Enhancing buildings' energy efficiency prediction through advanced data fusion and fuzzy classification

Marc Grossouvre, Didier Rullière, Jonathan Villot

Enhancing buildings' energy efficiency prediction through advanced data fusion and fuzzy classification

Marc Grossouvre*[1,2], Didier Rullière†[1], and Jonathan Villot‡[3]

[1]Mines Saint-Etienne, Univ Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158 LIMOS, F - 42023 Saint-Etienne France.
[2]U.R.B.S., Bâtiment des Hautes Technologies, 20 Rue Professeur Benoit Lauras, F - 42000 Saint-Etienne France
[3]Mines Saint-Etienne, Univ Lyon, CNRS, Univ Jean Monnet, Univ Lumiere Lyon 2, Univ Lyon 3 Jean Moulin, ENS Lyon, ENTPE, ENSA Lyon, UMR 5600 EVS, Institut Henri Fayol, F - 42023 Saint-Etienne France.

March 20, 2024

Abstract

This study proposes a method to predict buildings' energy efficiency based on available descriptive information and without a physical visit, by merging diverse datasets and employing advanced classification techniques. By integrating geographical, structural, legal, and socio-economic data with Energy Performance Certificate (EPC) observations, our approach yields a rich learning set. Through variable selection methods like forward selection with KNN and simultaneous perturbation stochastic approximation for fuzzy KNN, we refine model variables. Comparing fuzzy and hard classification using KNN, Kriging or Random Forest approaches, we find fuzzy classification more adept at capturing nuanced energy inefficiency indicators. Our study highlights the importance of mass energy efficiency prediction for sustainable renovation efforts.

Keywords—fuzzy classification, Kriging, constrained classification, spatial Prediction, energy efficiency, sustainability.

*marcgrossouvre@urbs.fr
†didier.rulliere@emse.fr
‡jonathan.villot@emse.fr

1 Introduction

European policies for sustainability (European Parliament, 2018; 2023) draw scientists, stakeholders, and politics to explore novel approaches to reduce energy consumption and minimise greenhouse gas emissions (GHG). Hence, European countries are defining strategies to enhance the energy performance of anthropogenic activities and address the urgent challenges of climate change. Pursuant to this roadmap, initiatives aimed at enhancing energy-efficient measures in the building sector are pivotal. Indeed, the building sector is one of the world's key energy consumers, accounting for 40% of European energy consumption (Buildings Performance Institute Europe, 2011 and European Commission, 2020). It contributes significantly to greenhouse gas emissions (36% of the total GHG), primarily CO_2, thereby altering our planet's climate, and has been experiencing an overall rising trend over the past decades.

As per the data provided by the French Ministry of Finance in 2023, France has 40 million residential buildings, encompassing a total land area of 12,000 square kilometres. Of these, 22 million dwellings were constructed before the first thermal regulations were introduced in 1975. These older buildings are highly demanding in energy, possess poor thermal properties, and lack insulation (ECOFYS, 2005). They represent 55% of the housing sector and contribute to over 75% of its energy consumption. Their renovation has therefore been a priority for the last fifteen years (Van de Maele, 2010) for several reasons: the building stock presents significant potential for energy saving (Ballarini et al., 2014); refurbishing buildings is the most profitable sector in terms of CO_2 decrease per invested Euro (Storck et al., 2023); the long lifespan of buildings amplifies the consequences of a wrong design. Studies also suggest that refurbishment, rather than demolition, is more effective based on time, cost, community impact, prevention of urban sprawl, reuse of existing infrastructure, and protection of established communities. By renovating buildings to high efficiency standards, ambitious climate change mitigation actions align with improvements in living quality.

However, undertaking energy efficiency refurbishment is a complex process involving many stakeholders. This aspect challenges city planners and municipal decision-makers, who have a crucial role to play (Hrabovszky-Horváth et al., 2013). Effective policy-making necessitates a comprehensive understanding of the building stock (Cappelletti et al., 2015). Yet, the technical literature (Johansson et al., 2017) and our empirical experience suggest that data collection is a barrier among the local administrations for strategic and political decisions. This is due to the dispersion of data among several municipal offices and other administrative entities, the lack of interoperability among the data collection systems, and notably the absence of energy efficiency assessments for each and every building and dwelling in the town (Swan and Ugursal, 2009; Caputo and Pasetti, 2015).

To overcome these difficulties, building stock energy prediction models have been developed over the last 15 years (Swan and Ugursal, 2009; Reinhart and Cerezo Davila, 2016). Simplified and data-driven approach models, also known as "bottom-up" models, represent a valuable method for assessing the consumption of a city's building stock (Ahmad et al.,

2017; Mastrucci et al., 2017). Bottom-up energy models are classified into two main subcategories: engineering models and statistical models (Fumo, 2014; Foucquier et al., 2013). On one hand, engineering, or numerical, models address the energy consumption questions with a dynamic approach based on equations describing the physical and thermal behaviour of the building (Wang and Zhai, 2016). However, these approaches are often driven by archetypes and sampling methods that rarely consider the local variability of building characteristics (Mastrucci et al., 2017), nor do they incorporate the incremental energy measures implemented in older buildings due to renovation strategies applied by municipalities. On the other hand, statistical models, based on machine learning algorithms, rely on numerous ground observations and can yield high prediction capabilities as provided that physical indicators describing a building are available, see Al-Shargabi et al. (2022) for a comprehensive review. While some efforts have been made in feature reduction Ali et al. (2020), the remaining features are still challenging to infer without a physical visit to the building (Schetelat et al., 2020).

Defining categories of buildings based on their energy efficiency is now widespread. In the USA, the Energy Star programme evaluates the energy efficiency of homes based on criteria such as insulation, windows, heating and cooling systems. China has adopted a holistic indicator known as the "Three Star" building rating system, which assesses the overall sustainability of a building in terms of land efficiency, energy efficiency, water efficiency, resource efficiency, and environmental quality, among others. In Canada, dwellings are classified using the EnerGuide rating system, which provides the energy consumption per square metre and per year. A similar building classification according to their energy performances, whether real or theoretical, has been defined in all E.U. countries. These labels are used to identify the energy sieves and target the renovation efforts; they may also be used to assess present and future energy needs. Some countries have opted for more qualitative indicators, while France has chosen an indicator based on quantitative measures. An Energy Performance Certificate (EPC) is defined in France as the building's energy consumption for standard use, associated with a qualitative labelling letter ranging from A to G. Similarly, a greenhouse gas emission label is defined. The final EPC label is the worst of both. For instance, if a building is labelled C for energy consumption and D for gas emissions, the EPC label is D.

The main goal of the present work is to quantitatively predict the EPC label of each building in France based on available descriptive information without requiring a physical visit. Unlike the publications cited previously, our interest extends beyond assessing the distribution of labels or energy consumption at the area level, such as a city; we aim for the most accurate prediction at the individual building level. It is important to note that EPC is determined by simulating a standard use for the building. Therefore, our work differs from those studies focusing on actual energy consumptions, as, for instance, Khafaga et al. (2023), which provides a comprehensive bibliography of recent works in this area. Furthermore, while real past energy consumption data is typically readily available and valuable for predictions in commercial buildings (Zhang et al., 2023), our focus lies on residential buildings, where access

to real consumption data is constrained by privacy regulations. **In terms of methodology, this paper aims to demonstrate the feasibility of estimating a building's EPC label without requiring a technician visit.** We illustrate that socio-economic features can compensate, to a certain extent, for the lack of technical information about a building. From a technical standpoint, we introduce fuzzy classification as a valuable tool in this context.

