



**HAL**  
open science

# Advancing Urban Building Energy Model Validation: A Comprehensive Multi-Period Approach for Dynamic Outputs

Chunxiao Wang, Thomas Berthou, Bruno Duplessis, Eric Peirano, Pascal Schetelat, Peter Riederer

► **To cite this version:**

Chunxiao Wang, Thomas Berthou, Bruno Duplessis, Eric Peirano, Pascal Schetelat, et al.. Advancing Urban Building Energy Model Validation: A Comprehensive Multi-Period Approach for Dynamic Outputs. Conférence IBPSA France 2024, IBPSA France, May 2024, La Rochelle - Ile d'Oléron, France. hal-04725841

**HAL Id: hal-04725841**

<https://minesparis-psl.hal.science/hal-04725841v1>

Submitted on 12 Oct 2024

**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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Advancing Urban Building Energy Model Validation: A Comprehensive Multi-Period Approach for Dynamic Outputs

Chunxiao WANG<sup>1,2,3</sup>, Thomas BERTHOUS<sup>3</sup>, Bruno DUPLESSIS<sup>3</sup>, Eric PEIRANO<sup>1</sup>, Pascal SCHETELAT<sup>2</sup>, Peter RIEDERER<sup>2</sup>

<sup>1</sup> Efficacity, France

<sup>2</sup> CSTB, Direction Energie et Environnement, France

<sup>3</sup> CEEP, Mines Paris, Université PSL, France

---

## RESUME.

*Cet article présente une nouvelle méthodologie pour valider les sorties dynamiques, telles que les courbes de charge, des modèles de simulation énergétique urbaine (MSEU) à travers une approche de validation multi-période. Cette approche consiste à évaluer les modèles sur différentes périodes afin d'assurer leur précision et fiabilité dans diverses conditions. Appliquée à Dimosim, un MSEU co-développé par le CSTB et Efficacity, via une étude de cas sur le réseau de chauffage de Blagnac en France, la méthodologie identifie efficacement les incertitudes au sein des périodes de validation et la validité relative du modèle sur différentes périodes, améliorant ainsi la fiabilité et l'interprétabilité de la validation des MSEU. Les perspectives de la méthodologie portent sur son application sur des contextes urbains différents et l'intégration de la calibration du modèle à la méthodologie actuelle.*

*MOTS-CLÉS : simulation énergétique urbaine, méthodologie de validation, sorties dynamiques.*

---

## ABSTRACT.

*This article presents a new methodology for the validation of the dynamic outputs, such as load profiles, of urban building energy models (UBEM) through a multi-period approach. This approach consists of evaluating models across different periods to ensure their accuracy and reliability under various conditions. Applied to Dimosim, a UBEM co-developed by the CSTB and Efficacity, through a case study of the Blagnac heating network in France, the methodology effectively identifies uncertainties within validation periods and the relative model validity over different periods, enhancing reliability and interpretability of UBEM validation. Perspectives for the methodology include its application to diverse urban contexts and the integration of model calibration into the current methodology.*

*KEYWORDS: urban building energy modelling, validation methodology, dynamic outputs.*

---

## 1. INTRODUCTION

Rapid urbanization significantly impacts energy consumption and sustainability, with urban areas, home to half of the global population and expected to rise to 58%, contributing to 36% of global final energy consumption and 40% of CO<sub>2</sub> emissions. This situation underscores the need for Urban Building Energy Models (UBEM) as tools for urban planners and policymakers to make informed decisions. Despite the potential of UBEMs, their effectiveness hinges on their accuracy and predictive capabilities, underscoring the importance of their validation. While validations of UBEM, are attempted in the literature (Lefort 2022; Oraiopoulos and Howard 2022), validations of dynamic

outputs are often limited by data availability and a traditional focus on singular time series (Hedegaard et al. 2019).

This approach, critiqued for its sensitivity to specific events (Pack 1990), contrasts with the forecasting field's recommendation for varied validation periods, like the rolling-origin approach, for validation reliability (Tashman 2000). Yet, UBEM studies primarily concern uncertainties in input parameters and data enrichment process (Prataviera et al. 2022; Jaeger 2021), neglecting validation period uncertainties. This study introduces a new UBEM validation methodology addressing these uncertainties, outlined in section 2 through methodology (2.1) and a case study (2.2), and evaluating the methodology's value via case study results in section 3.

