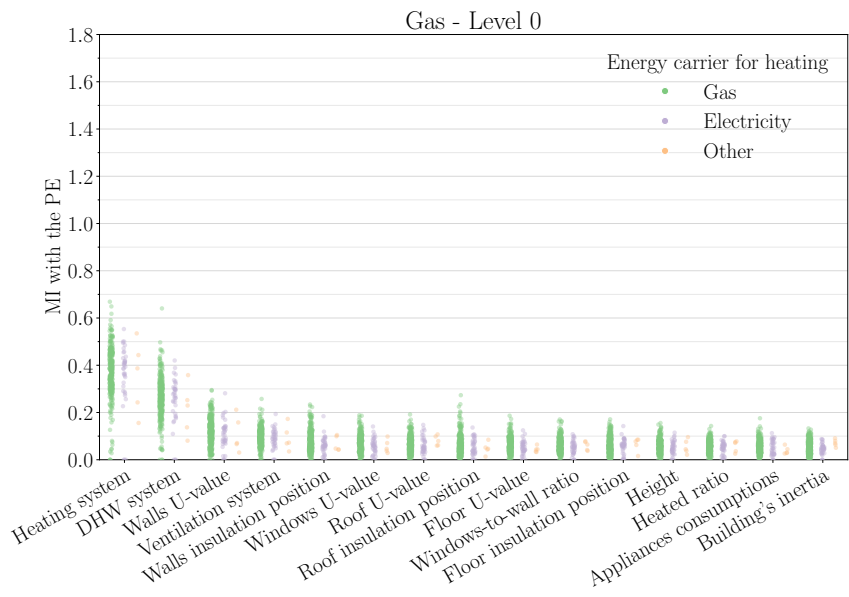


Figure 14: Vectors of mutual information between uncertain parameters and percentage of error on gas consumption obtained for the buildings of the validation dataset and for three levels of data availability with Calomel

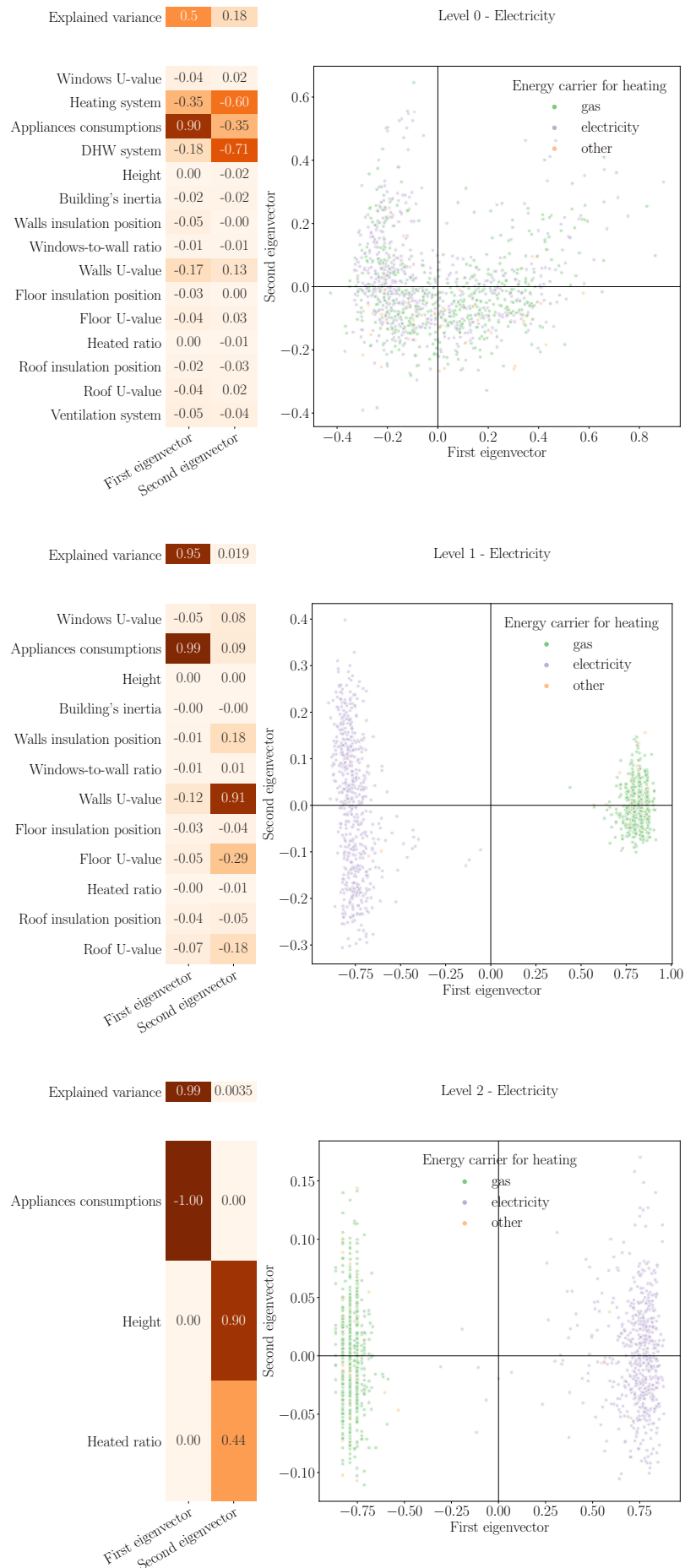


Figure 15: Results of principal components analysis applied to vectors of mutual information for the error on electricity consumption for three levels of data availability obtained for Calomel

