A framework to evaluate the reliability of energy performance assessment methods using buildings databases, application to the French Residential Energy Performance Certificate

Nicolas Bourbon\textsuperscript{1,2}, Thomas Recht\textsuperscript{2}, Rofaida Lahrech\textsuperscript{1}, Laurent Mora\textsuperscript{2}

\textsuperscript{1}CSTB - Centre scientifique et technique du bâtiment, Champs sur Marne, France
\textsuperscript{2}University of Bordeaux, CNRS, Arts et Metiers Institute of Technology, Bordeaux INP, INRAE, I2M Bordeaux, F-33400 Talence, France

Abstract
The reliability of Assessment Methods of Energy Performance in Housing (AMEPH) is a major stake to boost energy-efficient renovation. Energy Performance Certificates (EPCs) represent a relevant instrument supposed to be strongly structuring for the assessment of buildings energy performance and decision support both for public and private actor. However, a lack of reliability in the methods used by the EPC is reported in several countries depending on the assessor, input data and calculation tools, etc. The French EPC tends to show an overestimation, a significant dispersion and a lack of repeatability in consumption calculations, which has lead public authorities and users to demand improvements in its reliability. Various studies have been carried out on previous calculation method but most of these studies do not make for an accurate representation of the diversity of the building stock. In this article, we propose a framework to evaluate the reliability of AMEPHS which uses recent housing databases containing a few hundred to several thousand elements. We then present the adaptation of the framework and all the materials needed to apply this framework to the French EPC calculation method.

Key Innovations
- Proposition of a framework to evaluate the reliability of energy performance assessing methods for residential buildings by exploiting buildings databases ensuring the diversity of residential buildings.
- Implementation to the French residential EPC.

Practical Implications
The proposed framework is intended to be replicable with other AMEPHS, particularly those under the EU EPC directive. Applying this approach to regulatory methods could make it possible to advise public authorities and propose solutions to make AMEPHS more reliable.

The application of this framework to the actual 3CL method of the French EPC is finalized in order to validate the approach in a defined context. This framework could be applied by public authorities to improve the reliability of the EPC scheme at each occasion of its overhaul.

Introduction
Greenhouse gas (GHG) emissions by humans are responsible for climate change, which threatens humanity and biodiversity. The building sector is one of the main contributors and therefore one of the fields with most potential for reducing energy consumption and GHG emissions. To do so, the Energy Performance of Buildings Directive (EPBD) was introduced in 2002 with the obligation to implement national legislation creating an Energy Performance Certificate (EPC). The EPC has been enshrined in French law since 2006 by the “Diagnostic de Performance Énergétique” (DPE), and it establishes a legislative framework to produce a unified energy performance document by setting out the methods for evaluating this performance.

In France, 70\% of the expected housing stock for 2050 is supposedly already built, according to Trainsel (2010). The average level of consumption (heating, cooling, DHW, auxiliary and lighting) of the French housing stock is 186 kWh/(m\(^2\).year) of primary energy in 2012 (ADEME 2018), whereas the target set by law for 2050 is 80 kWh/(m\(^2\).year) of primary energy. These findings are similar across European countries (without being identical), it is therefore essential to increase the scale of energy efficient renovations. To refurbish the stock efficiently, it is useful to be able to assess the initial performance and the performance gains after renovation of existing buildings. Then, it is necessary to ensure the reliability of the methods for assessing the energy performance of housing (AMEPH) for existing dwellings.

The aim of the article is to propose a comprehensive approach to assess the reliability of AMEPHS, particularly in the context of the European EPC Directive. First of all, it is necessary to consider the complexity of defining and assessing the reliability of a AMEPH (problem statement) and then to develop a framework enabling its assessment through appropriate choices of studies, data bases and uncertainties (methods). Finally, the adaptation and implementation process of the approach to an AMEPH will be illustrated using the method of Calculating Conventional Housing Consumption (3CL) associated with the French EPC regulation.

