A Residential Case Study To Validate A New Default Detection Method Based On Discrepancies Between Simulated and Measured Data

Rozenn Josse-Buret[1,2], Frederic Wurtz[1], Philippe Marchal[2], Etienne Wurtz[2]
[1]Univ. Grenoble Alpes, CNRS, Grenoble INP*, G2Elab, Grenoble, France
[2]Univ. Grenoble Alpes, CEA, LITEN, INES, LGEB, Grenoble, France

Abstract

Guarantee of performance (GP) is a big research challenge that we propose to address in this paper. Even if different guides have been developed, a need in automation of defaults detection in buildings is felt. In this paper, we propose a method helping the GP by detecting these defaults, using the method of Morris and the ABC-PMC bayesian algorithm. We apply it on an inhabited house, focusing the analysis on absence periods. We test and compare 2 techniques for absence detection. Finally we study the impact of simulation on short periods on the Morris analysis.

Introduction

Nowadays, we can notice that various policies are developped in order to decrease the impact of our consumptions on the environment. To reach this objective, it is striking to realize that buildings are the major lever for action. As a matter of fact, this sector represents 40% of primary energy consumption and a third of CO_2 emissions (European Commission DG ENV (2007)).
As a result, regulation is more and more demanding in terms of energy efficiency which lead to development of positive energy houses. Even though Rosenberg and Eley (2013), EVO (2012) both developed guides to help the GP, there is no clear method that guarantees the performance of this kind of dwellings. Meanwhile, houses are more and more equipped with monitoring systems, giving occupants information about their home.
Hence, in this paper, we propose a new method non-intrusive (i.e. that, once the dwelling is inhabited, we do not need to disturb the occupants anymore), based on data analysis focused on absence periods, which will help to identify what went wrong between the design phase (simulation) and the occupation phase.

A method for the guarantee of performance

What is the guarantee of performance (GP)?

Speaking of GP, a contract is often implied between the constructor and the client. This contract aims to verify that construction is made regarding certain conditions (Erhorn and Erhorn-Kluttig (2011)). Hence, the contract stands as a guarantee.

*Institute of Engineering Univ. Grenoble Alpes

However, in order to assess performances of a building, we need to be able detect what is wrong inside. Are there differences ? If yes, where does it come from? In this paper, we try to detect defaults and its origin, when differences are noticed.
This will help to verify if terms within the contract are respected or not, and thus the GP .

Overview and objectives of the method

Our main objective is to understand the differences between the simulation model and the measurements when comfort and energy consumption are studied. These differences are key-factors to the GP.

In order to do that, we need to understand defaults linked to the construction. As the study case is a smart-house, we choose to use data arriving directly from sensors already set up so that we avoid the useless accumulation of sensors.

To identify these defaults, we study the error between, one one hand, the simulated and measured heating energy and, on the other hand, the simulated and measured comfort level.
As a result, detecting the sensible parameters is the first step of our methodology. We look for sensitive factors on the error above calculated. Thus, these parameters are most likely the source of the observed differences. To achieve this part, we use the method of Morris presented in Morris (1991) and updated in Campolongo et al. (2007).

After that, we need to know the range of values where our sensible parameters should be so that the simulation output (energy consumption or comfort) matches the measurements. The following step of our methodology consists in applying a bayesian algorithm called ABCPMC (Approximation Bayesian Computation - Population Monte-Carlo) described in Beaumont et al. (2009).

Studying measurements of a building is mainly influenced by three factors: (i) physics, (ii) occupancy (and occupants' behaviour) and (iii) the weather. In our case, physics is approximated thanks to the software EnergyPlus. To minimize the impact of occupancy, we decide to focus on periods of time where inhabitants are absents. We principally analyse electrical data to identify these periods. Finally, to get over the influence of the weather, we measure the external temperature and the solar flux on site that we reintegrate in the EnergyPlus model.

These steps enable the identification of the defaults of the building compared to the initial model, using basic sensors, minimizing the impact of the occupants and taking into account the weather.
This identification aims to help the guarantee of performance by responding to these three questions: (i) what are the gaps of consumption and comfort level between the initial model and the final building, (ii) what parameters are responsible, (iii) what values should they take to match the measurements (consumption or comfort)?

Description step-by-step

Detection of absence

Occupants have a significant impact on consumptions inside a building, as shown by Page (2007). Several methods are studied to estimate their behavior. Jiang et al. (2016) uses feature scaled extreme learning machine (FS-ELM) on indoor CO_2 concentration sensor to estimate the occupancy. Hong et al. (2016) present advances in terms of sensors techniques to monitor occupants behavior as well as analytical and modeling methods. Zhao et al. (2014) focuse on power consumption data-mining to model occupants behavior. Vorger (2014) developed statistical scenarii to implement in dynamic thermal simulation (DTS) software not to identify the absence but to approach at best the behaviour of inhabitants.

In our case, we decided to focus on absence period. This will enable us to minimize the impact of occupants behaviour and heat emission whilst taking into account their main consumptions when they are present. We also wanted a method simple to implement since detecting occupants' absence is not the main objective of this research work. This step might be further developed and implemented later.
As a result, we decided to analyse thermal data and power consumption data from sensors set up on site.