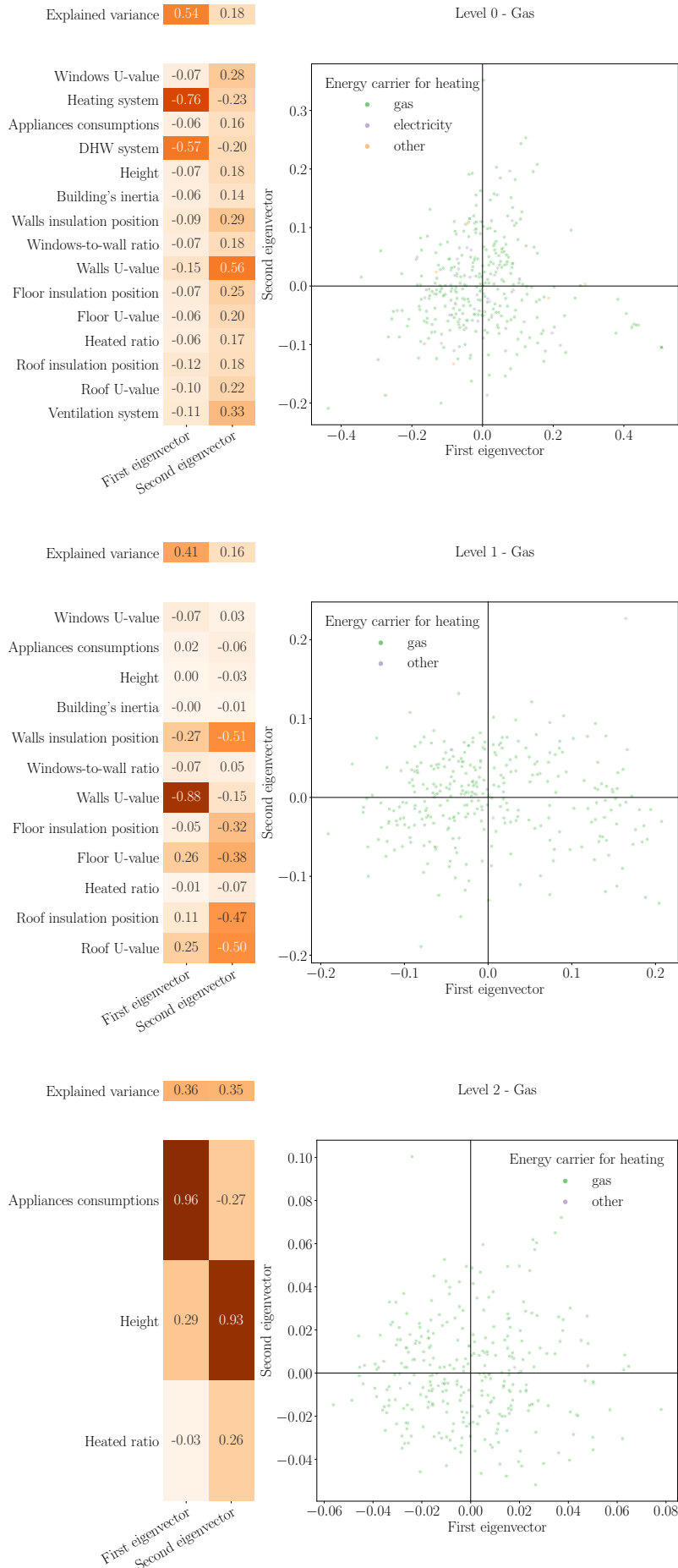


Figure 16: Results of principal components analysis applied to vectors of mutual information for the error on gas consumption for three levels of data availability obtained for Calomel