## 2. METHODOLOGY

### 2.1. METHODOLOGY PRINCIPLE

The proposed methodology, consisting of four stages, is based on the principle of multi-period validation, an approach that has been applied in time series forecasting (Pack 1990). Multi-period validation creates a sample of validation periods within an individual time series via a rolling-origin approach, to remove ‘the possibility that the arbitrary choice of time origin might unduly affect the [forecasting accuracy]’ (Fildes 1992). This methodology section provides a guide for conducting multi-period validation of UBEMs.

#### 2.1.1. Data pre-processing

This stage aims to prepare a reliable validation dataset. We must verify following aspects of collected data:

- Availability: At what temporal and spatial scales is the data available?
- Internal Consistency: Are measurements consistent within themselves?
- External Consistency: Are measurements consistent with other reference figures, such as heating degree day, and the modelling objectives?

The more detailed conditions for working with validation data at this stage have been discussed by Oberkampf, Roy and Jensen (Oberkampf and Roy 2013; Jensen 1995).

#### 2.1.2. Creation of validation period samples

In a multi-period approach, we can increase the effective size of validation period sample in the following three dimensions (Pack 1990):

- The number of time series studied, for example, more measurement sites;
- The duration of validation period, which can be determined by the uses of model;
- The origins of validation period, where a rolling-origin approach can be applied.

When possible, we should add an additional dimension, spatial scale which indicates the aggregation level of buildings, which is crucial in uses of UBEM.

#### 2.1.3. Classification of validation periods

Once the sample of validation periods is created, it is possible to classify periods to study the model's validity under different conditions. The size of each group of validation periods must remain large enough to produce interpretable distributions of validation results.

#### 2.1.4. Calculation of validation metrics in samples

In this step, one must select validation metrics that fit with the potential uses of the model. Normalized metrics should be preferred in this step as we may perform comparison of validation results across different measurement sites.

Then, errors should be calculated for all validation periods in each sample. Validation results are thus distributions of validation metrics, which allow interpreting the uncertainty related to the choice of the validation period.

## 2.2. CASE STUDY

A case study was chosen to evaluate the proposed methodology on the Dimosim tool (Garreau et al. 2021). The site is a district in Blagnac, France, where selected buildings are connected to a fourth-generation district heating network (Figure 1). The network features 4 kilometres of tubes, 36 substations and a nominal thermal power of 14 MW. Annually, the system delivers 19 GWh of heat for the space heating and the domestic hot water use of 72 buildings (2000 equivalent dwellings), while 55 of them are residential.



Figure 1: Layout of Blagnac heating network

Measurements from heat meters installed at substations: heating power (MW), flow rate ( $\text{m}^3/\text{h}$ ), supply and return temperature ( $^{\circ}\text{C}$ ), with an hourly time step, are used in this case study. The objective of this case study is to evaluate the model adequacy while simulating dynamic thermal behaviours of collective residential buildings.

### 2.2.1. Data pre-processing

In this phase, we initially identify substations with available dynamic data, then ensuring both internal and external data consistency. Thermal load profiles, in  $\text{W}/\text{m}^2$ , are derived from either direct measurement via heat meters or recalculated using flow rate and supply and return temperatures. Internal consistency checks involve comparing recalculated energy data with direct measurements from heat meters. Subsequently, profiles based on flow rates and temperature measurements are preferred for their superior availability and accuracy. External consistency is confirmed by correlating daily energy consumption with heating degree days calculated from local weather data using the COSTIC method (COSTIC 2021).

### 2.2.2. District modelling and inputs

The space heating demand of selected buildings is simulated by Dimosim using a single-zone model for each building. The simulation duration is from 01/01/2021 to 31/12/2022 at a time step of 1 hour. No calibration is performed.

Building geometrical data is provided by BDNB (CSTB 2024), and then corrected manually by researchers after visual verifications. Regarding building envelop, parameters are provided either by building-level Energy Performance Certificates (EPC) from BDNB or provided by an internal multi-source expert-curated database, which assigns typical values for thermal properties of a building according to building usage type and construction period. The heating temperature set point is chosen

as 19 °C constantly for all buildings and the internal gains are set as constant ratios (in W/m<sup>2</sup>) depending on building usage type. Weather data used in the simulation is collected in a weather station at the airport of Toulouse, right next to Blagnac.