This might appear very simple. However, having usable data is not as easy as it seems, especially in a new-built house where sensors just have been set up.

Modifications of the limit conditions of the model

In this part, we want to minimize the impact of factors which are not related to the construction, such as weather conditions or internal gains due to household appliances. As a matter of fact, Enertech (2011) shows that using weather data from a different site, even close, can induce an over-estimation in heating of 30%. As for internal gains, it can also represent an over-estimation of 30% under a warm climate as presented by Clevenger and Haymaker (2006).

Our case study is equipped with sensors that enables us to measure weather conditions. As a result, we decide to create a new weather file thanks to the Weather Converter Program (Energy Plus (2015)) to get rid of the gap between the weather condition values used in simulation and the one measured on-site in order to put the model in the same environment than the studied dwelling.
On top of that, we also have at our disposal the specific electricity consumption data (household appliances, plugs, lighting, and so on). Hence, we create new scenarii based on these measurements.

These steps help to recreate the excitation vector of the DTS model in order, once again, to place the model in the same conditions than the building.

Calculation of the coefficient of variation of the root mean square error ($C_v(RMSE)$) and the mean bias error (MBE)

There are many ways to calculate the error (or difference) between a serie of Measurements (M_i) and Simulations (S_i) as shown in Robillart (2015).
To fit our case, we decided to follow the guideline defined in Rosenberg and Eley (2013), Rosenberg and Eley (2016). Hence, we calculate the MBE (1) and the $C_v(RMSE)$ (3).

$$MBE[\%] = \frac{\sum_{i=1}^{n} S_i - M_i}{\sum_{i=1}^{n} M_i} \times 100 \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(S-M)_i^2}{n}} \tag{2}$$

$$C_v(RMSE)[\%] = \frac{RMSE}{A} \times 100 \tag{3}$$

With :

- $A = \frac{\sum_{i=1}^{n} M_i}{n}$
- S = Simulated data (timestep = 1 hour)
- M = Measured data (timestep = 1 hour)
- n = Number of hours on the studied period

Method of Morris

Speaking of sensitivity analyses (SA), we mainly consider three types of methods:

- Local sensitivity analyses
- Global sensitivity analyses
- Screening methods

As mentionned by Saltelli et al. (2008), local analyses calculate partial derivatives of the considered output against inputs around a reference point. Hence, it focuses on the local impact on the model.
When we face a non-linear model, local SA methods show some limits. Indeed, sensitivity indices would

not take the same value on all uncertainty intervals of the inputs. Thus, obtained information are no longer relevant.

In that case, the use of global sensitivity analyses is highly recommended. These evaluate global uncertainty and sensitivity of each input on the considered output. They rely on variance decomposition. As described in Saltelli et al. (2008), they are based on two main properties:

- *Inclusion of influence of scale and shape*

  "The sensitivity estimates of individual factors incorporate the effect of the range and the shape of their probability density function".
- *Mutidimensional averaging*

  "The sensitivity estimates of individual factors are evaluated varying all other factors as well".

Now focusing on screening methods, its main purpose is to answer the following question : which factors among all do actually matter ? In other words, screening methods aim to reduce the list of important factors.

The Morris screening method (Morris (1991)) can be defined as a "global sensitivity experiment" because the whole space of variation of factors is covered. This method is well appropriated when the model is computationally expensive or with a lot of input factors to study.

Sampling plan of this method is an One-At-a-Time (OAT) plan. For N inputs considered, we define a minimal and a maximal bound (i.e. uncertainty). These bounds create the variation space of our inputs. Each space is then divided in Q levels $\left\{0; \frac{1}{Q-1}; \frac{2}{Q-1}; ...; 1\right\}$. After that, r repetitions of a random trajectory are made, going throught $N+1$ nodes. As a result, one trajectory necessitates $N+1$ simulations. Figure 1 illustrates three potential trajectories when N, Q, r equal 3.
The uniformity of samples distribution is kept while defining a length Δ such as $\Delta = \frac{Q}{2 \times (Q-1)}$.

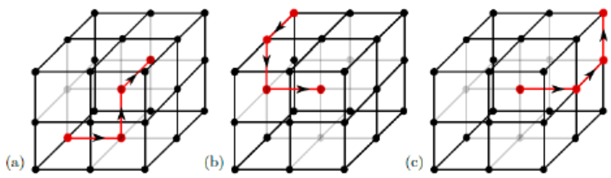

Figure 1: Trajectories, $r = 3$, $N = 3$, $Q = 3$ (Munaretto (2014)).

The method aims to identify which factors have :

- negligible effects (1)
- linear and additive effects (2)
- non-linear or interaction effects (3).