The next section introduces the dataset used in our study, consolidating information from various sources to comprehensively characterise residential buildings. We will then detail our methodology for variable selection and the process of learning and predicting EPC labels. Finally, the fourth section presents and discusses the results obtained.

2 Data presentation

This section describes available data sources and the way to merge them to obtain a table that can be used as input for a learning algorithm.

2.1 Information about dwellings

Among the various French institutions collecting information about dwellings, the Ministry of Finance (MoF) is a major player as it requires data to compute property taxes. MoF manages a database of all dwellings in France, geolocated by address and land plot identifier. It provides structural, historical, and qualitative information about each dwelling, including the surface area, number of rooms of each type (bedroom, kitchen, bathroom, etc.), construction materials for the roof and main wall, and year of construction. Some qualitative variables, such as the comfort level and the maintenance quality, are also provided. Another set of variables informs about ownership and occupancy, including date of acquisition, type of owner (private/public, individual/company), occupancy status (owner-occupied, rented, vacant), and rental value. This database has very limited information about energy consumption except for an indicator that identifies a connection of the dwelling to the city gas supply.

The main advantage of the MoF's database is its comprehensiveness, as it inventories all dwellings. Moreover, it provides up-to-date documentation that includes the reliability level of each variable. However, it has limitations, primarily stemming from missing data and a lack of data updating. These issues affect the performance of the algorithms that learn from this database.

In addition to this restricted-access database, open data is also available. The National Institute of Geography provides a 2D model of the territory, outlining the ground print of all buildings. And it is in the process of acquiring LIDAR (Laser Imaging, Detection, And Ranging) data for a 3D model. Other databases contain information about altitude, climate zone, and areas subject to specific regulations, such as heritage protection. Although historically, the government has collected a lot of information about dwellings, there has

been minimal focus on energy consumption. To address the need for improved national energy consumption management, all EPCs are now collected and available for research (see Subsection 2.2).

The above data bases are merged to gather all information about buildings themselves, not only dwellings, see Figure 1. This fusion process results in a table where each row is an address. This table comprises 37 million rows and 275 columns. Many of these columns have characteristics that prevent us from using them in a machine learning algorithm: they may have a specific type, like strings or geometries, or they may contain too much missing data. Among the columns that are valuable to us, we also have to select those that are truly relevant for predicting the EPC label, see Subsection 3.2.

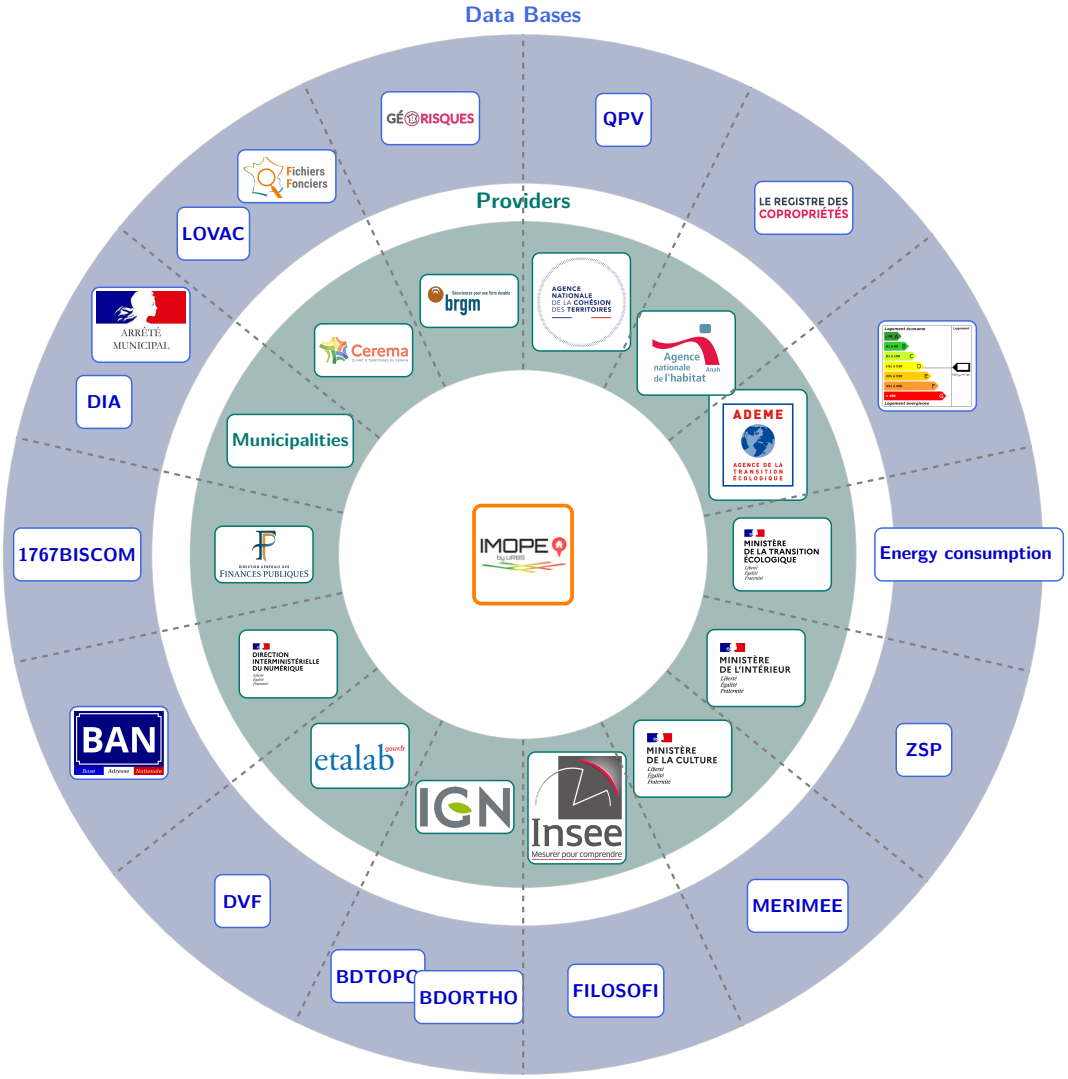

Figure 1: Outer circle: Main data bases that are merged to form the Imope data base. Inner circle: Data provider. The data bases' names as well as the providers' names are clickable.

2.2 Energy efficiency observations

When a dwelling or a building is sold or put up for rent, an Energy Performance Certificate (EPC) must be available for the buyer or tenant. To establish an EPC, a technician visits the dwelling or building, creates a floor plan, gathers information about construction materials, insulation type, windows' specifications and orientation, heating system, air conditioning (if any), hot sanitary water system, and other relevant indicators. This information is entered as input parameters into software that models energy consumption. It calculates a standardised energy consumption, making assumptions about the occupants and their behaviour, neighbours' behaviour (in the case of apartments), and climate conditions. An EPC presents two figures: energy consumption expressed in $kWh/m^2/year$ and greenhouse gas emissions given in $kgCO_2/m^2/year$. Two labels are derived from these quantities, ultimately resulting in an EPC label as described in the Introduction 1 and in Figure 2.