For domestic hot water (DHW) usage, we implemented precalculated stochastic DHW demand profiles in modelled buildings. These profiles are precalculated using Qiriél and Croniq for building profile enrichment and DHW demand simulation. (Ansanay-Alex et al. 2016).

### 2.2.3. Validation of period samples

In the current case study, we are capable to increase our sample size through all three dimensions mentioned in 2.1.2 while adding an additional spatial dimension. Table 1 resumes all four dimensions in our global sample. In the spatial dimension, we worked both at substation and aggregated level. For the three dimensions mentioned in 2.1.2, we took the following approach:

- Number of time series investigated: based on the application objective and results of data pre-processing, we selected 17 substations out of 32 to create substation level samples and 15 of them to generate 300 combinations of substations for aggregated level samples.
- Duration of validation periods: in each sample, we created 7 sub-samples corresponding to different period durations, ranging from 1 day to 3 months.
- The origins of the validation period: in each sub-sample, individual validation periods are created with a rolling-origin approach as origins are modified each time at a daily time step.

### 2.2.4. Clustering method

In this stage, a clustering algorithm has been applied to classify daily load profiles at the substation scale. This combination of dimensions was chosen for two reasons:

- Validity of classification approaches: We can refer to existing studies (Bourdeau et al. 2020), whose findings can guide our methodological choice.
- Independence of profiles: Unlike profiles of other durations generated with a rolling-origin approach, daily profiles are still independent of each other in substation level samples.

In this case study, we applied clustering analysis using the tslearn Python package (Tavenard et al. 2020) to time series from 01/01/2021 to 31/05/2023 to maximize sample size. Focusing on intra-day patterns rather than daily consumption, we normalized daily profiles using mean and standard deviation, executed K-means clustering, and determined the optimal cluster count with the elbow method.

In the customized elbow method, we first calculate the within-cluster sum of squares (WCSS) for number of clusters ranging from 1 to 8, where the maximum value is chosen after a primary clustering test, then identify the elbow point with a quantitative approach by comparing a decrease rate, as in equation (2), to a chosen threshold value. The optimal cluster number will be the largest number before the decrease rate reaches this threshold.

$$decrease\ rate = (WCSS[i + 1] - WCSS[i]) / (WCSS[i] - WCSS[0]) \quad (2)$$

In this study we set the threshold at 0.15 to best capture minor differences in heating set point management patterns. We also preferred the Euclidean distance metric for its effectiveness over Dynamic Time Warping, as confirmed through a series of clustering tests.

### 2.2.5. Calculation of validation metrics

Two commonly used metrics: Normalized Mean Bias Error (NMBE) and Coefficient of Variation of Root-Mean-Square Error (CVRMSE) are selected in this case study and these metrics are calculated at profile / cluster / substation / aggregated levels. Metrics are calculated by equations (3) and (4).

$$NMBE = \frac{1}{\bar{M}} \times \frac{\sum_{i=1}^n (P_i - M_i)}{n} \times 100\% \quad (3)$$

$$CVRMSE = \frac{1}{\bar{M}} \times \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - M_i)^2} \times 100\% \quad (4)$$

Where:

- $P_i$  represents the predicted values,
- $M_i$  represents the measured values,
- $\bar{M}$  represents the mean of measured values,
- $n$  is the number of measurements.

	Nb of time series	Validation period lengths	Period origins
Substation level	17 substations	1/3/7/14/30/60/90 days	01/01/2021 – 31/12/2022
Aggregated level	300 combinations of 5 substations out of 15	1/3/7/14/30/60/90 days	01/01/2021 – 31/12/2022

*Table 1: Four dimensions in created validation samples.*

## 3. RESULTS

### 3.1. CLUSTERING RESULTS

Optimal cluster numbers range from 3 to 5, showcasing distinct daily heating patterns influenced by temperature setpoints variations. Noteworthy patterns include Substation 20's increase in power supply at 23h during January 2021 (Figure 2) and Substation 3's temperature decrease every Monday night (Figure 3a). Weather conditions significantly impact clustering where temperature setpoints are consistent among periods, with variations in heating degree day and Global Horizontal Irradiance (GHI) affecting cluster formation (Figure 3b).