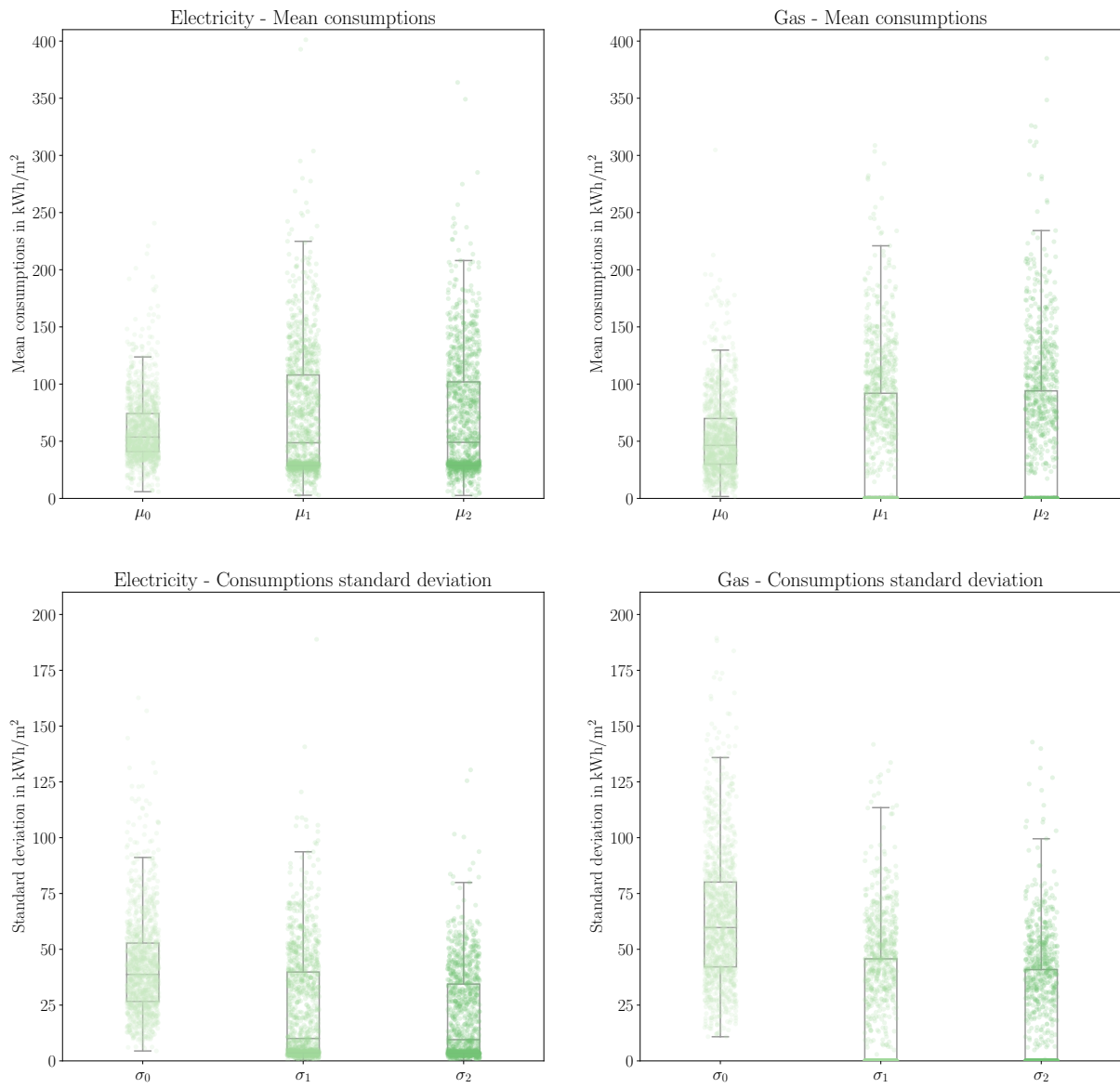


Figure 17: Evolution of the mean and standard deviation of the consumptions obtained with Dimosim for each building between the three levels of data availability

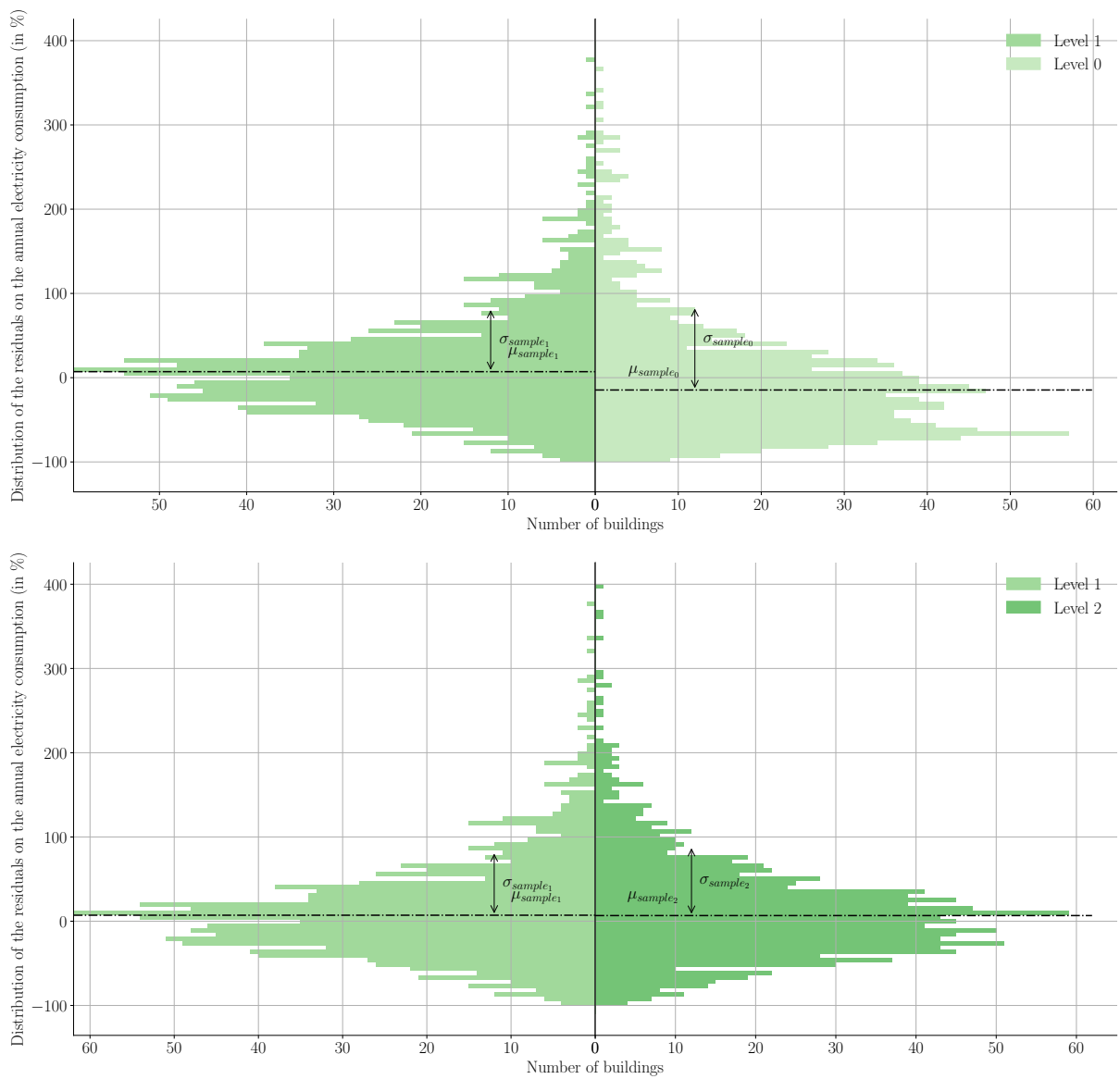


Figure 18: Distributions of residuals obtained with Dimosim for the three levels of data availability on electricity