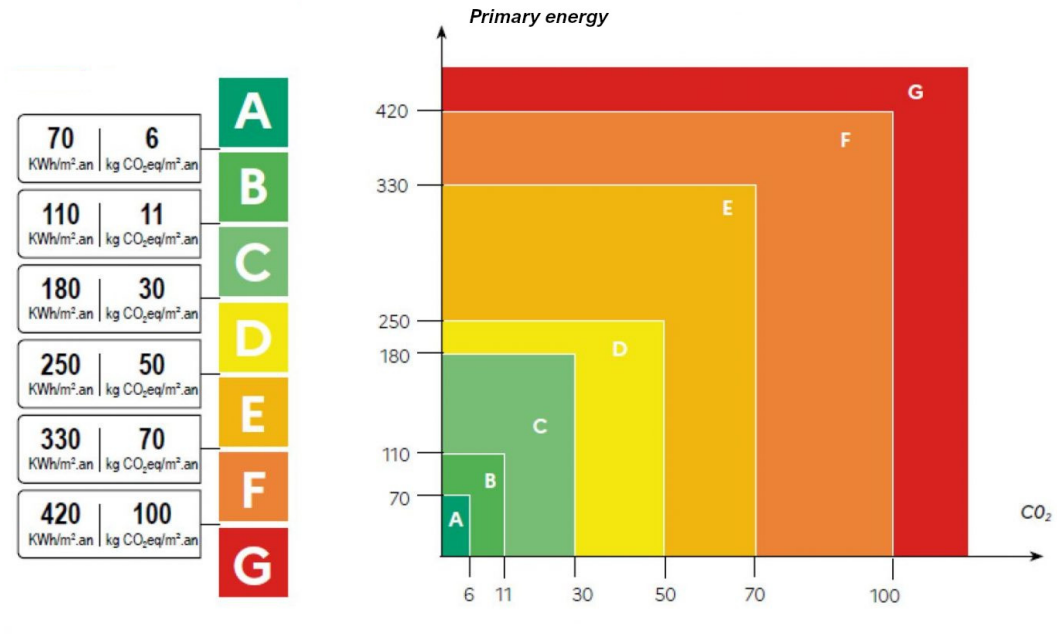

Figure 2: Double threshold process used to determine the EPC label of a building.

The database containing all diagnostics, encompassing every structural, quantitative, and qualitative detail about dwellings, is publicly available as open data and continuously updated by the French Agency For Energy Transition (ADEME). These observations are important in several aspects. They result from direct visual inspections of the buildings, which enhances their reliability. They are used to generate a legal document for which the technician is held accountable. Additionally, they include a set of recommendations to improve energy efficiency. The main limitation of this database is the difficulty in precisely geolocating the visited buildings. This is because only an address is provided, without any land plot specification. Technicians input this address manually, without connecting to the national address database (see Figure 3). Consequently, addresses may lack standardisation, contain

ambiguities, and have missing information.

EPC observed at
“La Montagne 22350 Plumaudan”

Ministry of Finances says that there are
6 houses numbered 1, 2, 3, 4, 6, 8.

Figure 3: Problematic case: Matching EPC observations with Ministry of Finances database. Technician lists visit to "La Montagne" hamlet, which comprises 6 separate houses.

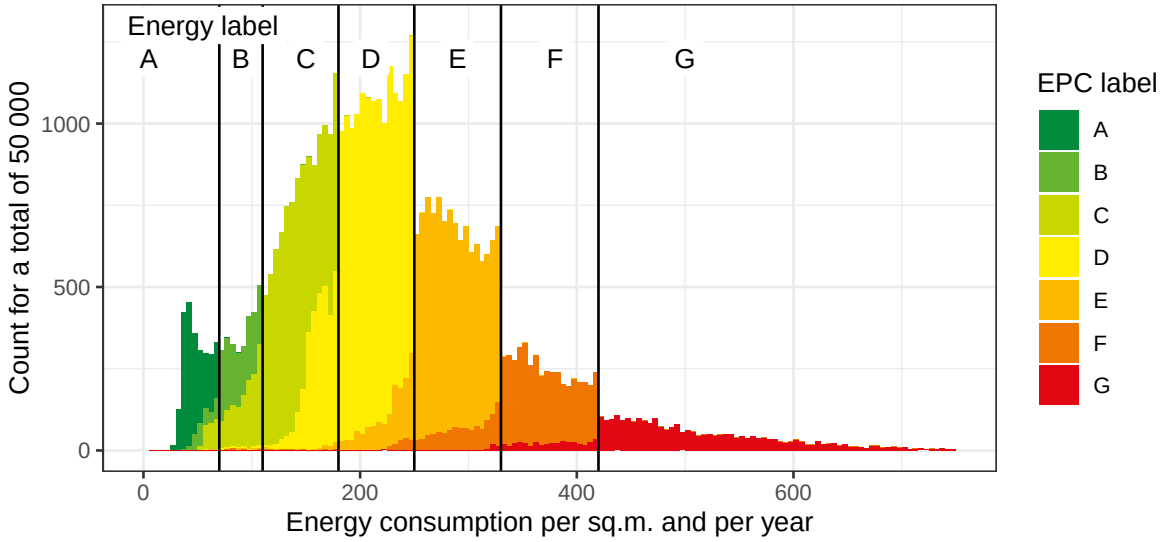

Figure 4: Histogram of observed energy consumption for 50,000 buildings in the French region of Pays de la Loire. Vertical lines indicate energy label thresholds. Colors represent the worst label between energy and GHG labels. Threshold effects are evident between energy labels D to G.

These observations of dwellings' energy efficiency form the learning set for attempting to predict the energy efficiency of all French buildings. In the following, we focus on the French region called Pays de Loire, in the west of France. This region presents a homogeneous climatic environment, and comprises a mix of mid-size cities and rural areas. The distribution of observations is presented in Figure 4 for a representative sample of $50\,000$ buildings.

3 Methodology

Two approaches are possible when trying to predict the EPC. The first is to treat it as regression problem, in which case the target variable is the standardised energy consumption or the GHG emission quantity. The second approach is to consider it as a classification problem. The main goal of this work is to be able to detect energy-inefficient dwellings, also known as energy sieves (EPC labels F and G), and energy-saving dwellings (EPC labels A and B). We do not intend to compute energy consumption or a confidence interval. In fact, considering the known variability of the EPC depending on the technician, we can assume that predicting energy consumption would come with such a large confidence interval that it would cover more than a label span. In this section, we therefore treat the EPC prediction problem as a classification problem. We intend to predict the EPC label at the address level as accurately as possible, respecting, as much as possible, the overall distribution of EPCs over a territory. To measure the model's performance, quantitative indicators are introduced in Subsection 3.1.

The data fusion process summarised in Section 1 produces a table of addresses that contains more than 250 variables. A lot of them are either irrelevant for the EPC, such as the distance of the address to the nearest school, or impossible to value, such as the identifier of the census tract where the address is located. Among the numerical or categorical variables that can be used in a predictive model, there is still some variable selection to perform to reduce as much as possible the noise in the input data. This process is described in Subsection 3.2. Eventually, we propose two supervised fuzzy classification models in Subsections 3.3 and 3.4; one is based on KNN, the other on Kriging.

3.1 Performance indicators

Following early works on fuzzy sets, Ruspini introduced in 1969 a fuzzy classification approach where an individual, in our case a vector in the feature (input) space, is assigned a "degree of membership" for each of the possible classes (fuzzy sets), which are in our case labels A to G. Ruspini introduces the condition for membership degrees to be of sum equal to 1, so that they represent the probability of each class knowing the feature vector. However, his approach has turned out to be very computationally intensive (see Amo et al., 2004). The algorithms presented in this work, using the KNN or Kriging approaches, reach suboptimal results as compared to Ruspini's but in a very reasonable time.

For supervised hard classification, meaning classification that predicts a single class, the base indicator is the confusion matrix, after which other indicators are computed. However, it is seldom used because of its complexity, and, depending on the problem to solve, more synthetic performance indicators can be derived. As far as hard classification is concerned, the confusion matrix is defined to be the matrix of which element (i, j) counts the number of EPCs observed as label i and predicted as label j. We propose here a new definition of the confusion matrix in order to generalise it to fuzzy classification, which involves predicting

membership degrees between 0 and 1 for each class instead of predicting classes. For a given EPC, the seven membership degrees associated with the seven classes sum to 1.