Problem statement
Assessing the energy performance of a dwelling is a difficult exercise, due to the complexity of the “building” system, the diversity of the notion of performance, the multiplicity of assessment methods or the presence of uncertainties, all of which taint the final result with errors.

Housing: a complex system
A building is a complex system because it is a dynamic, multifactorial system, with a great diversity in design, whose behavior results from the interaction between the envelope, energy installations, the environment and occupants. The envelope - all the walls, floors and ceilings - and the energy installations - all the systems that produce, transform, store and distribute energy - are the two components of a building. It is subject to external environmental solicitations - temperatures, radiation, etc. - as well as internal solicitations from occupants - behavior, presence, etc. - see Figure 1. The energy consumption of a housing unit is therefore dependent on the characteristics of these components, but is also highly dependent on the solicitations that vary according to the spatial (geographical position, etc.) and temporal (time of year/day, etc.) context.

![Figure 1: Interactions generating energy consumption in housing - separation according to energy use](image)

In a housing unit, consumption is linked to comfort and health needs and can be separated into six categories: heating, air conditioning, domestic hot water (DHW), lighting, auxiliary equipment (equipment to assist in meeting needs: fans, pumps, etc.), and specific uses (in the case of housing, this comprises all domestic and entertainment appliances) - see Figure 1. If renewable energy is produced on site, production can be higher than consumption - see Figure 1, which can improve the energy balance of the unit. Legally, the final energy consumption is converted into primary energy, the coefficients of which are a function of the energy vector (environment). Finally, the method for obtaining the data is chosen by the entity that wishes to characterize the performance of the housing, taking into account the expectations set out above. This also requires detailing the characteristics of the components to be collected and the methods for obtaining this data. The notion of performance is also used to compare the housing units. A common structure (range of uses, conventions on needs, etc.) is then necessary. In addition, the spatio-temporal context surrounding the housing unit has an impact on the level of performance. Indeed, external solicitations present hourly, seasonal or annual variations (climate change), but also regional or even local variations. In the same way, standard behavior tends to vary over time: the notion of comfort differs depending on the occupants, and from one country and period to another. In addition, building components are often specific to the local context but are also subject to ageing and deterioration. Finally, if there is a desire to communicate the resulting performance to a non-expert public or to exploit it for legal purposes (financial incentives, sanctions, etc.), it is important to generate indicators that are understandable, appropriate and reliable. The reliability of these indicators must therefore be evaluated.

**The multiplicity of AMEPHs**

In view of the above-mentioned criteria, the great diversity of AMEPHs is due to the complexity of the building system and the diversity in the notion of performance. They are nevertheless represented according to a common principle of assessment, illustrated by the block diagram in Figure 2, even if many disparities exist. Depending on the methods, performance is not always qualified for the same number of energy needs. In France, three energy needs are considered for the 3CL method of the 2012 French EPC (heating, DHW, air-conditioning), five energy needs for the Th-BCE method of the 2012 Thermal Regulations and all energy needs for the EPC invoice method.

![Figure 2: Principle in a block-diagramm and sources of errors in AMEPH](image)
and on the physical parameters used in the chosen modeling. The diversity of data collection strategies contributes to the diversity of available AMEPHs (see examples in Table 1).

Table 1: Examples of diversity in data collection strategies for an AMEPH

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Measure</th>
<th>Computing</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Envelope (surface)</td>
<td>Tape</td>
<td>Using architectural plan</td>
<td>Statistics on dwelling surface and ceiling height</td>
</tr>
<tr>
<td>Installation (air flow)</td>
<td>Flowmeter</td>
<td>Mandatory air flow</td>
<td>Estimation using dwelling volume</td>
</tr>
<tr>
<td>Environment (air temperature)</td>
<td>Thermometer</td>
<td>Numerical model of local climate</td>
<td>Heating degree days (HDD) using 30 years mean value</td>
</tr>
<tr>
<td>Occupants (Water volume)</td>
<td>Water meter</td>
<td>Number of occupant, age, etc.</td>
<td>Statistics using dwelling surface</td>
</tr>
</tbody>
</table>