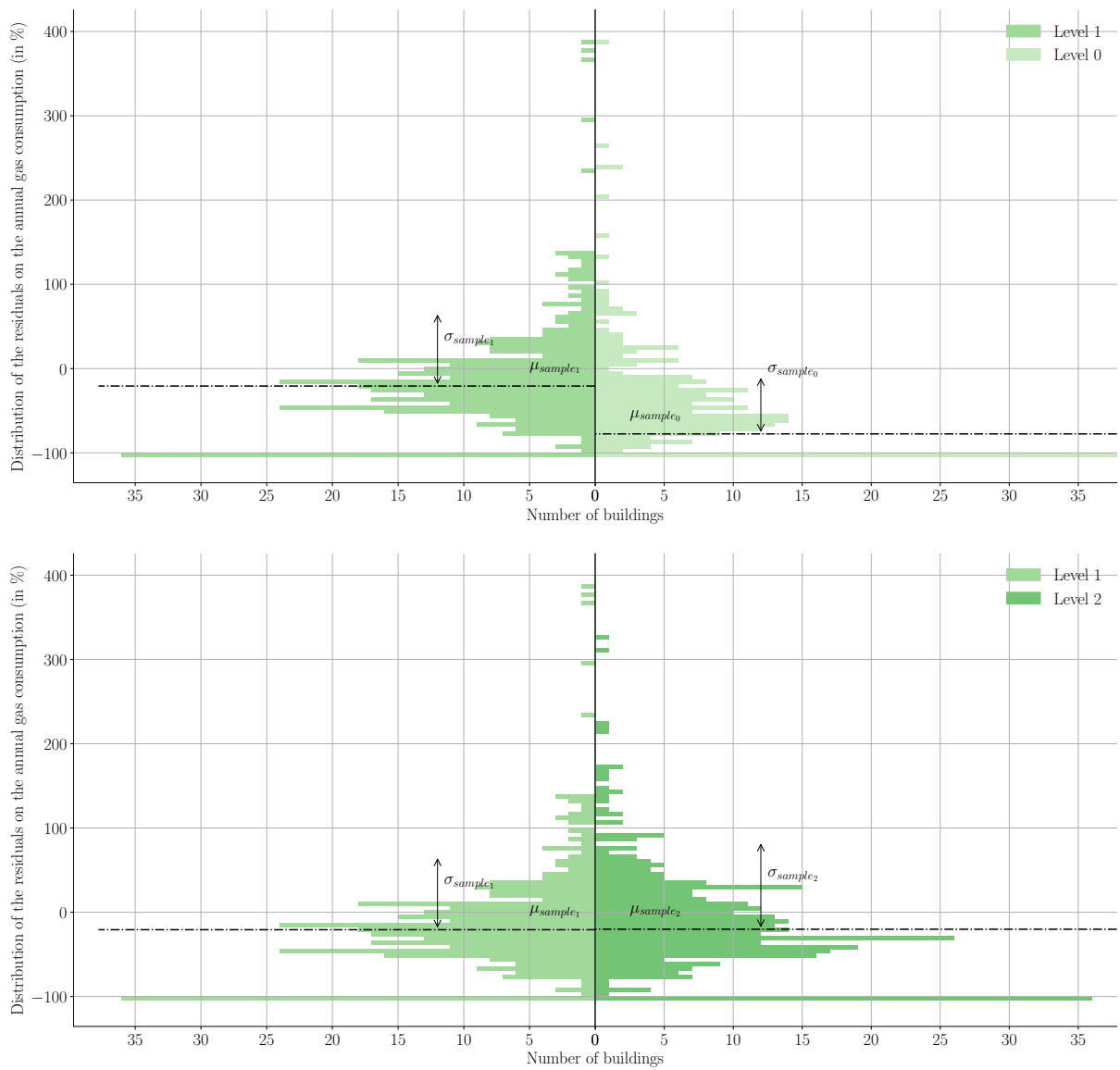


Figure 19: Distributions of residuals obtained with Dimosim for the three levels of data availability on gas

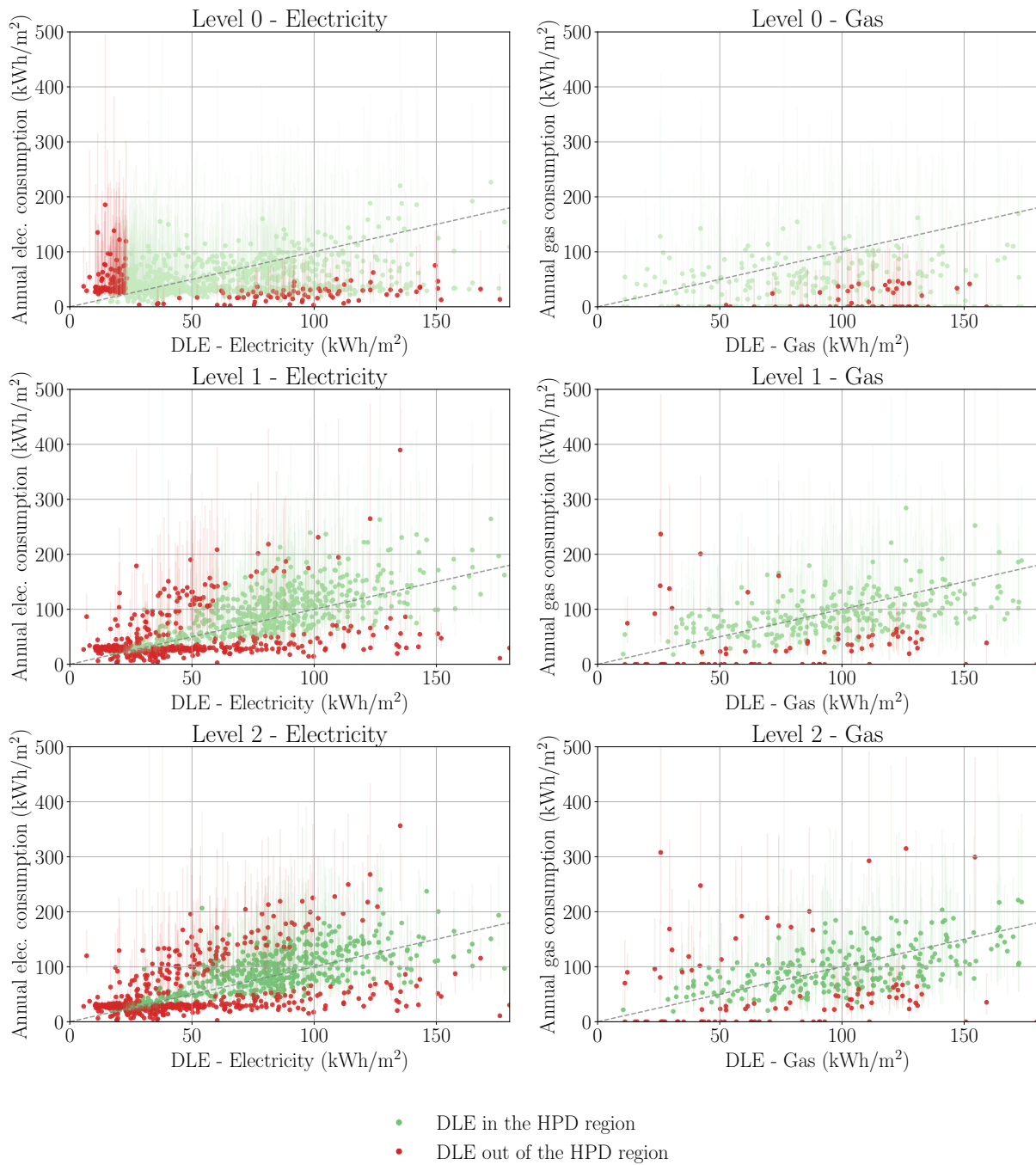


Figure 20: μ_{bldg} and the HPD region of each building against its DLE for three levels of data availability for Dimosim