Definition 1 (Confusion matrix, accuracy, balanced accuracy)**.** *Let* $\mathbb{M}$ *and* $\hat{\mathbb{M}}$ *two matrices associated respectively with true membership degrees and predicted membership degrees of a given set of buildings, with one building per row and* $c = 7$ *columns each. The associated confusion matrix is:*

$$\mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}} = \mathbb{M}^\top \hat{\mathbb{M}}$$

The accuracy of the prediction is the sum of the diagonal elements of $\mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}}$ *divided by the sum of all its elements.*

$$Acc_{\mathbb{M},\hat{\mathbb{M}}} = \frac{\operatorname{diag}\left[\mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}}\right] \mathbb{1}_c}{{\mathbb{1}_c}^\top \mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}} \mathbb{1}_c} \qquad where \mathbb{1}_c \text{ is a vector of } c \text{ ones.}$$

The balanced accuracy of the prediction is the mean value of each label's accuracy, which is an element of $\mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}}$*'s diagonal divided by the sum of the elements in its row.*

$$BA_{\mathbb{M},\hat{\mathbb{M}}} = \frac{1}{c}{\mathbb{1}_c}^\top \left(\frac{\operatorname{diag}\left[\mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}}\right]}{\mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}} \mathbb{1}_c} \right) \quad \textit{where the second fraction denotes a term-wise division.}$$

For instance, let us assume that we have $c = 3$ classes and 5 observations:

$$\text{If } \mathbb{M} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \ \hat{\mathbb{M}} = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.7 & 0.1 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.8 & 0 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}, \text{ then we have } \mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}} = \begin{pmatrix} 1.2 & 0.4 & 0.4 \\ 0.4 & 1.4 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$$

$$Acc_{\mathbb{M},\hat{\mathbb{M}}} = \frac{1.2 + 1.4 + 0.3}{5} = 0.58 \text{ and } BA_{\mathbb{M},\hat{\mathbb{M}}} = \frac{1}{3}\left(\frac{1.2}{2} + \frac{1.4}{2} + \frac{0.3}{1}\right) = 0.53\,.$$

In the case of hard classification, true classes and predicted classes are specific instances of fuzzy classification, wherein one membership degree is 1 and the others are null. One can verify that Definition 1 aligns with the usual definitions of accuracy and balanced accuracy for hard classification.

It is worth noting that, as depicted in Figure 4, labels C, D, and E are much more frequent than labels A, B, F, and G. However, decision-makers have a particular interest in identifying buildings labelled F or G (energy sieves) and A or B (energy efficient buildings). A model that exclusively predicts labels C, D, and E may exhibit good accuracy but could still be irrelevant for decision-makers. Therefore, the balanced accuracy indicator aids in identifying models with both good accuracy and relevance.

Since EPC labels A to G are ordered, it is also pertinent to assess the proximity of predictions to the true values. We define the accuracy $\pm$ 1 label ratio as follows:

$$Acc_{\pm 1} = \frac{\sum_{-1 \leq i-j \leq 1} \mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}}[i,j]}{\mathbb{1}_c{}^\top \mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}} \mathbb{1}_c}.$$

For example, if an observation is classified as C, we are interested in knowing if the prediction falls within the range of B, C, or D, and not A, E, F, or G. Similarly, we can extend this concept to define accuracy for ranges beyond $\pm$ 1 label, such as $\pm$ 2, 3, 4, 5, or 6 labels.

In the above example, it yields:

$$Acc_{\pm 1} = \frac{1.2 + 1.4 + 0.3 + 0.4 + 0.4 + 0.4 + 0.2}{5} = 0.86$$

Eventually, since we are interested in producing predictions that reflect the population according to the overall distribution of predicted labels, we compare the label distribution for a representative sample with the distribution of predicted labels for the same sample. This distribution is estimated by computing the membership degree's mean value for each label.

3.2 Variable selection

After completing the data fusion process, each building exhibits a large number of features, not all of which are usable or relevant for predicting the EPC. We have identified the potentially useful features as presented in Table 1.

Variable selection is first implemented using forward selection with KNN for hard classification: Each variable is tested separately; the best one is selected, say v_1; each of the remaining variables is tested with v_1; and the best pair v_1, v_2 is selected. And so on, as long as the performance indicator, in our case, balanced accuracy, improves. The process stops when the balanced accuracy does not improve any more. This variable selection process has been performed for each one of the 12 French regions separately using fifty thousand EPC observations, forming a representative sample of the building stock with regard to the construction period and status (house/apartment building). For each feature, we have identified whether it was selected and its rank. Those who have been selected only once, never, or only in the last steps of the selection process have been ignored.

Based on this first subset of variables, a second variable selection was implemented using the algorithm presented as "A stochastic approximation approach to simultaneous feature weighting and selection for nearest neighbour learners" in (Yeo and Aksakalli, 2021), maximising the balanced accuracy. This algorithm optimises the variables' weights in the distance measure that is used to compute the distance between two individuals, in our case, two buildings. It computes an approximated gradient based on the averaging of multiple directional derivatives. The algorithm performs simultaneously weights optimisation and variable selection, therefore providing a powerful tool, especially calibrated for KNN. In this process, the

Type of feature	Features
Geographic location	Latitude, longitude, and altitude.
Physical features	Roof and main walls material, total living space, number of storeys, house/apartment building.
Descriptive features	Number of dwellings, average living space of dwellings, number of apartments of each type (with 1, 2, 3, 4+ rooms), year of construction for the oldest part of the building, type of energy saving regulation at construction time (identified by an integer increasing for each new regulation), year of construction for the newest part of the building.
Heating system	Individual or collective system, heating source of energy (electricity, city gas, wood, oil, other), availability of a city gas connection in the building.
Dwellings' quality	Comfort level, maintenance level.
Surroundings	Indication of a nearby national heritage building, indication that the building is located in a priority area (meaning a qualified underprivileged area).
Inhabitants & owners	Type of owner (individual, private company, state), type of occupant (owner/tenant), number of vacant dwellings, tax status regarding the occupation (occupied for free, occupied by a farming worker, rented free of furniture, rented as a fully furnished dwelling), number of dwellings that are unfit for renting, indication that a dwelling has been sold in the last year, price per square metre, number of social housing units, number of different owners owning dwellings in the building.

Table 1: Features identified as potentially useful to predict the EPC.

weight of some variables tends towards 0, making it handy for variable selection. Out of the 47 variables, 18 coefficients end up nearly null, while 29 coefficients have non-null values, see Figure 5. In the following, we work with those 29 variables.

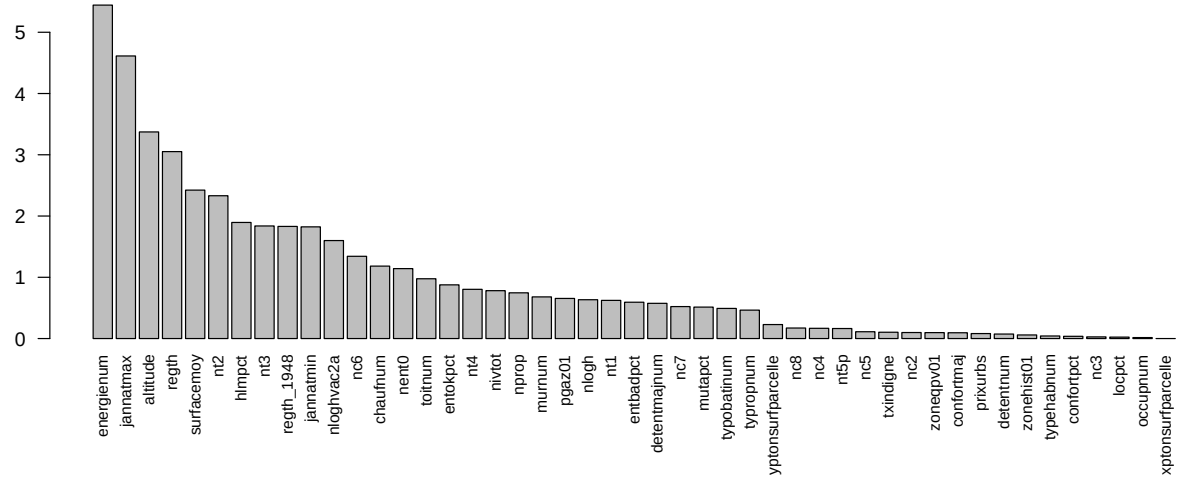

Figure 5: Optimal features' weights for KNN algorithm. See the dictionnary of variables in Appendix A.