### 3.2. VALIDATION OF CLUSTERED DAILY PROFILES

In the validation of clustered daily profiles, we assessed cluster-level validation metrics alongside the distributions of individual profile metrics, as partially illustrated in Table 2 and Figure 3c. The findings indicate that the metrics at the cluster level are consistent with the average values of the samples. However, variations become evident when comparing metrics across different clusters and against substation-level metrics, highlighting how operational settings and weather conditions influence model performance.

The impact of load pattern on model validity is quite different: long-duration setpoint adjustments significantly impact validation metrics, notably in cluster 1 in Substation 3 where NMBE indicates daily consumption overestimation (Table 2) and CVRMSE suggests a poorer model validity compared to cluster 0 periods under similar weather conditions. Meanwhile, the minor adjustments show negligible impact on the validation metrics.

Weather conditions also play a critical role in model validity, as differentiated clusters, based on heating degree days and GHI, show varying model performance. This is particularly true for periods

with same setpoints but diverse weather conditions, such as cluster 0, 2 and 3 in Substation 3 (Table 2). In Substation 3, we obtained a substation-level NMBE of 2.5% while cluster-level NMBE range from -18.7% to 25.4% and none of them is as good as substation-level result. Similar results in other substations indicate a tendency for the current model to perform better during full heating seasons and overestimate energy usage during milder seasons

### 3.3. VALIDATION WITH DIFFERENT PERIOD DURATIONS

The variability of validation results calculated under different validation period durations is measured through standard deviation respectively at substation and aggregated level.

#### 3.3.1. Substation level results

At the substation level, standard deviations generally decrease as the validation period duration increases from 1 day to 3 months. This trend suggests an increasing stability of validation metrics over longer periods. Despite of the global trend of decreasing variability, the maximal standard deviations in sub-samples don't necessarily follow such trend as illustrated in Figure 4. However, due to the limited number of sub-samples at substation level (17), it is hard to interpret more on such phenomena.

#### 3.3.2. Aggregated level results

Similar trends at substation level have been observed again at aggregated level, such as the non-monotony of maximal standard deviations in samples (Figure 5). While the trend of decreasing variability with increasing validation period duration remains evident, the aggregated level results exhibit less variability than those at the substation level. Besides, when the period duration comes to 2 months, variabilities for all metrics at aggregated level are significantly reduced, which is not the case at substation level.