To do that, the following steps are carried out.
For each i, i.e. trajectory, we calculate the elementary effect $E_j^{(i)}$ (5) for each input factor θ_j on the variation of the output M (4) between two consecutive timesteps.

$$\begin{aligned} VAR_j^{(i)} = & M(\theta_1^{(i)}, ..., \theta_j^{(i)} + \Delta, ..., \theta_N^{(i)}) \\ & - M(\theta_1^{(i)}, ..., \theta_j^{(i)}, ..., \theta_N^{(i)}) \end{aligned} \tag{4}$$

$$E_j^{(i)} = \pm \frac{VAR_j^{(i)}}{\Delta} \tag{5}$$

Once simulations are over, we determine the following sensitivity indices :

- mean effect:

$$\mu_j = \frac{1}{r} \sum_{i=1}^{r} E_j^{(i)} \tag{6}$$

- absolute mean effect, added by Campolongo et al. (2007) solving the effects of opposite signs:

$$\mu_j^* = \frac{1}{r} \sum_{i=1}^{r} |E_j^{(i)}| \tag{7}$$

- standard deviation :

$$\sigma_j = \sqrt{\frac{1}{r-1} \sum_{i=1}^{r} (E_j^{(i)} - \mu_j)^2} \tag{8}$$

The absolute mean μ^* (7) and the standard deviation σ (8) are then used to identify the important factors. Indeed, a high mean means that the factor has a significant influence on the output and a high standard deviation that the factor has either interaction or non-linear effects.
An easy way to represent results from the analysis, and to visualize it, is shown on Figure 2.
This method fits our purposes. Indeed, we need a SA method computationally cheap, ranking the input factors by importance order. This enable us to decrease the list of factors for the bayesian algorithm.

Dispersion of uncertainties

The purpose of this step is to find a range of values where the important factors, identified thanks to the method of Morris, should stand to match the measurements.
Here we use the ABCPMC (Approximate Bayesian Computation - Population Monte-Carlo).

Wilkinson (2013) presents a state of the art of bayesian algorithms. Thus we choose to follow the

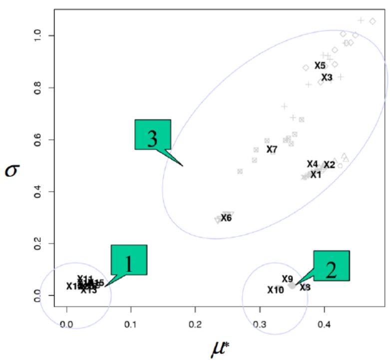

Figure 2: Typical representation of Morris analysis results (Iooss (2009)).

one explained by Beaumont et al. (2009). There, he explains that, when the likelyhood function is not available, bayesian algorithms are used as rejection technique.
We define θ as a studied factor. We first observe $y \sim f(y|\theta)$, the prior distribution on the factor θ. Originally, the ABC algorithm simulates $\theta^{'} \sim \pi(\theta)$ and the resulting distribution $x \sim f(x|\theta^{'})$. Then, it accepts the simulated factor $\theta^{'}$ only if $x = y$.

$$\omega_i^{(t)} \propto \frac{\pi(\theta_i^{(t)}) \times L_{t-1}(\theta^*|\theta_i^{(t)})}{\pi(\theta^*) \times K_t(\theta_i^{(t)}|\theta^*)} \tag{9}$$

The weight, i.e. the probability, is defined in (9), where:

- $\pi(\theta)$ is the prior distribution of the factor θ,
- K_t is a Markov transition kernel,
- L_{t-1} is a random transition kernel.

Beaumont et al. (2009) shows that the weight (9) is skewed since it is only function of y when the tolerance bound ϵ equals zero. In all other cases, ()9) is false since the output is not distributed from $\pi(\theta|y)$.

As a result, they tried to correct this bias by using an importance sampling. Thus, algorithm 1 should be applied.

This algorithm has already been written in Python 2. As a result, the package *abcpmc* (Akeret et al. (2015)) will be used.
It should be pointed out that we have chosen to use this algorithm because it presents good results (Beaumont et al. (2009), Lenormand et al. (2011), Robillart (2015)) and above all, because it can be parallelized, which considerably decreases the computationnal time.

Case study

Thermal description

Our case study is a single house of roughly 200 m^2 whose 3D model is presented in Figure 3 (realistic

for t = *1* **do**
  for $i \in [1 : N]$ **do**
    while $\rho(x, y) > \epsilon_t$ **do**
      Simulate $\theta_i^{(t)} \sim \pi(\theta)$ and $x \sim f(x|\theta_i^{(t)})$;
    end
    Set $\omega_i^{(t)} = \frac{1}{N}$;
  end
  Take τ_t^2 as twice the empirical variance of the $\theta_i^{(t)}$'s
end
for $t \in [2 : T]$ **do**
  for $i \in [1 : N]$ **do**
    Pick θ_i^* from the $\theta_j^{(t-1)}$'s with probabilities $\omega_j^{(t-1)}$;
    while $\rho(x, y) > \epsilon_t$ **do**
      Generate $\theta_i^{(t)}|\theta_i^* \sim \mathcal{N}(\theta_i^*, \tau_t^2)$ and $x \sim f(x|\theta_i^{(t)})$;
    end
    Set $\omega_i^{(t)} \propto$
    $$\frac{\pi(\theta_i^{(t)}}{\sum_{j=1}^{N} \omega_j^{(t-1)} \times \varphi\{\tau_t^{-1} \times (\theta_i^{(t)} - \theta_i^{(t-1)})\}};$$
  end
  Take τ_{t+1}^2 as twice the empirical variance of the $\theta_i^{(t)}$'s
end

Algorithm 1: ABCPMC algorithm.

one) and Figure 4 (simplified one).