3.3 Fuzzy classification with KNN

The fuzzy k-nearest neighbour classifier known as FKNN and presented in Mailagaha Kumbure et al. (2020) assigns a membership degree to each class of a given categorical feature for any unobserved individual. In the case of EPC prediction, this means that for each unobserved building, membership degrees can be predicted for all 7 EPC labels from A to G. These membership degrees are positive and sum to 1.

The algorithm proceeds as follows:

1. Begin with a labelled set consisting of buildings with known EPC labels, forming the observed buildings.
2. Select an unobserved building y.
3. For each EPC label L ranging from A to G, find the k-nearest neighbours of y that have label L in the labelled set. They are denoted $x_1^L, ...x_k^L$.
4. For each EPC label L, compute the membership degree $u_L(y)$ of y in class L:

$$u_L(y) = \frac{\sum_{j=1}^{k} 1/||y - x_j^L||^2}{\sum_{L=A}^{G} \sum_{j=1}^{k} 1/||y - x_j^L||^2} \tag{1}$$

In Equation 1, $||.||^2$ represents the squared Euclidean distance. However, refining this model by applying rescaling factors to feature variables is of interest. In this case, each feature is

divided by a positive number. These factors are optimised with the same stochastic method employed for variable selection, maximising the balanced accuracy. Additionally, if one is interested in hard classification, the label with the largest membership degree is attributed to y.

Although membership degrees in FKNN are not strictly defined as probabilities, they possess properties akin to probabilities, allowing for interpretation as such. Notably, each class has a strictly positive membership degree. However, given the ordered nature of classes in our scenario, if an individual is very likely to be of class A, it might be likely to be of class B, but it should be very unlikely to observe it in class G. The model presented in Subsection 3.4 introduces the possibility of having negative membership degrees.

3.4 Fuzzy classification with Kriging

KNN is a classification algorithm that predicts the class of a given individual by considering a finite number of its neighbours. It is reasonable to assume that results could improve if we consider all neighbours, assigning them decreasing importance as they are further from the individual being predicted. Thus, instead of considering the number of nearest neighbours, Kriging replaces this with a characteristic distance, often referred to as range. While the statistical interpretation of KNN is challenging, Kriging minimises the predicted mean error, ensuring the best predictor in a well-defined sense. Moreover, Kriging can be expressed with a close formula that allows for the inclusion of constraints. However, Kriging is not inherently a classification algorithm, and additional conditions are needed to use it as a fuzzy classifier, which is the objective of Joint Kriging.

The Joint Kriging algorithm (Rullière and Grossouvre, 2023b,a) assigns a classification score for each EPC label to an unobserved building. These classification scores sum to 1 and, when positive, can be interpreted as probabilistic membership degrees. However, these classification scores may also take negative values, indicating both positive and negative confidence levels for each class. When assessing the model, one can encounter a confusion matrix with negative elements. For instance, considering the example presented to illustrate Definition 1, one could observe a confusion matrix such as presented in equation 2. In this context, predicting a negative classification score for label G for a given individual is interpreted as the individual being a counter-example of label G. In terms of ordered classes, the individual is "far" from being in class G.

$$\mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}} = \begin{pmatrix} 1.4 & 0.8 & -0.2 \\ 0.4 & 1.4 & 0.2 \\ -0.1 & 0.4 & 0.7 \end{pmatrix} \tag{2}$$

When predicting a set of unobserved buildings, the model can be further constrained so that the average classification scores for each label are defined by the user. This is particularly

valuable for predicting EPC labels because the labelled set is large enough to extract a subset that is representative of the complete building stock. Consequently, an estimated average distribution of labels can be derived. Moreover, as mentioned in Subsection 3.1, it is desirable to minimize the risk of exclusively predicting labels C, D and E. By constraining the average membership degree for each label, the model ensures sufficient weight is assigned to each label.

This model offers flexibility, as it may be constrained to simulate multiple scenarios of the predicted output which FKNN can't do. It takes into account the ordered nature of the classes. However, the probabilistic aspect is lost due to the negative classification scores.

4 Results and discussion

Here is the learning/test process for each algorithm:

KNN For the region Pays de Loire in France, a FKNN model coupled with SPSA pseudo-gradient descent was run based on the 29 selected variables. The model was trained on a sample of 15,000 randomly selected buildings from observations. Predictions were made using 10-fold cross-validation, selecting the 3 nearest neighbours for each label. The resulting weights were tested using the learning set to predict a test set of 50,000 observations, representative of the complete building stock. Membership degrees (positive and summing to 1) were predicted for fuzzy classification and binarized for hard classification.

Joint Kriging Based on the 29 selected variables, a joint Kriging model was run on a balanced sample (same number of observations of each label) of $5\,000$ observations. A preliminary step of variable selection reduced the number of variables to 9. The learning sample was then used to predict a test set of $6\,000$ observations, representative of the building stock. Joint Kriging predicts classification scores,which can also be binarized.

Random Forest The same learning and test sets as for Joint Kriging were used. Variable selection was performed based on the same 29 selected variables using VSURF algorithm (Genuer et al., 2015) resulting in 9 selected variables. Random Forest is a hard classifier and does not predict membership degrees.

The list of selected variables can be found in Appendix B and the dictionary of variables is available in Appendix A. The selection of variables informing about the building's age or about the heating system and source of energy is expected. However, neither Joint Kriging nor Random Forest select variables informing about the building's material (walls, roof). Instead, these models favoured socio-economic variables such as the number of owners occupying their dwellings and the percentage of dwellings under the social housing system.

Fuzzy classification							
Indicator	FKNN + SPSA			Joint Kriging			Random Forest
Model optimization criterion	Balanced accuracy			Balanced accuracy			Gini impurity
Balanced accuracy	0.269			**0.434**			N.A.
Accuracy	0.284			**0.387**			N.A.
Accuracy ± 1 label	62.5%			**82.6%**			N.A.
Accuracy of A or B	45.0%			**94.3%**			N.A.
Accuracy of C, D or E	66.7%			**85.1%**			N.A.
Accuracy of F or G	35.0%			**46.7%**			N.A.
Adequacy between learnt and predicted distributions		True	Predicted		True	Predicted	N.A.
	A	0.040	0.036	A	0.034	0.034	
	B	0.035	0.028	B	0.027	0.027	
	C	0.191	0.181	C	0.191	0.191	
	D	0.334	0.370	D	0.347	0.347	
	E	0.229	0.242	E	0.234	0.234	
	F	0.109	0.093	F	0.104	0.104	
	G	0.060	0.050	G	0.063	0.063	

Table 2: Performances of the 3 compared models for fuzzy classification.