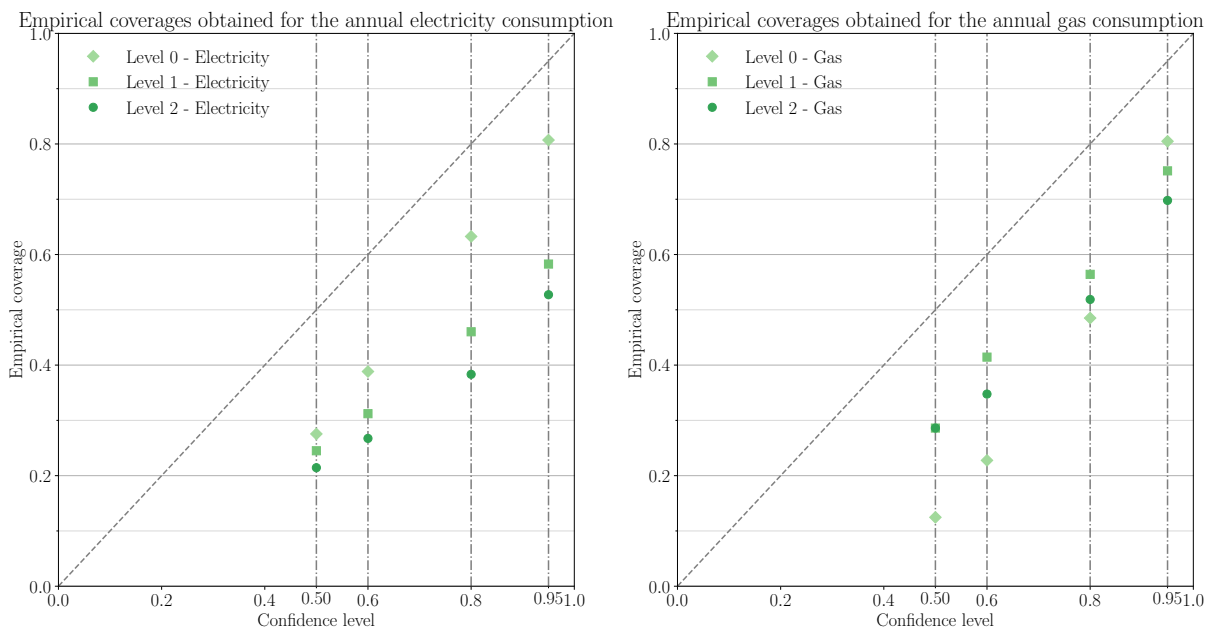


Figure 21: Evolution of the empirical coverage depending on the data availability for Dimosim