Figure 3: Realistic 3D model.

It is composed of 4 half-floors. At level 0, we find 3 bedrooms and their own bathroom. At level 1, a living-room as well as the cellar where you can find all the electrical cabinet. At level 2, there are the living space with the dining-room and the kitchen. Finally, at level 3, there is the master suite.

Thermal characteristics are presented in Table 1. Windows are composed of double glazing with a 15-millimeter argon layer in the between as shown in Table 2 . Regarding infiltrations, the value of 0.6

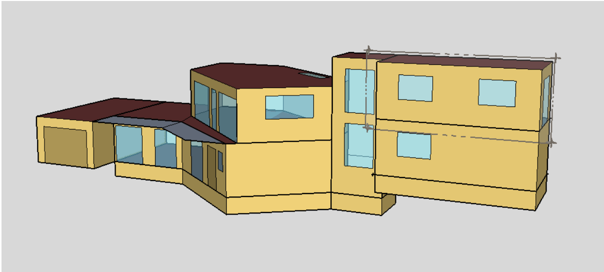

Figure 4: EnergyPlus 3D model.

$m^3/h/m^2$ was considered.

Table 1: Thermal characteristics.

Wall type	**R** $[m^2.K/W]$	**U** $[W/(m^2.K)]$
External wall	6.12	0.16

Table 2: Windows characteristics.

Wall type	$\mathbf{U}_g$ $[W/(m^2.K)]$	$\mathbf{f}_s$ [-]	$\mathbf{T}_L$ [-]
Windows	1	0.5	0.71

Heat emitters are electrical heaters with inertia. The hot-water is produced thanks to a thermodynamic water heater with a tank of 270 liters. The ventilation system is a single-flow (HygroB) by extraction.

Monitoring and available data

The construction of this house ended at the end of summer 2015 and was inhabited since then. Data from the beginning of October till the end of April were collected.

The air temperature and the humidity are measured in every room. On top of that, CO_2 and COV sensors are set up in the master suite and the living-room.

A weather station located on the roof enables the measurement of the environment conditions. Solar flux is also measured thanks to a sensor located on the south side of the house, and another temperature sensor, located on the north side, enables a second measurement of the outside temperature.

This house is quite remarkable regarding the instrumentation of systems. In fact, an impressive amount of data is provided.
For instance, sub-meters for heating of each room (bedrooms, living-room, office), for household appliances (oven, fridge, dishwasher, washer) as well as for the plugs and lights are installed.
On top of their consumption, temperatures (in and out) of the hot water system and of the ventilation system are measured.

If we take a look at the energy production, we can see that the dwelling has photovoltaic panels and a wind turbine.
To measure the amount of energy originated from the PV panels, a sub-meter is directly set up in the electrical cabinet. As for the energy produced by the wind turbine, it can be read thanks to the energy management device. This device enables the control of the use of the produced energy. On this device, we can also read the amount of energy which is reinjected on the grid, as well as various data of the batteries (capacity : 51.8 kWh), and particularly the state of charge.

As mentionned before, the instrumentation of this dwelling is very wide, and specifically regarding the monitoring of occupants usage.
As a matter of fact, most of lights are equipped with state sensor. All televisions dispose of a state sensor and an energy meter. Motion detector have been set up in the hallway, the living-room and the stairs. The state of the openings is also registered.

Chosen data for the study

When we had to choose what data we were going to study, the first problem that occured was the avaibility of measurement. Indeed, as the house was brand new, some data were still not available as we started our study.

As a result, we decided to concentrate our analysis on the room containing the most data, that is to say the master suite. Figure 5 shows detail of material used in its construction.

On top of that, as the dwelling does not have a cooling system, we can not study our methodology on cooling consumption. As a result, we focus on comfort (wintry and summery) and on heating consumption. By comparing summery and wintry data, we expect to have different results for the analysis of Morris.

Application of the method

Identification of absence period of time

In order to identify these periods, our first approach was to rely on only "ambiant" sensors. We defined a period of absence when the following conditions were met:

- CO_2 concentration < 500 ppm,
- State of light changes (from 1 to 0) and stays at 0,
- Motion detector stays at 0.

In February, we obtained the results presented in Table 3 . 10 periods have been identified. 3 are remarkable since they last more than a day.
To consolide these results, an analysis of the different consumptions on winter 2015-2016 was carried out.