Hard classification									
Indicator	FKNN + SPSA			Joint Kriging			Random Forest		
Balanced accuracy	0.358			0.383			**0.451**		
Accuracy	**0.409**			0.387			0.371		
Accuracy ± 1 label	**79.5%**			73.6%			72.4%		
Accuracy of A or B	49.7%			63.3%			**78.3%**		
Accuracy of C, D or E	**86.6%**			76.1%			64.2%		
Accuracy of F or G	36.5%			45.2%			**66.0%**		
Adequacy between learnt and predicted distributions		True	Predicted		True	Predicted		True	Predicted
	A	0.040	0.036	A	0.034	0.066	A	0.037	0.075
	B	0.035	0.028	B	0.027	0.044	B	0.032	0.083
	C	0.191	0.181	C	0.191	0.130	C	0.198	0.161
	D	0.334	0.370	D	0.347	0.411	D	0.337	0.187
	E	0.229	0.242	E	0.234	0.150	E	0.229	0.201
	F	0.109	0.093	F	0.104	0.082	F	0.105	0.148
	G	0.060	0.050	G	0.063	0.119	G	0.062	0.146

Table 3: Performances of the 3 compared models for hard classification.

Moreover, both Joint Kriging and Random Forest select latitude and longitude as meaningful variables, indicating that EPCs are geographic information in the sense that a building located near an observed building is likely to have the same EPC label as the observed building.

In addition to the results given in Tables 2 and 3, the complete confusion matrices are available in Appendix C. These results demonstrate a diversity of behaviours among models. When considering balanced accuracy, our key indicator for this study, Joint Kriging performs best (0.434) for fuzzy classification, while Random Forest performs best (0.451) for hard classification. However, while Joint Kriging maintains its superior performance in terms of accuracy for fuzzy classification, Random Forest is surpassed by both KNN and Joint Kriging in the case of hard classification. This suggests that Random Forest struggles to predict the more common classes accurately but performs well in rare classes. Consequently, Random Forest predicts labels A, B, F, and G more frequently than their actual occurrence, with labels F and G being predicted 76% more frequently than their actual frequency.

Joint Kriging demonstrates strong scores across all indicators for fuzzy classification and predicts a label distribution identical to the actual population distribution, making it the most effective model overall. However, when classification scores are binarized for hard classification, there is a decrease in performance for Joint Kriging, although its performance remains consistent between KNN and Random Forest. The frequency of predicted energy sieves (labels F and G) is only 18% higher than their actual frequency.

While FKNN is outperformed by Joint Kriging for fuzzy classification, its performance significantly improves when membership degrees are binarized for hard classification. Particularly, it achieves the best accuracy within one label. Although FKNN underestimates the frequencies of labels A, B, F, and G, the discrepancy with observed frequencies is much smaller than that of Random Forest.

Overall, it is noteworthy that these models can reasonably predict the EPC label with a minimal number of variables compared to the parameters required to compute a building's energy efficiency.

We are also interested in extracting information from the fuzzy classification predictions, membership degrees for FKNN and classification scores for Joint Kriging. Table 4 illustrates that fuzzy classification effectively captures class orders. For a set of buildings with a given label, we compute the mean values of the fuzzy indicators, membership degree of KNN and classification score for Joint Kriging. The mean fuzzy indicator is consistently highest for the true label, with the true label's neighbours being given the two next largest values. For example, according to Joint Kriging, buildings with true label F have a mean classification score of 0.24 for F, 0.23 and 0.16 for G and E, with the four other scores considerably smaller. This observation raises questions about the probability of finding the true label among the top two or three fuzzy indicators. In the case of FKNN, the true label is among the top three membership degrees in 83% of the sample studied. Similarly, for Joint Kriging, 76% of energy sieves (true labels F or G) have F or G among their top two classification scores.

These results highlight the added value of fuzzy classification for detecting potential energy sieves.

	mean classification score in Fuzzy Joint Kriging						
true EPC	A	B	C	D	E	F	G
A	**0.87**	**0.28**	0.05	0.00	-0.03	-0.02	-0.02
B	**0.25**	**0.44**	**0.10**	0.01	-0.04	-0.05	-0.05
C	0.02	**0.16**	**0.40**	**0.20**	0.11	0.10	0.06
D	0.11	0.19	**0.33**	**0.43**	**0.36**	0.29	0.22
E	-0.03	0.05	0.15	**0.25**	**0.32**	**0.27**	0.23
F	-0.11	-0.05	0.01	0.08	**0.16**	**0.24**	**0.23**
G	-0.10	-0.07	-0.03	0.03	0.11	**0.16**	**0.35**

	mean membership degree in Fuzzy KNN						
true EPC	A	B	C	D	E	F	G
A	**0.32**	**0.14**	0.07	0.07	0.07	0.06	0.08
B	**0.16**	**0.28**	**0.09**	0.07	0.06	0.06	0.06
C	0.14	**0.20**	**0.34**	**0.18**	0.13	0.12	0.12
D	0.15	0.16	**0.23**	**0.31**	**0.22**	0.20	0.18
E	0.11	0.10	0.13	**0.19**	**0.26**	**0.22**	0.20
F	0.07	0.07	0.08	0.11	**0.15**	**0.20**	**0.19**
G	0.05	0.05	0.05	0.08	0.11	**0.15**	**0.17**

Table 4: Fuzzy classification in relation with true labels.

5 Conclusion

After presenting the scientific context and the main goals of this work, a data fusion approach has been implemented to construct a data table that gathers all available information about dwellings, including geographical, structural, legal, or socio-economic aspects. This data table has been matched with EPC observations, creating a learning set comprising millions of observations and hundreds of features. To learn from this dataset, variable selection was necessary. Forward selection with KNN reduced the number of variables to 47. SPSA for FKNN reduced this subset to 29. Forward selection with Joint Kriging further reduced the number of variables down to 9, with the same number of variables selected by the VSURF algorithm. The results of fuzzy classification and hard classification were compared using the same parameters, thanks to a generalization of confusion matrices.

Results indicate that for hard classification, if an EPC label is predicted, there is a 70 to 80% probability that the true label matches the predicted label or one of the two adjacent labels. While this may not be sufficient for legal purposes, it is adequate for identifying buildings likely to be energy-inefficient or energy efficient, which is the primary objective of this article. Although Random Forest appears to have promising results for hard classification,

it significantly distorts the distribution of labels, resulting in a considerable overestimation of energy-inefficient buildings, which is a notable drawback. Joint Kriging and FKNN fairly reproduce the overall distribution of labels, but energy-inefficient buildings remain challenging to predict, as half of them are not detected. These challenges justify the decision to employ fuzzy classification, which proves efficient in capturing secondary information indicative of energy-inefficient buildings.

Despite our efforts, we were unable to find any quantitative results to compare this work with. However, the mass prediction of buildings energy efficiency for enhancing renovation efforts is undeniably a major challenge, and we hope that this work will encourage other teams to publish their methodologies and results. Only then can we truly advance sustainability.

Acknowledgments This work has been jointly funded by Mines Saint-Étienne School of Engineering, U.R.B.S. company, and a Ph.D. grant from ANRT (French National Agency for Research and Technology). The data fusion process presented in Section 2 is the result of teamwork in which U.R.B.S. engineers played a crucial role, and their contribution is acknowledged. The authors also thank Nathan Seychal for implementing the SPSA algorithm for fuzzy KNN. Additionally, they would like to express their gratitude to Maximilien Brossard for his advice and support.

References

Europe's buildings under the microscope, BPIE - Buildings Performance Institute Europe, Sept. 2011. URL https://www.bpie.eu/publication/europes-buildings-under-the-microscope/.

Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency (Text with EEA relevance), May 2018. URL http://data.europa.eu/eli/dir/2018/844/oj/eng.