Taking into account a temporal dependence of the parameters - whether it be the envelope, energy installations or solicitations - and/or a temporal scheme in the model allows for a characterization of the AMEPH as a dynamic method (i.e. in France Th-BCE method of the RT2012, STD software). When the parameters do not adapt to the temporal evolution, the method is then called static (i.e. in France the 3CL-method of the EPC). Intermediate methods may exist. The choice of modelling as well as the collection of associated parameters strongly depends on the building stage: design, construction, exploitation, or destruction. At the design stage, it is fairly easy to collect the numerous characteristics of the envelope and the installations because these are then selected by the designer, but this does not mean that the data is any more reliable. On the other hand, when the building is already constructed, the data on the characteristics of a component is more difficult to collect because of deterioration due to ageing. The measurement of characteristics also becomes technically more difficult - especially for envelope elements - because the measurement must be non-destructive. The choice of the principle of data collection for AMEPHs therefore implies the integration of technical constraints.

The sources of errors in an AMEPH

The methods exploit different strategies for collecting housing data, as mentioned above, and so they necessarily involve measurement errors, approximations in the models used to calculate them, or biases (statistical, for instance) in the proposed estimates. Uncertainties are therefore associated with housing data. In addition, the spatial, temporal and behavioural variability of internal and external solicitations also involves uncertainties. Finally, the models make choices and simplifications of reality; this is referred to as model biases. All these sources of error are shown in Figure 2. Thus, in order to assess the degree of confidence in an AMEPH, it is necessary to identify all existing sources of error and then to evaluate their impact. In particular, the reliability of an AMEPH must be considered in light of the wide diversity of buildings in terms of diversity of characteristics.

In order to support public policies, and in view of the need to define reliability for AMEPHs, we wish to propose an adaptable and coherent evaluation framework. The EPBD imposes a broad regulatory base for EPCs but leaves it to each state to develop its own AMEPH. Public authorities, particularly in France, encounter difficulties regarding the reliability of their method. The similarity of the problems as well as a common regulatory base leads us to propose a framework for the evaluation of reliability which is more particularly adapted to EPCs methods

Methods

Reliability: a vague and multifaceted notion

Many definitions of reliability exist and can be summarized as the ability of an entity (physical system, method, theory) to perform its required purpose under fixed conditions. Because of the multiplicity of AMEPHs as well as the diversity of actors involved, the notion of reliability is very versatile. Reliability can be defined as the confidence of interested parties in the entity under study - in our case, the energy performance assessed. This perceived reliability then depends on the actor's point of view. Thus for the occupants of a dwelling, reliability is mostly associated with the reference value of interest, which may be the actual consumption or the energy bill. They want the energy performance obtained by calculation to be close to their actual consumption. For developers of an AMEPH, reliability can also be questioned by assessing its feasibility by the people using it (assessors for EPC). Reliability for these creators may also imply the ability to tamper with the AMEPH in order to improve the perceived resulting performance. As for the actors and experts in the building industry as well as the creators of the AMEPH, technical reliability is expected (a detail of this notion is provided below). As far as the public entity carrying out the method is concerned, the objectives associated with the method are not only technical, but they also aim for awareness (informing the population), incentives (i.e. financial) or regulations (thresholds to be reached). The reliability of the assessment method can then be evaluated in terms of its ability to meet these objectives - as mentioned in the proposed definition.

The definition of technical reliability can be understood as the quantification of the ability to correctly assess energy performance under given conditions (that could be standard). This requires defining the variable of interest used to compute reliability indicators: is it the final consumption, the performance indicators, intermediate values (i.e. quantity by type of needs), or else? Once the variable(s) of interest has been chosen, we define the reliability of an estimation method following the ISO 5725-1 standard on measurement methods. Thus, reliability results from the evaluation of trueness, precision and repeatability.