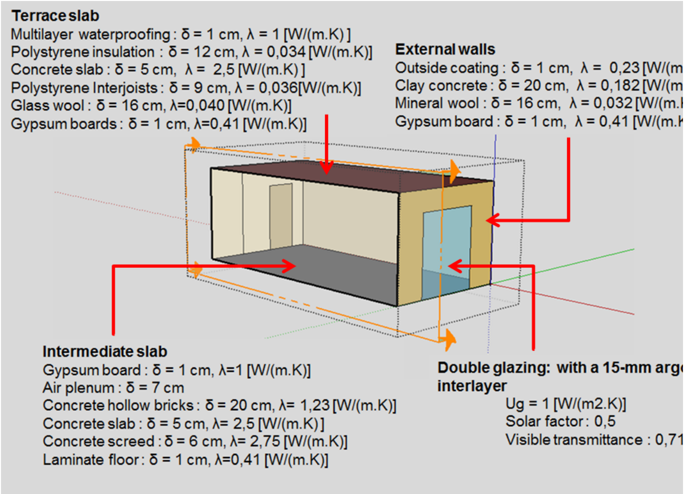

Figure 5: Construction details - Master suite.

Table 3: Absence periods identified thanks to sensors.

N^o	Beginning	End	Duration
1	02-01 08:25	02-01 20:10	11:45
2	02-02 07:50	02-02 21:20	13:30
3	02-04 07:20	02-07 18:45	3 days 11:25
4	02-05 08:45	02-05 17:50	09:05
5	02-06 09:50	02-06 19:25	09:35
6	02-08 07:25	02-09 18:35	1 day 11:10
7	02-12 09:30	02-14 20:45	2 days 11:15
8	02-19 11:00	02-19 19:40	08:40
9	02-23 14:05	02-23 19:25	05:05
10	02-29 10:30	02-29 17:40	07:10

Figure 6 shows consumptions of February with:

- The green curve as the electric consumption of the heating,
- The blue one as the consumption for the production of hot water,
- Finally, the red one as the consumption due to the ventilation system.

As it can be seen, the 3 curves show null consumptions at some point of the month. It can be assumed that it represents either daily or extended absences.

It is important to note that both methods detect the absence period in mid February. The analysis of systems consumptions detects a period of absence from the 8^{th} till the 13^{th} of February whereas the one on "ambiant" sensors detects 2 absence periods on this duration: from the 8^{th} till the 9^{th} of February (period n^{o}6 in Table 3) and from the 12^{th} till the 14^{th} of February (period n^{o}7).

Clearly, if there has been some hot water consumption, it means that someone was inside the house. As a consequence, we can exclude periods n^{o}1, 3, 4 and 9. Period n^{o}2 is partially true. Indeed, there is hot

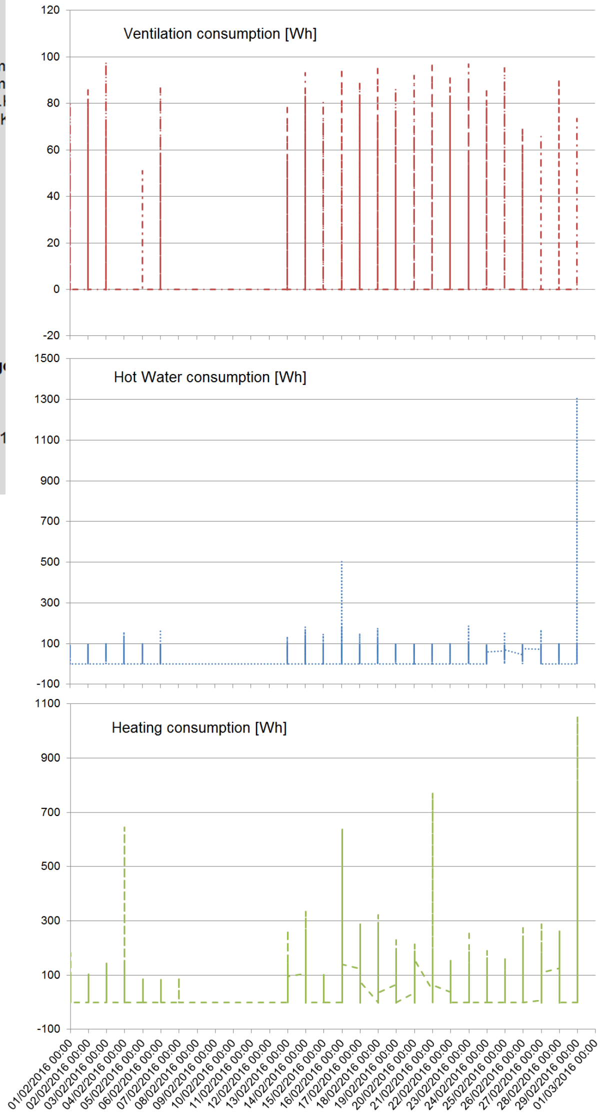

Figure 6: Absence periods identified thanks to meters.

water consumption from 6.30 a.m. till 1 p.m. and nothing more till morning of the 3^{rd} of February. Periods n^{o}5 and 10, are also partially true.

To go further in our method, we decided to follow the analysis of systems consumption for two main reasons :

- first, consumptions data processing is more reliable as it was shown previously and it is also easier,
- second, monitoring systems consumptions is very common, and even mandatory in thermal regulation for new buildings.

Modifications of the limit conditions of the model

As said previously, we want to place our model in the same conditions than our studied building. Hence, we inserate here the measured external temperature and solar flux in order to reproduce at best the envi-

ronmental characteristics. Moreover, we create a new scenario of electric equipment based on measurements made inside the house.
Indeed, this is an indispensible step if we want to be able to identify a potential default due to the construction and not due to the use or the environment.