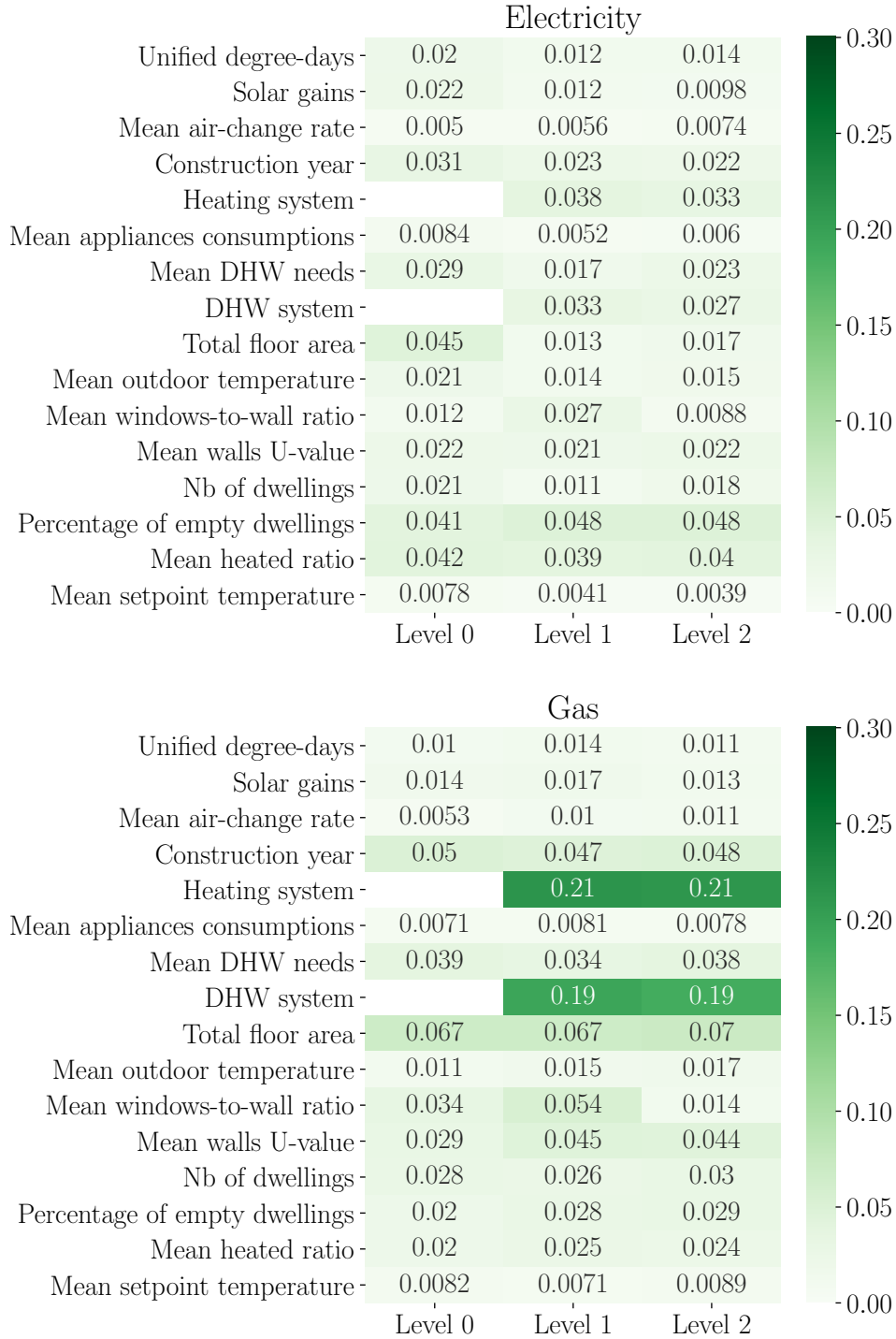


Figure 22: Mutual information between the residual and building's characteristics for three levels of data availability for Dimosim

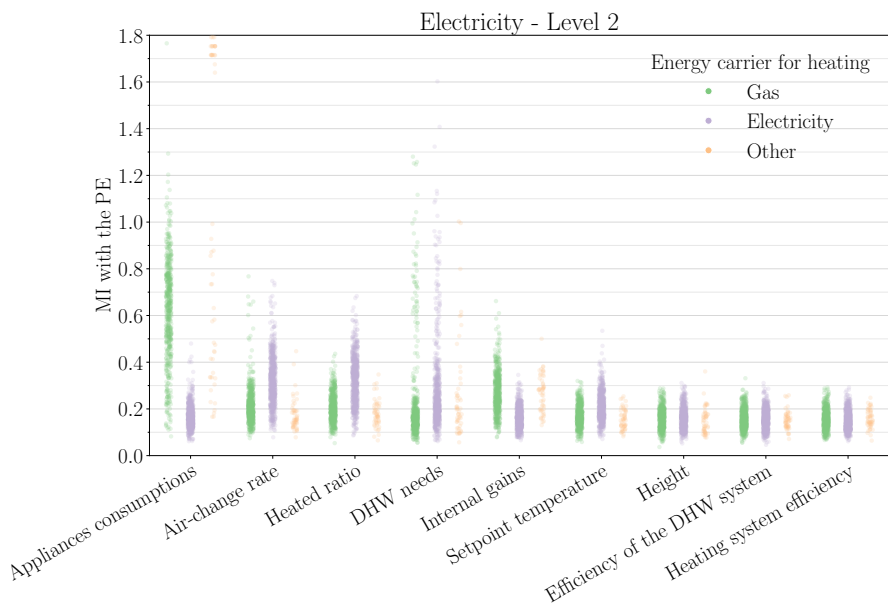
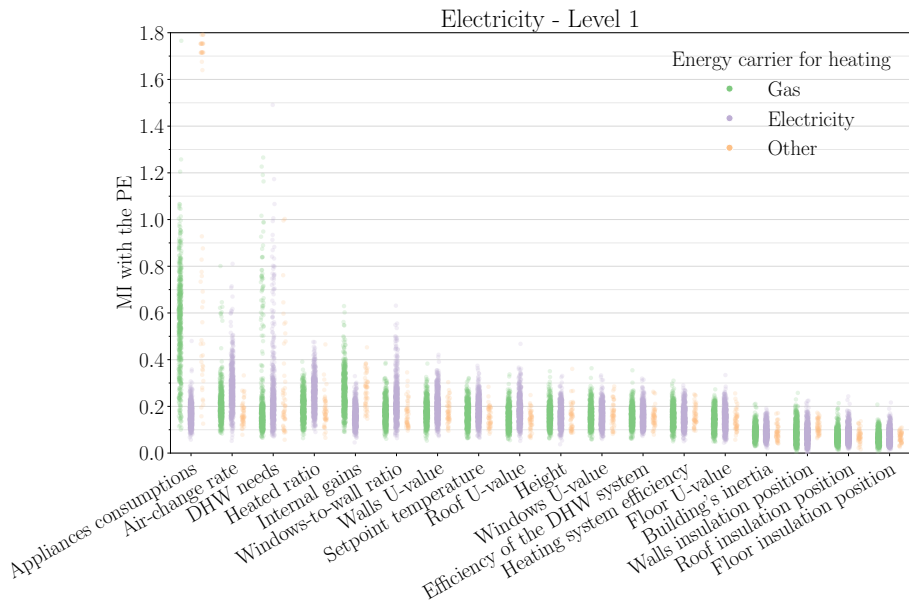
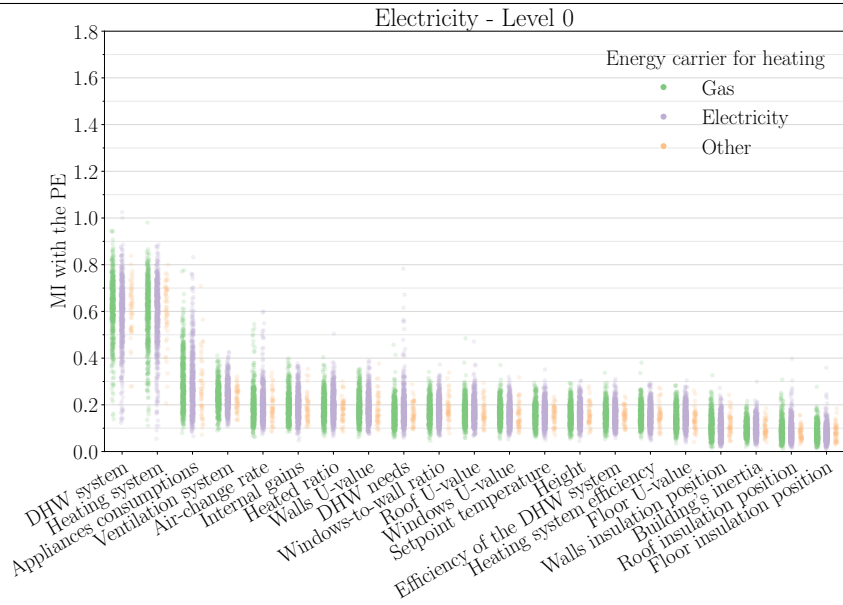