A definition for each of the terms is proposed to qualify the reliability of AMEPHs. Trueness is "the closeness of
agreement between the average value obtained from a large series of test results and an accepted reference value". The indicator generated therefore corresponds to the value of the deviation between the mean value and the reference value. This means that a reference value must be defined. Precision is "the closeness of agreement between independent test results obtained under stipulated conditions", so it is necessary to quantify the standard deviation of the many values obtained using the method. Repeatability is "precision under repeatability conditions", which is to say that results are "obtained with the same method on identical test items". This means that for a chosen method, with identical housing units, a repeatability deviation is quantified. If a lack of reliability is found, the causes for this lack of reliability must then be identified.

Within the framework of perceived reliability, the responsibilities may be more diffuse because they depend, on one hand:

- on the subjectivity of the actors who consciously or unconsciously set their own confidence threshold;
- on the behaviour of the actors: many elements of this perceived reliability are linked to the psychological or socio-economic behaviour of the different actors;
- on enlightened communication: relevance of indicators, education and explanation of associated physical phenomena.

As for technical reliability, responsibilities can be assigned to various main categories:

- Input data: values and/or methods of collection;
- Model: relevance of physical models;
- Solicitations (i.e. given / standard conditions): values / models allowing their quantification;

Because of this great diversity in the definition of reliability, many studies can be carried out in this way. We have chosen to propose a framework to assess technical reliability.

**Studies to assess technical reliability**

By evaluating technical reliability in terms of trueness, precision and repeatability, we generate indicators on the final performance obtained. However, it is necessary to identify the causes of unreliability, with these indicators and by means of appropriate studies.

Trueness is then understood as the quantification of the deviation between the average performance obtained by the calculation method and a reference value. The reference value must be obtained under the same conditions, which means that the consumptions studied must be studied over the same period, according to similar solicitations and taking into consideration the same energy needs. Thus, this approach allows us to evaluate the selected model with the ultimate goal of validating the model. A detail of the methods, constraints and reference values is provided in Table 2. In the case of an inter-comparison method, the principle of evaluation is illustrated in Figure 3. The choice of solicitations requires special attention to ensure that they remain identical across both methods, so that the comparison is carried out on the models only without the influence of solicitations.

**Table 2: Methods of trueness evaluation**

<table>
<thead>
<tr>
<th>Methods</th>
<th>Reference value</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical</td>
<td>Actual consumption (measured one)</td>
<td>All energy needs must be modeled. Actual (measured) solicitations must be used. Uncertainties of collected data must be characterized.</td>
</tr>
<tr>
<td>Numerical by inter-comparison</td>
<td>Estimated consumption by an AMEPH of reference</td>
<td>AMEPH of reference must be at least true. Solicitations must be identical. Uncertainties of collected data must be similar.</td>
</tr>
</tbody>
</table>

Precision is the quantification of the standard deviation after many values obtained by the AMEPH on similar systems. It therefore amounts to carrying out uncertainty analyses (see Figure 4). By taking into account the uncertainties regarding the input data, one can thus assess the relevance of the model and start to quantify the relevance of the input data retained - and their associated uncertainties. By taking into account the variability of solicitations, the consequences of the choice of solicitations can be assessed. To enrich the results, a sensitivity analysis can be conducted to identify the data responsible for the unreliability in each case. In addition, conducting the same study protocol on the same method used to evaluate trueness by inter-comparison, allows to identify and evaluate differences in terms of overall precision between methods. Many approaches exist for uncertainty analysis and sensitivity analyses (UASA), the choice depending on computation time constraints, among other factors. The advantage of probabilistic methods is that they allow for an uncertainty analysis and a sensitivity analysis to be carried out at the same time (Saltelli et al. 2000). A sensitivity analysis enables the most influential parameters to be identified by quantifying the share of uncertainty in the final result due to each of the parameters (Figure 4). The repeatability of a method is evaluated by quantifying the dispersion of the results obtained for an identical case. In the context of a performance measurement method, this means carrying out measurements of consumption on a dwelling over an identical period and erasing disparities in solicitations.