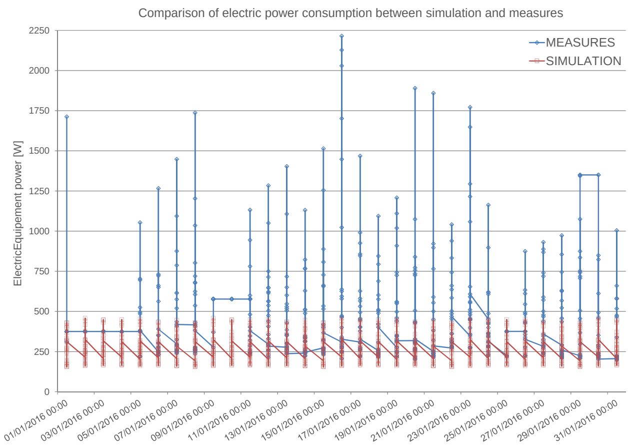

Figure 7: Comparison between the simulated and the measured consumptions.

For instance, Figure 7 shows the usual input scenario for power use (red curve) and the measured one (blue curve). As expected we can see that the measured electric consumption is significantly higher. The same can be represented for weather conditions (outdoor temperature and global radiation rate).

Preliminary results of the Morris method

First of all, we apply the method of Morris between the temperature measured in the master suite and the one simulated by EnergyPlus. We apply the method of Morris to the whole set of factors (materials, electric equipment, ventilation, occupation). In total, we analysed the influence of 143 input factors. A quick description of these factors is given in table 4.

Two trajectories are studied, hence 20,736 simulations are needed to generate the results of the Morris' analysis presented on Figure 8. Theses simulations are carriout on a whole-year period, and the coefficient of variation of the RMSE is calculated on the main absence period detected thanks to the analysis performed on the meters, i.e. from the 8^{th} till the 13^{th} of February. It should be noted that, each time we mention an absence period in following paragraphs, we actually refer to this main absence period.
It appears that factors (1) to (9) are supposed to have an important effect on the output ($C_v(RMSE)$ on the air temperature in the master suite). The U-factor of windows (7) is seperated from the others mainly because of its absolute mean value.

Table 4: Absence periods identified thanks to sensors.

Category	Factor	Details
Materials	All material used in the construction	Thickness Conductivity Specific heat Density
Windows	All window used in the construction	Factor solar U-Factor
People	Occupation rate (nightzone)	People/Area
Electric Equipment	Power	Watt/Area
Infiltration	Flow (master suite)	Flow rate
Ventilation	Flow (master suite)	AirChanges/hour

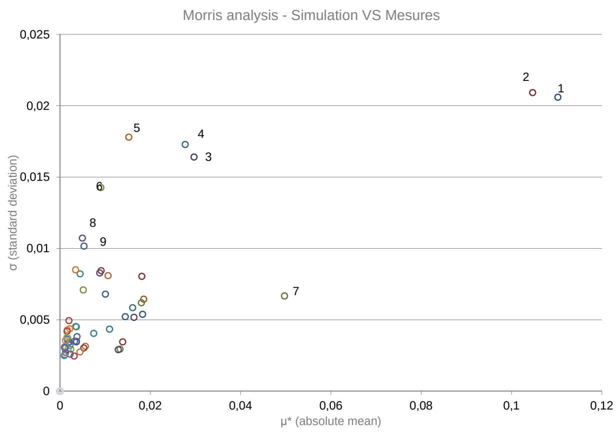

Figure 8: Analysis of Morris on the $C_v(RMSE)$ on the air temperature - Measurements vs. Simulation

The following factors - clay concrete (thickness (3), conductivity (4) and density (5)), glass wool (thickness (8) and conductivity (9)) and screed density (6) - have a high value of standard deviation which significate a linked influence. The mineral wool thickness (1) and conductivity (2) have both high mean and standard deviation values which indicate an overall influence on the output. This makes sense since the mineral wool is the insulation used in the external walls as it can be seen on Figure 5.

Then, we tested this analysis between two EnergyPlus files in order to control every factor inside our model. Hence, we are able to deteriorate on purpose one model by changing for instance the thickness of insulation of a wall. The purpose here is to test our method to find out if it can detect the factor that we have changed.