Figure 23: Vectors of mutual information between uncertain parameters and percentage of error on electricity consumption obtained for the buildings of the validation dataset and for three levels of data availability with Dimosim

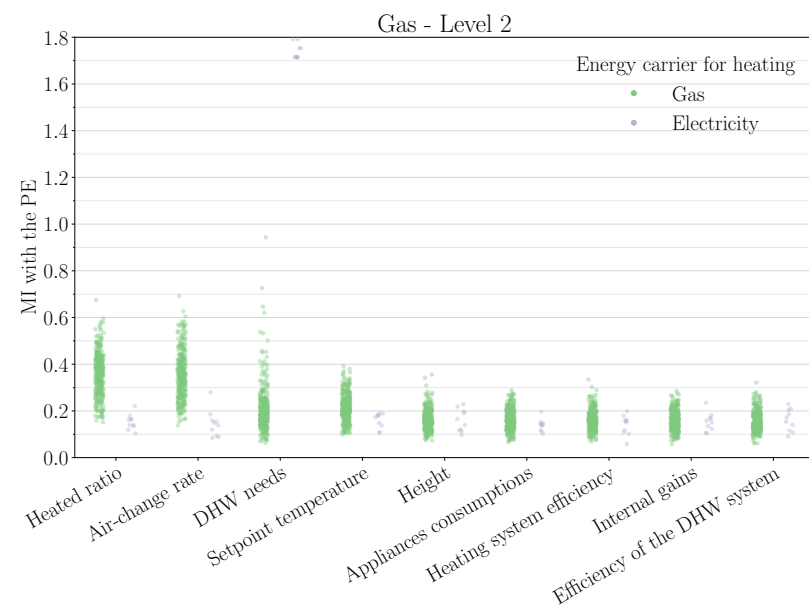
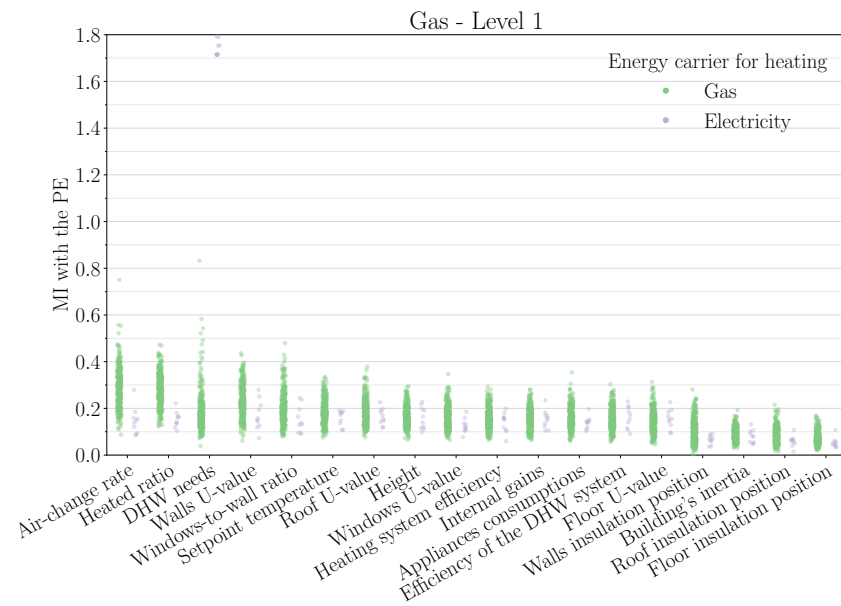
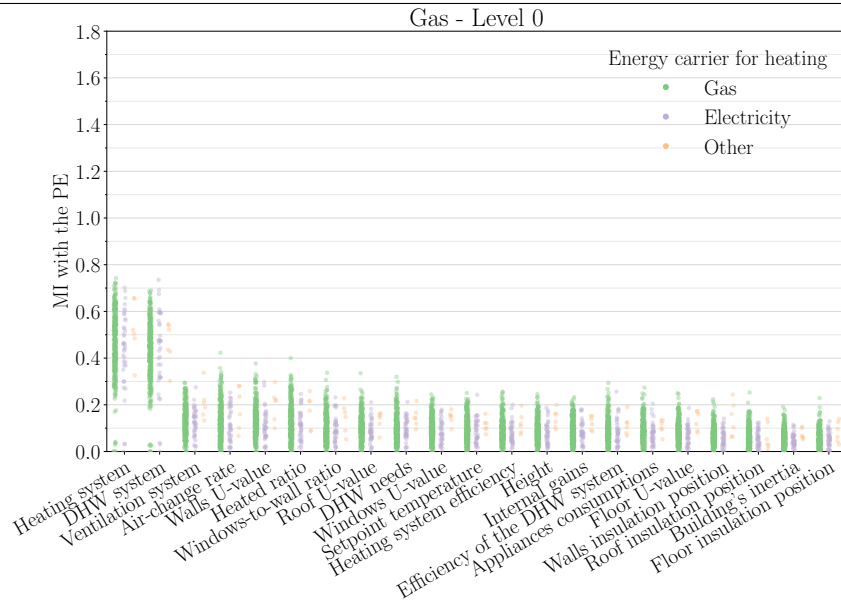


Figure 24: Vectors of mutual information between uncertain parameters and percentage of error on gas consumption obtained for the buildings of the validation dataset and for three levels of data availability with Dimosim



Figure 25: Results of principal components analysis applied to vectors of mutual information for the error on electricity consumption for three levels of data availability obtained for Dimosim