In focus: Energy efficiency in buildings - European Commission, Feb. 2020. URL https://commission.europa.eu/news/focus-energy-efficiency-buildings-2020-02-17_en.

Directive (EU) 2023/1791 of the European Parliament and of the Council of 13 September 2023 on energy efficiency and amending Regulation (EU) 2023/955 (recast) (Text with EEA relevance), Sept. 2023. URL http://data.europa.eu/eli/dir/2023/1791/oj/eng.

M. W. Ahmad, M. Mourshed, and Y. Rezgui. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, 147:77–89, July 2017. ISSN 0378-7788. doi: 10.1016/j.enbuild.2017.04.038. URL http://www.sciencedirect.com/science/article/pii/S0378778816313937.

A. A. Al-Shargabi, A. Almhafdy, D. M. Ibrahim, M. Alghieth, and F. Chiclana. Buildings' energy consumption prediction models based on buildings' characteristics: Research trends, taxonomy, and performance measures. *Journal of Building Engineering*, 54:104577, Aug. 2022. ISSN 2352-7102. doi: 10.1016/j.jobe.2022.104577. URL https://www.sciencedirect.com/science/article/pii/S2352710222005903.

U. Ali, M. H. Shamsi, M. Bohacek, C. Hoare, K. Purcell, E. Mangina, and J. O'Donnell. A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings. *Applied Energy*, 267, June 2020. ISSN 0306-2619. doi: 10.1016/j.apenergy.2020.114861. URL https://www.sciencedirect.com/science/article/pii/S0306261920303731.

A. Amo, J. Montero, G. Biging, and V. Cutello. Fuzzy classification systems. *European Journal of Operational Research*, 156(2):495–507, July 2004. ISSN 0377-2217. doi: 10.1016/S0377-2217(03)00002-X. URL https://www.sciencedirect.com/science/article/pii/S037722170300002X.

I. Ballarini, S. P. Corgnati, and V. Corrado. Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project. *Energy*

Policy, 68:273–284, May 2014. ISSN 0301-4215. doi: 10.1016/j.enpol.2014.01.027. URL https://www.sciencedirect.com/science/article/pii/S0301421514000329.

F. Cappelletti, T. D. Mora, F. Peron, P. Romagnoni, and P. Ruggeri. Building Renovation: Which Kind of Guidelines could be Proposed for Policy Makers and Professional Owners? *Energy Procedia*, 78:2366–2371, Nov. 2015. ISSN 1876-6102. doi: 10.1016/j.egypro.2015.11.189. URL https://www.sciencedirect.com/science/article/pii/S1876610215019219.

P. Caputo and G. Pasetti. Overcoming the inertia of building energy retrofit at municipal level: The Italian challenge. *Sustainable Cities and Society*, 15:120–134, July 2015. ISSN 2210-6707. doi: 10.1016/j.scs.2015.01.001. URL http://www.sciencedirect.com/science/article/pii/S2210670715000025.

A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay. State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, 23:272–288, July 2013. ISSN 1364-0321. doi: 10.1016/j.rser.2013.03.004. URL https://www.sciencedirect.com/science/article/pii/S1364032113001536.

N. Fumo. A review on the basics of building energy estimation. *Renewable and Sustainable Energy Reviews*, 31:53–60, Mar. 2014. ISSN 1364-0321. doi: 10.1016/j.rser.2013.11.040. URL https://www.sciencedirect.com/science/article/pii/S1364032113007892.

R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal*, 7(2):19–33, Dec. 2015. URL https://hal.archives-ouvertes.fr/hal-01251924.

S. Hrabovszky-Horváth, T. Pálvölgyi, T. Csoknyai, and A. Talamon. Generalized residential building typology for urban climate change mitigation and adaptation strategies: The case of Hungary. *Energy and Buildings*, 62:475–485, July 2013. ISSN 0378-7788. doi: 10.1016/j.enbuild.2013.03.011. URL https://www.sciencedirect.com/science/article/pii/S0378778813001795.

T. Johansson, T. Olofsson, and M. Mangold. Development of an energy atlas for renovation of the multifamily building stock in Sweden. *Applied Energy*, 203:723–736, Oct. 2017. ISSN 0306-2619. doi: 10.1016/j.apenergy.2017.06.027. URL http://www.sciencedirect.com/science/article/pii/S0306261917307857.

D. Khafaga, E.-S. El-kenawy, A. Alhussan, and M. Eid. Forecasting Energy Consumption Using a Novel Hybrid Dipper Throated Optimization and Stochastic Fractal Search Algorithm. *Intelligent Automation & Soft Computing*, 37(2):2117–2132, 2023. ISSN 1079-8587, 2326-005X. doi: 10.32604/iasc.2023.038811. URL https://www.techscience.com/iasc/v37n2/53228. Publisher: Tech Science Press.

M. Mailagaha Kumbure, P. Luukka, and M. Collan. A new fuzzy k-nearest neigh-

bor classifier based on the Bonferroni mean. *Pattern Recognition Letters*, 140:172–178, Dec. 2020. ISSN 0167-8655. doi: 10.1016/j.patrec.2020.10.005. URL https://www.sciencedirect.com/science/article/pii/S0167865520303792.

A. Mastrucci, P. Pérez-López, E. Benetto, U. Leopold, and I. Blanc. Global sensitivity analysis as a support for the generation of simplified building stock energy models. *Energy and Buildings*, 149:368–383, Aug. 2017. ISSN 0378-7788. doi: 10.1016/j.enbuild.2017.05.022. URL http://www.sciencedirect.com/science/article/pii/S0378778816320023.

C. Petersdorff, T. Boermans, J. Harnisch, O. Stobbe, S. Ullrich, and S. Wartmann. Cost-Effective Climate Protection in the EU Building Stock. Technical report, ECOFYS, Cologne, Germany, Feb. 2005. URL https://www.eurima.org/uploads/files/modules/articles/1577099802_ecofysIII_report_EN.pdf.

C. F. Reinhart and C. Cerezo Davila. Urban building energy modeling – A review of a nascent field. *Building and Environment*, 97:196–202, Feb. 2016. ISSN 0360-1323. doi: 10.1016/j.buildenv.2015.12.001. URL https://www.sciencedirect.com/science/article/pii/S0360132315003248.

D. Rullière and M. Grossouvre. A Joint Kriging Model with Application to Constrained Classification, Sept. 2023a. URL https://hal.science/hal-04208454.

D. Rullière and M. Grossouvre. On Multi-Output Kriging and Constrained Classification. In *Séminaire Statistique LMA*, Avignon (FR), France, Oct. 2023b. URL https://hal.science/hal-04227155.

E. H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, July 1969. ISSN 0019-9958. doi: 10.1016/S0019-9958(69)90591-9. URL https://www.sciencedirect.com/science/article/pii/S0019995869905919.

P. Schetelat, L. Lefort, and N. Delgado. Urban data imputation using multi-output multi-class classification. *Building to Buildings: Urban and Community Energy Modelling*, Nov. 2020.

M. Storck, S. Slabik, A. Hafner, and R. Herz. Towards Assessing Embodied Emissions in Existing Buildings LCA—Comparison of Continuing Use, Energetic Refurbishment versus Demolition and New Construction. *Sustainability*, 15(18):13981, Jan. 2023. ISSN 2071-1050. doi: 10.3390/su151813981. URL https://www.mdpi.com/2071-1050/15/18/13981. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

L. G. Swan and V. I. Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8):1819–1835, Oct. 2009. ISSN 1364-0321. doi: 10.1016/j.rser.2008.09.033. URL http://www.sciencedirect.com/science/article/pii/S1364032108001949.