![Figure 3: Principle of the inter-comparison method to evaluate AMEPH trueness](https://doi.org/10.26868/25222708.2021.30786)
on the condition that the data is previously known in full. It thus allows for the identification of the technical difficulties inherent in obtaining the data from the dwelling (and possibly the solicitations). If a large number of dwellings is studied, trends in agents’ behaviors can also be identified in the way that they collect data.

Figure 4: Principle of UASA
Assessing reliability in terms of trueness, precision and repeatability is therefore not enough. It is necessary to set up specific studies using the calculation of indicators associated with these notions to validate and/or identify the share of responsibility of the different sources of error.

The use of building databases
Housing and residential buildings show great diversity due to architectural and construction arrangements that depend on the standards, time periods, regulations and location in the local territory. In addition, the systems installed also present great technological and performance diversity. Therefore, in order to evaluate the reliability of a method, it is of primary importance to conduct these studies on a wide variety of buildings. In the first place, one must identify discriminating characteristics that differ between dwellings and which then make it possible to reflect the diversity of the dwellings. For this purpose, a characterization of the housing stock on which the AMEPH will be used is strongly recommended. Among the discriminating characteristics, the following are the most common: the type of housing, the energy vector used, the construction date, the climate zone or, a little less conventionally, the "true" (meaning actual) performance level of the housing. Ensuring diversity therefore requires having housing in each of the subcategories (or typologies) that can be created. Under these conditions, the use of housing databases becomes relevant. The databases used can include real dwellings or hypothetical dwellings created for that purpose.

If the stock studied is characterized and the distribution according to typologies is known, it is interesting to have a database of dwellings that is representative of the stock. In fact, the conclusions of reliability studies can be used to provide trends according to the typologies and above all to aggregate error indicators or to prioritize reliability actions depending on the importance of certain typologies within the housing stock. In any case, if representativeness is sought, it is very important to ensure it, particularly with specific organizations.

Characterization of uncertainties and their diversity
The data exploited, whether for components or stresses, can exist under three different formats: continuous, discrete or categorical. Associating uncertainties can be more or less difficult depending on the format. Determining an uncertainty on categorical variables is particularly difficult. The realization of UASA must therefore take these problems into account.

First of all, an analysis of the data set must be carried out. First by defining the certainty or otherwise of the data (e.g. the type of habitat can be considered certain), then the format of the associated data must be identified, and finally the uncertainty associated with the data must be evaluated. If the parameters are measured, then the uncertainty is that of the measurement. If the parameters are evaluated (either by calculation or by estimation) it is then necessary to establish the associated uncertainty before anything else, by studying the model used for the calculation or estimation. Finally, as a last resort, the uncertainty can be provided by means of a literature review and, if not, by expertise. Given that some data is interrelated or redundant, it is important to ensure that there are no conflicts in the uncertainties provided. Uncertainty distributions may also vary depending on the value of the parameter itself, or on the discriminating characteristics of the buildings. For example, the uncertainties regarding the characteristics of materials in older dwellings are more important than in recent buildings because of ageing and the lack of traceability of the quality of the materials used.

Particular attention must be paid to solicitations, because although they are not part of the intrinsic characteristics of housing, they have a significant impact on the discrepancies between calculated and actual consumption (Laurent 2013 for the case of a few European EPCs). However, an AMEPH wishes to calculate the consumption that is closest to the actual consumption. As has already been evoked, a great variability of solicitations is noted in the literature. Concerning external solicitations, in spite of a strong variability and an evolution due to global warming, the collection methods are well defined and the uncertainties are therefore known. On the other hand, internal solicitations depend strongly on occupant behavior and their evaluation requires either numerous hypotheses whose relevance is questionable, or an unacceptable invasion of privacy. These latter are therefore often conventions derived from statistical information, and the associated variabilities are difficult to quantify even if studies provide answers.