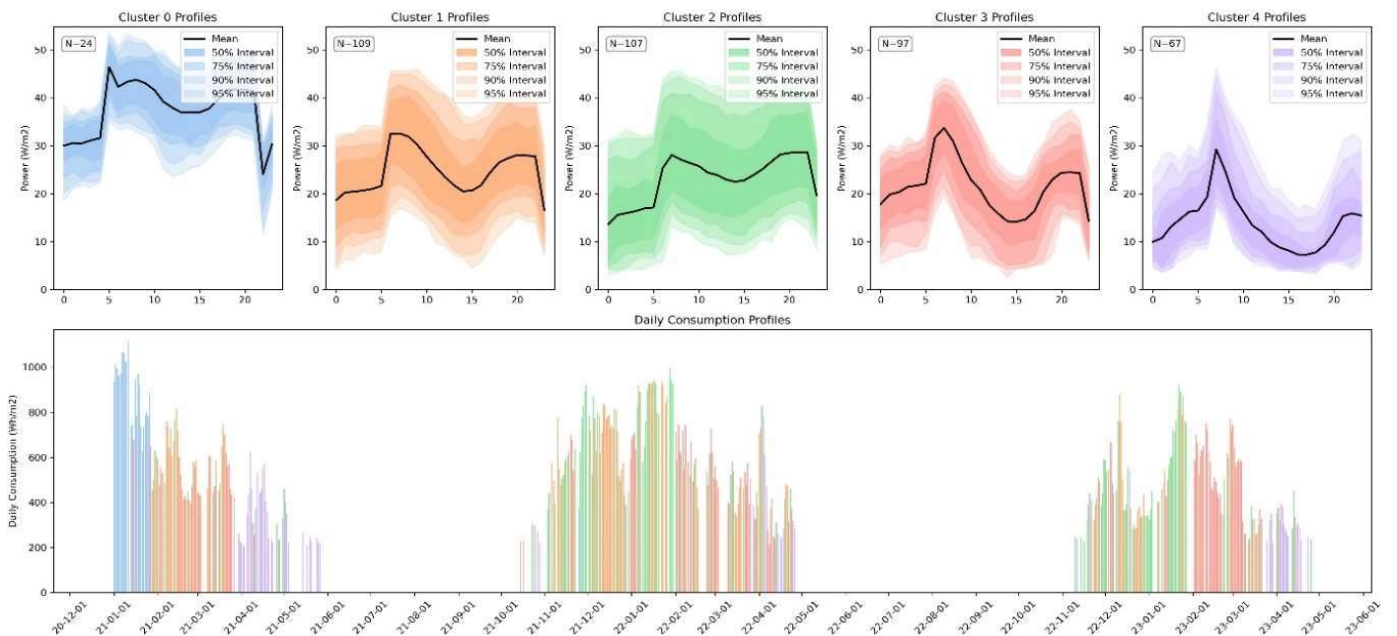


Figure 2: Clustering results in Substation 20. (Upper: Mean profiles and confidence intervals. Lower: Daily consumptions.)

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	SST3
NMBE	0%	-18.7%	9.4%	25.4%	2.5%
CVRMSE	64.5%	84.1%	78.3%	97.1%	73.4%

Table 2: Validation results at cluster and substation level for substation 3

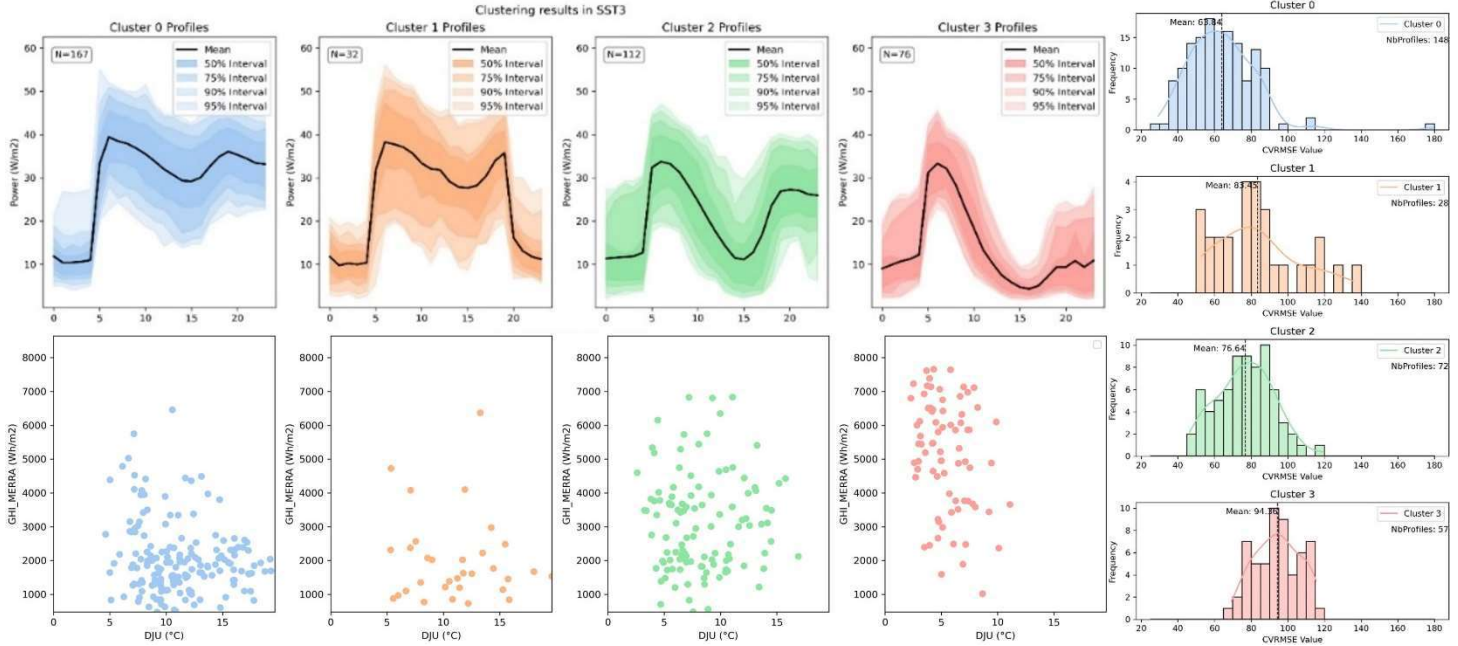


Figure 3: Clustering and validation results in Substation 3 (Upper left 3a: comparison of measured profiles in each cluster; lower left 3b: distribution of profiles in each cluster and related weather conditions; right 3c: distribution of individual level CVRMSE in each cluster).