The results presented here aim to detect the decrease from 18 down to 15 cm of the thickness of the mate-

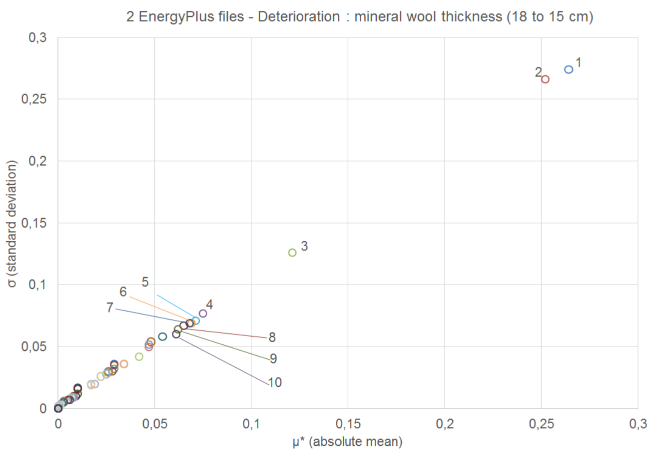

Figure 9: Analysis of Morris on the $C_v(RMSE)$ on the air temperature - 2 EnergyPlus files

rial named "mineral wool". The $C_v(RMSE)$ on the ambient temperature in the master suite is considered as the output.
Figure 9 shows the results of the method of Morris when simulation is running on a whole year with the $C_v(RMSE)$ calculated on data of the period of absence.
Comparing Figure 9 and Figure 8, 9 factors have a significant influence on the $C_v(RMSE)$. We decided then to reduce the list of studied factors to these factors in order to reduce computional time. Moreover, as we wanted to analyse the effects of household appliances consumption and of ventilation flow, we added these two.

To sum up, the Morris analyses will be further conducted on : the mineral wool, the clay concrete, the glass wool, occupancy level, household appliances power and ventilation flow.

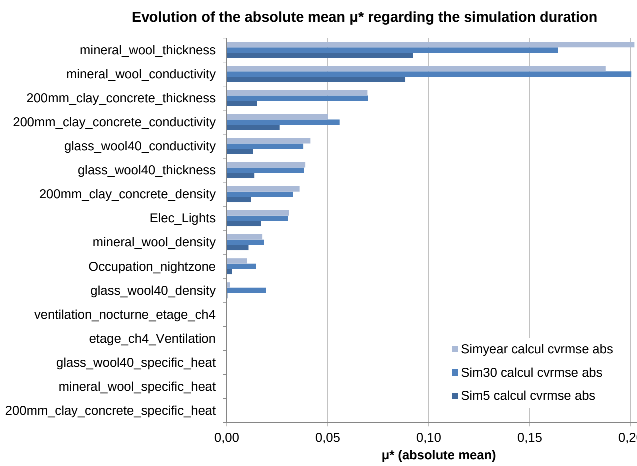

Figure 10: Comparison on μ^*: simulation on 1 year vs. on a short periods

The next step was to see the impact of the simulation on a short period of time. Indeed, STD softwares

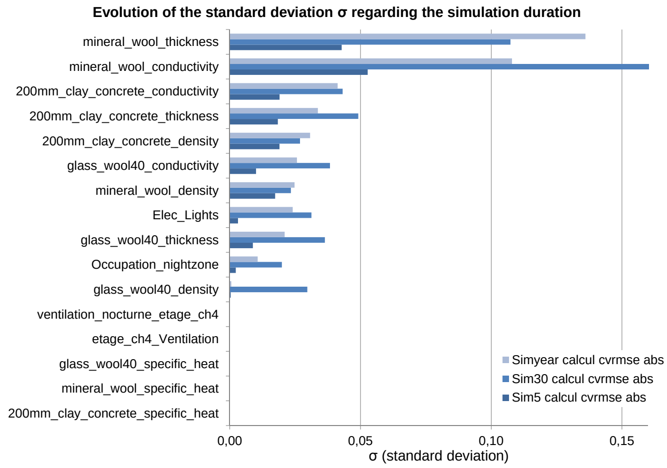

Figure 11: Comparison on σ: simulation on 1 year vs. on a short periods

need a period of initialization so that results can be relevant. Figures 10 and 11 show differences between three analyses of Morris, respectfully with:

1/ simulation running on a whole year - "Simyear",
2/ simulation beginning one month before the start of the absence period (i.e. on the 8^{th} of January) - "Sim30",
3/ simulation starting at the beginning of the absence period (i.e. on the 8^{th} of February) - "Sim5".

The mention "calcul. cvrmse abs" means that $C_v(RMSE)$ was calculated on the absence period. Figure 10 presents the impact on the absolute mean (μ^*) whereas Figure 11 presents the one on the standard deviation (σ).
As it can be seen on these two Figures, when simulation is running on a year or on a period starting 30 days before the absence, σ and μ^* stay in the same order of magnitude, whereas values, when simulation is only running on the absence period, are a bit lower. However, the order of importance stays globally the same.

The first conclusion that can be made is that we must not start our simulation directly at the beginning of the absence period if relevant results are wanted.
Secondly, it is not necessary to simulate on the whole year since the results are really similar between a simulation on a year ("Simyear") and a simulation starting one month before the beginning of the absence period ("Sim30").
On top of that, simulate on a whole year is quite time consuming. Indeed, two hours and a half are needed to achieve Morris analysis in that case whereas 25 and 15 minutes only are needed for "Sim30" and "Sim5", respectively.

It is then clear that it is wiser to start our simulation one month before the absence period, so that both results and calculation time are optimized.