Summary: A framework for assessing the reliability of AMEPHs using a multi-study design
If we summarize our proposals, the framework for assessing the reliability of AMEPHs is structured into four main stages, which are described in Figure 5. The stages, while requiring an order (as in Figure 5), may require a looping process. Indeed, it is necessary to adapt the studies to the limitations of the data stored in the databases as well as to the constraints imposed by the format of the data and the quality of the uncertainty laws. Adapting the approach and defining the studies to be carried out is therefore the last stage. The results of some of the studies enables to reduce the technical complexity of others (e.g. the identification of influential parameters limits the time spent studying the problems of study repeatability). The phasing of the studies must therefore
be reflected upon, as well as the phasing in the use and exploitation of the databases, in particular to limit the calculation time. The studies will always be carried out initially on test cases or small database extractions before being applied to the databases in their entirety.

**Figure 5: Stages of the proposed framework**

In table 3, the main studies that can be carried out according to the causes of potential unreliability are referenced. The constraints that should ideally be met are detailed. The proposed framework is intended to be adaptable to many AMEPHs, and in particular to that associated with the European EPC regulation.

**Table 3: Proposition of potential studies according to sources of errors causing a lack of reliability.**

<table>
<thead>
<tr>
<th>Sources of errors</th>
<th>Principle of the study</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Trueness evaluation (inter-comparison)</td>
<td>Data for components and solicitations should be identical</td>
</tr>
<tr>
<td>Collected Data</td>
<td>Precision evaluation and most influential parameters (UASA + inter-comparison)</td>
<td>Qualified uncertainties should be similar for identical parameters</td>
</tr>
<tr>
<td>Solicitations</td>
<td>Evaluation of precision (UASA + inter-comparison)</td>
<td>Similar variability, identical model</td>
</tr>
<tr>
<td>Conception Defaults</td>
<td>Repeatability evaluation (in situ, or theoretical analysis)</td>
<td>Unique housing. If by measurement, ensure similar solicitations</td>
</tr>
</tbody>
</table>

**Results**

We present the adaptation of the framework to the 3CL-method. The application of the approach for each stage is detailed but the execution is not developed.

**The unreliability of the French EPC**

The choice of the French EPC is because:
- it concerns a very large majority of the national stock: it is mandatory in cases of sale or rental
- it has an economic impact: the appearance of a green value on the real estate market is observed
- it has a political impact: it is a public tool that provides information on the quality of the national stock and can be used to boost renovation work
- it is informative and educational: it addresses the issue of carbon energy for the population through labels. However, it is a tool that lacks perceived reliability. First, it is not well understood by occupants. Public authorities and experts regularly criticize it - there have already been two reliability actions in 2012 and 2020. In terms of technical reliability, many problems are noted. With regards to repeatability, a survey by Chesnais and Bourcier (2017) found a flagrant lack of repeatability in some buildings. Raynaud et al. (2011) had already addressed this finding of a design flaw in the previous method which resulted in a lack of repeatability due to the diversity of methods of collection for the parameter U. As for Laurent et al. (2013), a quasi-systematic over-estimation of consumption is observed when comparing it with actual consumption, which raises questions about the validity of the model used. Finally, Raynaud et al. (2011) and Osso et al. (2019) observe a significant dispersion of the result on two versions of the DPE calculation method, by propagating uncertainties on the envelope data but only on a few theoretical test dwellings. This raises questions about the quality of the input data and model, and thus about the reliability of the method.

Although each of the reliability indicators have been mentioned and some causes of the lack of reliability are addressed by these actors, it is clear that there are limitations to these studies. Indeed, they were conducted on two versions of the DPE computational method (the previous and current one) or on a limited number of test cases (a little less than ten). The application of our framework would make it possible to partially overcome these limitations and to potentially clarify and bring together the conclusions obtained.