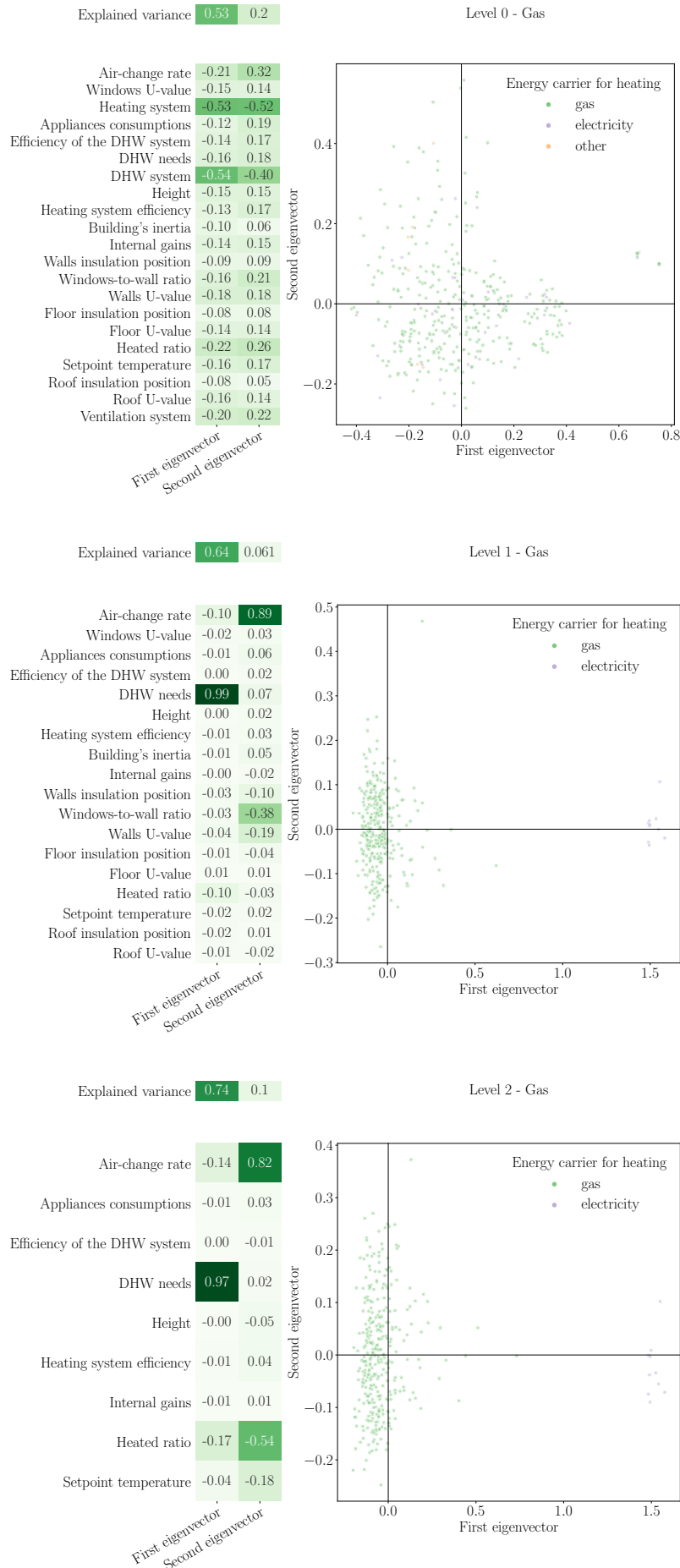


Figure 26: Results of principal components analysis applied to vectors of mutual information for the error on gas consumption for three levels of data availability obtained for Dimosim

RÉSUMÉ

Afin d'aider les décideurs à choisir les meilleures mesures d'économie d'énergie pour les bâtiments à l'échelle urbaine, la communauté scientifique a développé des outils d'aide à la décision parmi lesquels figurent les outils de simulation énergétique urbaine (UBEM). Un UBEM est capable de calculer les consommations d'énergie pour chaque bâtiment d'un quartier ou d'une ville et permet ainsi aux décideurs d'avoir une compréhension claire des impacts de différentes mesures d'économie d'énergie telles que la rénovation de certains bâtiments ou la mise en place d'actions d'effacement des consommations.

Cependant, pour s'assurer que les UBEM fournissent une aide réelle au processus de prise de décision, leur exactitude doit être vérifiée par un processus de validation approfondi. La principale question de recherche abordée dans cette thèse peut donc être décrite comme suit : Comment vérifier l'adéquation des UBEM au processus de décision pour lequel ils sont utilisés ?

Un processus de validation composé d'une étape d'évaluation et d'une étape de diagnostic a été développé et est présenté dans cette thèse. Une application de ce processus de validation est proposée sur des cas d'usages pour lesquels des données accessibles librement peuvent être utilisées pour composer un jeu de données de validation. L'étape d'évaluation et l'étape de diagnostic sont ensuite décrites en profondeur pour cet ensemble de cas d'usages de même que les outils statistiques sélectionnés pour évaluer et analyser les performances des UBEM.

Le cadre d'évaluation proposé a été pensé pour être applicable à un large éventail d'UBEM et à un large éventail de cas d'usages, mais de futurs travaux devraient permettre de vérifier son applicabilité à une plus large échelle.

MOTS CLÉS

Simulation énergétique urbaine, validation

ABSTRACT

In order to help decision makers choose the best energy saving measures for buildings at the urban scale, the scientific community developed decision-aiding tools among which lie the urban building energy models (UBEM). A UBEM is able to compute energy consumptions for each building of a district or a city and thus enables decision-makers to have a clear understanding of the impacts of different energy saving measures such as buildings refurbishment or flexibility.

However, to ensure that UBEM provide a real aid in the decision-making process, their accuracy should be verified through a thorough validation process. The main research question tackled in this thesis can then be described as follows: How to verify the adequacy of UBEM to the decision process they are used for?

A validation framework composed of an assessment stage and a diagnostic stage has been developed and is presented in this dissertation. An application of this assessment framework is proposed on use cases for which open data can be used to compose a validation dataset. The assessment stage and the diagnostic stage are then described in depth for this set of use cases along with the statistical tools selected to assess and analyse the UBEM performances.

The proposed assessment framework has been thought to be applicable to a large range of UBEM and to a large range of use cases, but future works should verify its applicability on a broader range.

KEYWORDS

Urban-scale Building Energy Modeling, Validation