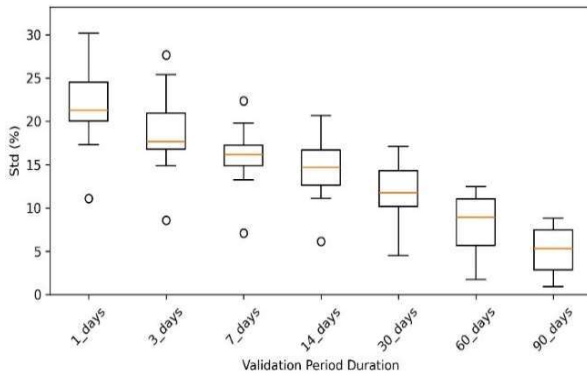


Figure 4: Distributions of STD for NMBE at substation level in 7 sub-samples.

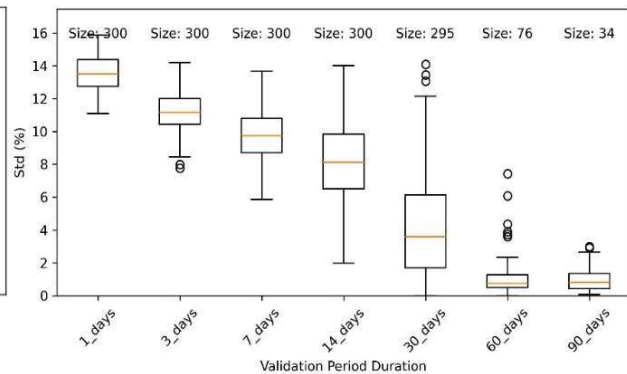


Figure 5: Distributions of STD for CVRMSE at aggregated level in 7 sub-samples.

## 4. DISCUSSIONS

### 4.1. CLUSTERING IN UBEM VALIDATION

The clustering results suggest that the presence of specific heating load patterns over a long period is not uncommon, occurring regularly (Substation 3) or uniquely during specific periods (Substation 20). These findings enhance our understanding of actual heating system operation patterns in buildings, and such patterns, with their impacts on simulation and validation results, help explain the causes of validation errors related to inappropriate temperature setpoint parameters. Additionally, deviations in errors across clusters indicate that a classification process in a validation workflow can prevent poorer results from being obscured by the average over a longer validation period.

By analysing the features of clusters without influence of operational patterns, we can particularly associate the deviation of errors across clusters with the model validity under different weather conditions. This analysis can help to deduce the validity domain of the current model and provides insights into model diagnostics. An important conclusion from this case study regarding the



performance of DIMOSIM is that its validity decreases under milder weather conditions with lower heating degree day and higher GHI.

We also find that while the current clustering approach with time series normalization captures minor differences in operational patterns, it does not perfectly differentiate clusters through weather conditions since the distributions overlap (Figure 3b). Therefore, to better study the model validity under various weather conditions, a different clustering approach, such as feature-based clustering or the same approach without normalization, might be more suitable.

In conclusion, a clustering-based approach is valuable in model validation as it provides crucial insights into the causes of validation errors and aids in UBEM diagnostics. However, an appropriate clustering approach should be carefully chosen based on the objectives of the study.

#### 4.2. IMPACTS OF VALIDATION PERIOD DURATION

The analysis clearly demonstrates the significant impact of validation period duration on the variability of validation metrics. Longer periods tend to provide more stable validation outcomes at both the substation and aggregated levels. Moreover, aggregating data across multiple substations tends to mitigate some of the fluctuations observed in individual substations, as evidenced by reduced variability among periods longer than two months (Figure 5). However, such conclusions may not apply universally to all UBEM uses. For instance, the maximal values of variabilities, which may show non-monotone characteristics, deserve more attention in sizing studies.

Based on these findings, we recommend longer validation periods for validating UBEM dynamic outputs. Nonetheless, the choice of validation duration for other use cases or models should also consider the study objectives, data availability, and the modelling approach. Despite better stability achieved with longer periods, the dispersion in samples created by the rolling-origin approach is still significant. Therefore, a multi-period validation approach should be adopted when data is available.