**Stage 1: AMEPH Analysis**

The first stage consists in analysing the 3CL-method. For the French EPC, a method using invoice data coexists with the calculation method called 3CL-method. We are studying the latter. The description of this AMEPH is included in the JORF decree (2012). The 3CL-method is a static annual method determining the consumption associated with the three energy needs: heating, DHW, and air conditioning. The model of air conditioning is very limited and depends only on the climate zone, the type of energy vector and the air-conditioned surface. The DHW is only determined by a characterization of the systems, the needs being obtained conventionally. The heating model is much more detailed. However, the dynamics of the buildings is only taken into account through a reduction of the impact of the internal and external contributions by a non-linear law resulting from the characterization of inertia according to four categories. All the solicitations data are annualized and defined by conventions. The spatial variation of external solicitations is partially taken into account (different values depending on the department and altitude). For internal solicitations, a single scenario of occupation and temperature is considered, the quantity of hot water is derived from statistical values of consumption depending solely on the surface and type of habitat.

The data collected (DC) for housing components is obtained either by measurement (rare), or evaluated by calculation or estimation (a little less rare) or simply described, and is therefore categorical data format (in majority). The DCs undergo a data treatment to assign...
values to the physical parameters necessary for the models (Intermediate Data – DI) using tabulated values set by regulation (JORF 2012). There are certain DIs for which several methods of collection are possible. This is the case in particular for the parameter U, which exacerbates the risk of non-repeatability.

Finally, the indicators are generated from the conversion coefficients set by normative choices. A schematization of the 3CL-method is proposed in Figure 6.

Figure 6: Block diagram of the principle of 3CL-method
The 3CL-method is therefore an AMEPH in which all the studies seem necessary. Lack of reliability in terms of repeatability, trueness and precision seems to be expected. The causes are multifactorial: input data, models and solicitations are involved. The very design of the method seems to cause a risk of non-repeatability. However, since the solicitations are defined by conventions in their entirety with strong similarities with the thermal regulation in newly constructed housing (RT2012 and its Th-BCE computational method), a study of trueness by inter-comparison seems relevant.

Stage 2: Identification of exploitable databases
We have identified several databases presenting a satisfying diversity (or even representativeness) of housing. Moreover, the associated data is compatible with the implementation of the 3CL-method.

1. PHEBUS
The Survey of Housing Performance, Equipment, Energy Needs and Uses (PHEBUS in French) conducted in 2013 is a statistical survey that studies the energy performance of housing, occupants’ behavior and energy consumption. One of the objectives was to have a vision of the energy characteristics of the French housing stock in 2012. The dwellings are real and extracted from the master sample of the 2011 annual census survey which is representative of the distribution according to regions, climate zones, type and years of construction. A representativeness weight is thus assigned to each selected building. The determination of energy performance was carried out using the 3CL-method on 2369 dwellings and the data collected was retained. This database known as PHEBUS is representative of the residential stock in 2012. The practices of the assessors in this context are very specific; indeed, the level of scientific expectation was high and only one company provided the EPCs (limiting the economic constraints). The analysis of data collection practices by assessors is then potentially biased, limiting the analysis of technical defects in the method design.

2. EPC Observatory
All of the EPCs carried out are referenced in this database, i.e. several million housing units. Given that the collected data is stored (although partially), the assessors’ practices in a real context can be characterized. Moreover, it has the potential to identify technical flaws in the design of the method: technical unfeasibility and impact of the diversity in methods of collection for a parameter on a large scale.

3. CSTB database
One of the databases of the “Centre Scientifique et Techniques du Bâtiment” (CSTB) is a grouping of about a hundred theoretical housing units. These dwellings are modelled using the PHEBUS database. Based on distributions of certain parameters deemed to be discriminating and the identification of trend values, typologies of theoretical dwellings have been created to represent this diversity. The level of data was consolidated to reach a higher level of information to allow calculation using the Th-BCE method. Thus this base will be useful to assess the trueness of the 3CL-method and especially the validity of the model.