P. Van de Maele. French know-how in the field of energy efficiency in buildings; Le savoir-faire francais dans le domaine de l'efficacite energetique des batiments. Sept. 2010. URL https://www.osti.gov/etdeweb/biblio/22777397.

H. Wang and Z. J. Zhai. Advances in building simulation and computational techniques: A review between 1987 and 2014. *Energy and Buildings*, 128:319–335, Sept. 2016. ISSN 0378-7788. doi: 10.1016/j.enbuild.2016.06.080. URL https://www.sciencedirect.com/science/article/pii/S0378778816305692.

G. F. A. Yeo and V. Aksakalli. A stochastic approximation approach to simultaneous feature weighting and selection for nearest neighbour learners. *Expert Systems with Applications*, 185:115671, Dec. 2021. ISSN 0957-4174. doi: 10.1016/j.eswa.2021.115671. URL https://www.sciencedirect.com/science/article/pii/S0957417421010605.

Y. Zhang, B. K. Teoh, M. Wu, J. Chen, and L. Zhang. Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence. *Energy*, 262:125468, Jan. 2023. ISSN 0360-5442. doi: 10.1016/j.energy.2022.125468. URL https://www.sciencedirect.com/science/article/pii/S0360544222023507.

A Dictionnary of main variables

Variables are presented in the order in which they were selected by the KNN algorithm.

altitude	Altitude of the building.
chaufnum	Type of heating system for the dwelling/building.
detentmajnum	Most frequent level of maintenance in the building, ranging from 1 to 5.
energienum	Main source of energy for heating.
entbadpct	Percentage of dwellings with bad level of maintenance (4 or 5 out of 5).
entokpct	Percentage of dwellings with good level of maintenance (1, 2, or 3 out of 5).
hlmpct	Number of dwellings in the building that are under the social housing system.
jannatmax	Year of construction of the most recent part of the building/dwelling.
jannatmin	Year of construction of the oldest part of the building/dwelling.
lat	Latitude of the building.
lon	Longitude of the building.
murnum	Main walls material.
mutapct	Percentage of dwellings in the building that have been sold in the last year.
nc6	Number of dwellings with comfort level 6 on a scale ranging from 1 to 8, from good to bad.
nc7	Number of dwellings with comfort level 7 on a scale ranging from 1 to 8, from good to bad.
nent0	Number of dwellings with unknown level of maintenance. Maintenance levels range from 1 to 5, from good to bad.
nivtot	Number of floors in the building.
nlogh	Number of dwellings in the building.
nloghvac2a	Number of dwellings in the building that have been vacant for at least 2 years.
nprop	Number of owners that occupy their dwelling in the building.
nt1	Number of dwellings in the building comprising 1 rooms and a bathroom.
nt2	Number of dwellings in the building comprising 2 rooms and a bathroom.
nt3	Number of dwellings in the building comprising 3 rooms and a bathroom.
nt4	Number of dwellings in the building comprising 4 rooms and a bathroom.
pgaz01	Availability of city gas connection in the building/dwelling.
regth	Thermal regulation at the time of construction.
regth_1948	Boolean indicating that the building construction started before 1948.
surfacemoy	Average surface area of the dwellings in the building.
toitnum	Roof material.
typobatinum	Type of occupation: only housing or also with some professional activities.
typropnum	Typology of dwellings' owners: only one private owner, multiple private owners, miw of private and public owners, only public owners.

B Details of selected variables per model

Variables selected by FKNN: energienum, jannatmax, altitude, regth, surfacemoy, nt2, hlmpct, nt3, regth_1948, jannatmin, nloghvac2a, nc6, chaufnum, nent0, toitnum, entokpct, nt4, nivtot, nprop, murnum, pgaz01, nlogh, nt1, entbadpct, detentmajnum, nc7, mutapct, typobatinum, typropnum.

Variables selected by Joint Kriging: lat, lon, jannatmax, jannatmin, energienum, chauffagenum, nlogh, entbadpct, nprop.

Variables selected by Random Forest: jannatmin, energienum, surfacemoy, lon, lat, nprop, nivtot, hlmpct, detentmajnum.

C Confusion matrices

True values	Predicted values							
	A	B	C	D	E	F	G	
A	648.7	320.2	288.2	297.2	217.4	150.6	104.8	2027
B	249.2	491	361.2	283.7	183.1	121.4	82.4	1772
C	667.6	844.5	3262.9	2236.1	1228.9	785.9	512	9538
D	1097.8	1100.5	3018.4	5255	3098.8	1873.3	1266.2	16710
E	750.7	702.3	1527	2560.4	2947.7	1764.3	1211.5	11464
F	338.7	304.9	637.9	1071.7	1224.5	1092.6	801.8	5472
G	231.4	185	348.8	554.7	616.3	569	511.9	3017
	3984.1	3948.4	9444.3	12258.8	9516.7	6357.1	4490.6	50000

Table 5: Confusion matrix of the Fuzzy KNN model.

True values	Predicted values							
	A	B	C	D	E	F	G	
A	933	189	231	343	237	64	30	2027
B	225	542	447	368	144	24	22	1772
C	205	348	4353	3227	1028	255	122	9538
D	219	184	2671	8558	3623	982	473	16710
E	134	79	890	3978	4312	1415	656	11464
F	53	27	306	1350	1849	1230	657	5472
G	49	15	148	676	918	689	522	3017
	1818	1384	9046	18500	12111	4659	2482	50000

Table 6: Confusion matrix of the binarized KNN model, used for hard classification.

True values	Predicted values							
	A	B	C	D	E	F	G	
A	179.8	50.6	3.4	22.1	-5.5	-23.1	-21.2	206
B	45.5	71.1	25.5	30.7	8.5	-8.4	-10.9	162
C	53.5	114	454.7	375.3	173.8	14.2	-39.5	1146
D	-5	24.2	417.6	885.9	523	175.2	60.1	2081
E	-44	-49.9	159.8	500.1	447.6	228.9	159.4	1402
F	-14.4	-28.2	63.5	182.9	169.1	149.6	100.5	623
G	-9.3	-19.8	21.4	83.9	85.6	86.6	131.6	380
	206	162	1146	2081	1402	623	380	6000

Table 7: Confusion matrix of the Fuzzy Joint Kriging model.

True values	Predicted values							
	A	B	C	D	E	F	G	
A	120	18	4	27	10	11	16	206
B	39	56	17	32	7	3	8	162
C	89	86	423	360	103	35	50	1146
D	96	74	262	1092	284	100	173	2081
E	33	17	51	607	340	147	207	1402
F	16	5	15	233	109	137	108	623
G	3	7	5	113	44	56	152	380
	396	263	777	2464	897	489	714	6000

Table 8: Confusion matrix of the binarized Joint Kriging model, used for hard classification.

True values	Predicted values							
	A	B	C	D	E	F	G	
A	147	36	4	6	10	7	14	224
B	27	115	20	8	15	2	4	191
C	56	173	513	192	131	64	58	1187
D	109	111	340	610	430	240	180	2020
E	67	48	69	221	436	284	249	1374
F	17	9	13	59	140	211	181	630
G	25	4	6	22	46	80	191	374
	448	496	965	1118	1208	888	877	6000

Table 9: Confusion matrix of Random Forest hard classification.

Notations

$\mathbb{M}$ A fuzzy classfication matrix, with one row per individual and one column per class. Each row should typically sum to 1. And for membership degrees, all elements are positive.

$\mathcal{C}_{\mathbb{M},\hat{\mathbb{M}}}$ A confusion matrix comparing true classifications with predicted classifications generated by a fuzzy classification algorithm.

$\mathbb{1}_n$ A column vector of n ones.

$\mathrm{diag}[]$ Diagonal of a matrix.

Acc Accuracy.

BA Balanced accuracy.