### 5. CONCLUSION AND PERSPECTIVES

In the current article, a new methodology for the validation of UBEM dynamic outputs is proposed and applied to a case study in Blagnac, France. The proposed methodology incorporates a classification step and multi-period validation that aims to address the uncertainty in validation periods due to specific events. K-means algorithm applied on normalized time series using Euclidean distance is demonstrated to be an adequate method to capture minor difference between thermal load profiles. Clustering results also suggest that a classification process of measured load profiles can be beneficial for model validation. Furthermore, the application of methodology successfully demonstrated that even for long validation periods (monthly or quarterly), it is crucial to consider such uncertainty while validating the performance of a UBEM. Such conclusion is extensible to aggregated level validation.

For future research, we look forward to: (1) expand the validation methodology to a wider range of urban contexts and building types to test its generalizability; (2) use classification results in model calibration; (3) place model calibration in the current methodology while considering additional temporal dimension that could be brought to the sample size, such as lead time; (4) integrate current methodology in a larger validation framework for UBEMs while considering their use cases.

## 6. BIBLIOGRAPHY

- Ansanay-Alex, Guillaume, Yassine Abdelouadoud, and Pascal Schetelat. 2016. “Statistical and Stochastic Modelling of French Households and Their Energy Consuming Activities.”
- Bourdeau, Mathieu, Philippe Basset, Solène Beauchêne, David Da Silva, David Werner, and Elyes Nefzaoui. 2020. “Classification of Daily Electric Load Profiles of Non-Residential Buildings.”
- COSTIC. 2021. “Services Degrés-Jours Unifiés 2021-2022.”
- CSTB. 2024. “Base de Données Nationale Des Bâtiments.” 2024. <https://bdnb.io/>.
- Fildes, Robert. 1992. “The Evaluation of Extrapolative Forecasting Methods.” *International Journal of Forecasting*. Vol. 8.
- Garreau, Enora, Yassine Abdelouadoud, Eunice Herrera, Werner Keilholz, G E Kyriakodis, Vincent Partenay, and Peter Riederer. 2021. “District MODeller and SIMulator (DIMOSIM) – A Dynamic Simulation Platform Based on a Bottom-up Approach for District and Territory Energetic Assessment.” *Energy and Buildings* 251 (March).
- Hedegaard, Rasmus Elbæk, Martin Heine Kristensen, Theis Heidmann Pedersen, Adam Brun, and Steffen Petersen. 2019. “Bottom-up Modelling Methodology for Urban-Scale Analysis of Residential Space Heating Demand Response.” *Applied Energy* 242 (March): 181–204.
- Jaeger, Ina De. 2021. “On the Impact of Input Data Uncertainty on the Reliability of Urban Building Energy Models.”
- Jensen, Soren Ostergaard. 1995. “Validation of Building Energy Simulation Programs: a Methodology.” *Energy and Buildings* 22: 133–44.
- Lefort, Lucie. 2022. “Methodological Developments for the Validation of Urban Building Energy Models.”
- Oberkampf, William L, and Christopher J Roy. 2013. “Model Validation Fundamentals.” In *Verification and Validation in Scientific Computing*, 371–408. Cambridge University Press.
- Oraiopoulos, A, and Howard B. 2022. “On the Accuracy of Urban Building Energy Modelling.” *Renewable and Sustainable Energy Reviews* 158 (March): 111976.
- Pack, David J. 1990. “In Defense of ARIMA Modeling.” *International Journal of Forecasting*.
- Prataviera, Enrico, Jacopo Vivian, Giulia Lombardo, and Angelo Zarrella. 2022. “Evaluation of the Impact of Input Uncertainty on Urban Building Energy Simulations Using Uncertainty and Sensitivity Analysis.” *Applied Energy* 311 (March): 118691.
- Tashman, Leonard J. 2000. “Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review.” *International Journal of Forecasting*.
- Tavenard, Romain, Johann Faouzi, Gilles Vandewiele, Felix Divo, Guillaume Androz, Chester Holtz, Marie Payne, Roman Yurchak, Marc Rußwurm, Kushal Kolar, and Eli Woods. 2020. “Tslearn, A Machine Learning Toolkit for Time Series Data.” *Journal of Machine Learning Research* 21: 1–6.