4. Energy Performance Observatory (OPE)
The dwellings of the previous bases offer a great diversity of buildings but do not take into account dwellings built after 2012. The RT2012 imposes a very low level of consumption for new housing: 50 kWh/m².(a). The OPE database references tens of thousands of dwellings built under the RT2012 regulation and presents more complete data than for an EPC, but compatible data. All reliability studies can thus be evaluated on housing with high energy performance. In particular, the capacity of the 3CL-method’s models to correctly represent the physics of high-performance housing can be evaluated.

Stage 3: Characterization of uncertainties
DCs for the 3CL-method are mostly categorical, especially for energy installations. As mentioned in the description of the method, the data processing uses tabulated values set by regulations to assign values to physical parameters (DIs). The uncertainties associated with the tabulated values are not given. By default, we decide to assign theoretical uncertainties to the DI values considering that the value resulting from the processing is the true value. This results in a loss of the causal link between the true uncertainties and the uncertainties used, which are the result of an arbitration between the bibliography (Osso et al. 2019, Mastrucci et al. 2017, Zhang et al. 2016) and the expertise. In particular, Mastrucci et al. (2017) detail the dependencies between the uncertainties of certain parameters.

Similarly, conventional values for solicitations are not associated with any uncertainty. For external solicitations, an estimate of the variability can be calculated from the method used to obtain the climate data (annual average over 30 specific years of measurement of this data). However, it seems much more complicated to characterize their variability of internal solicitations.

Synthesis : Adaptation of the framework
We have selected four studies to examine the reliability of French EPC by using the databases mentioned, following the phasing below. In view of the difficulties in qualifying the variability of the parameters related to solicitations,
we will not deal with this problem here. To evaluate the validity of the 3CL-method, we choose to quantify trueness by inter-comparison with the Th-BCE method, whose reliability has been checked by the Standard 140-2001 (ANSI/ASHRAE 2001). It will be applied based on CSTB and an extraction of the OPE. In order to quantify the dispersion of the final result and to attribute the responsibility for this dispersion to the most influential parameters, UASAs will be executed based on a Morris method. This method will be applied on the PHEBUS and CSTB bases. To evaluate the repeatability of the method in a theoretical way, we study the trueness bias due to the different collection methods on the U and the nominal power of combustion generators. This study is also conducted on the CSTB and PHEBUS bases, and possibly on the EPC Observatory. Finally, an analysis of data collection practices is carried out in order to assess the technical feasibility of the method. A statistical analysis of the distribution of data entry modes and the rate of identifiable errors is performed on the PHEBUS database and the EPC Observatory.

Conclusion

Part of the problem with the reliability of AMEPHs is the diversity and complexity of the different concepts involved. In this article defining these notions, we propose a framework to assess the reliability of AMEPHs, in particular computational and regulatory one. This framework is based on the use of housing databases that reflect the diversity of the housing stock to assess reliability. The viability of this approach is demonstrated with its adaptation to the French EPC. The next step is to carry out these studies to conclude on the reliability of the French EPC. The reproducibility of the adaptation of the global approach to other AMEPHs still remains theoretical for other computational methods, whether they are derived from EPC regulations or not, but is one of the major perspectives.

Acknowledgement

This work was supported by the CSTB through a Ph.D. thesis conducted with the University of Bordeaux (I2M Bordeaux).

Nomenclature

3CL: Calculation of Conventional Housing Consumption
AMEPH: Assessment Methods of Energy Performance in Housing
DPE: Diagnostic Performance Énergétique (French EPC)
DC: Collected Data
DI: Intermediate Data
DHW: Domestic Hot Water
EPC: Energy Performance Certificate
EU: European Union
GHG: Greenhouse Gas
U: Thermal transmission coefficient

UASA: Uncertainty Analysis and Sensitivity Analysis

References