Conclusion and further work

In this paper, we have proposed a methodology aiming the identification of defaults in a building. It is based on 4 crucial steps:

- First, we identify periods where occupants are absent to focus on data which are the less influenced by occupants behaviour. The analysis of electric data was pointed out as a reliable way to do that.
- Secondly, we reinject inside our model weather parameters and electrical equipment consumptions that are measured in order to minimize this bias.
- Then, we perform an analysis of Morris on the $C_v(RMSE)$ of comfort and energy consumption in order to reduce the list of input factors for the calibration process.
- Finally, we will apply the bayesian algorithm ABCPMC to evaluate the range where the factors identified thanks to the Morris's method should be to fit the measurements.

We applied this methodology on a case study. We first demonstrated that it is possible to identify absence period thanks to the study of consumption meters. Secondly, we showed that the insulation has a significant effect on the inside temperature. We tested different simulation duration and concluded that a whole year is not necessary to obtain relevant results. Indeed, starting one month before the detected absence period is enough.

However, some programming steps are still needed. The final stage (the ABCPMC program) is currently studied and we are hoping to achieve it before the end of the year 2016 so that we can analyse results as soon as possible.

Finally, we will first test it between 2 EnergyPlus files. This will enable us to deteriorate intentionally one of them, e.g. changing the thickness of insulation, while controling every data (effects of : occupants, weather, and so on). By comparing these two files, we are hoping to succeed to detect the created default and then have conclusive results when comparing simulation and measurements.

Acknowledgement

We would like to acknowledge the support of the SNEF[1] for funding this research through the G2eLab laboratory. This work took place within the project COMEPOS supported by the ADEME[2].

[1]Societe Nationale d'Espaces Ferroviaires

[2]French Agency for the Environment and Energy Management

References

Akeret, J., A. Refregier, A. Amara, S. Seehars, and C. Hasner (2015). Approximate Bayesian computation for forward modeling in cosmology.

Beaumont, M. A., J.-M. Cornuet, J.-M. Marin, and C. P. Robert (2009, December). Adaptive approximate Bayesian computation. *Biometrika 96*(4), 983–990. arXiv: 0805.2256.

Campolongo, F., J. Cariboni, and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software 22*, 1509–1518.

Clevenger, C. M. and J. Haymaker (2006). The impact of the building occupant on energy modeling simulations. Montreal, Canada.

Energy Plus (2015). Auxiliary EnergyPlus Programs - Extra programs for EnergyPlus.

Enertech (2011, April). Evaluation par mesure des performances nergtiques des 8 btiments construits dans le cadre du programme europen Concerto. Technical report, Enertech.

Erhorn, H. and H. Erhorn-Kluttig (2011, January). Terms and definitions for high performance buildings. Technical report, Concerted Action Energy Performance Buildings.

European Commission DG ENV (2007). *Buildings and climate change: status, challenges, and opportunities*. Nairobi, Kenya : United Nations Development Programme ; Paris, France: UNEP DTIE, Sustainable Consumption and Production Branch. OCLC: ocn137238226.

EVO (2012, January). Concepts and Options for Determining Energy and Water Savings.

Hong, T., S. C. Taylor-Lange, S. DOca, D. Yan, and S. P. Corgnati (2016, March). Advances in research and applications of energy-related occupant behavior in buildings. *Energy and Buildings Volume 116*, Pages 694–702.

Iooss, B. (2009). Analyses dincertitudes et de sensibilit de modles complexes - Applications dans des problmes dingnierie. In *Rencontres "Maths-Mto" Toulouse - http://www.math.univ-toulouse.fr/baehr/meteo_SMAI/Pres/Pres_Iooss.pdf*.

Jiang, C., M. K. Masood, Y. C. Soh, and H. Li (2016, November). Indoor occupancy estimation from carbon dioxide concentration. *Energy and Buildings 131*, Pages 132–141.

Lenormand, M., F. Jabot, and G. Deffuant (2011, November). Adaptive approximate Bayesian computation for complex models. *arXiv:1111.1308 [math, stat]*. arXiv: 1111.1308.

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics 33*, 161–179.

Munaretto, F. (2014). *Etude de l'influence de l'inertie thermique sur les performances nergtiques des btiments*. Ph. D. thesis, MINES Paristech.

Page, J. (2007, October). *Simulating Occupant Presence and Behaviour in Buildings*. Ph. D. thesis, Ecole Polytechnique Federale de Lausanne.

Robillart, M. (2015). *tude de stratgies de gestion en temps rel pour des btiments nergtiquement performants*. Ph. D. thesis, MINES Paristech.

Rosenberg, M. I. and C. Eley (2013, May). A Stable Whole Building Performance Method for Standard 90.1. *ResearchGate 55*(5).

Rosenberg, M. I. and C. Eley (2016, June). A Stable Whole Building Performance Method for Standard 90.1-Part Ii. *ASHRAE Journal, 58(6):28-42*.

Saltelli, A., K. Chan, and E. M. Scott (2008, October). *Sensitivity Analysis*. Wiley.

Vorger, E. (2014, December). *tude de l'influence du comportement des habitants sur la performance nergtique du btiment*. phdthesis, Ecole Nationale Suprieure des Mines de Paris.

Wilkinson, R. D. (2013). Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Statistical Applications in Genetics and Molecular Biology 12*(2), Pages 129–162.

Zhao, J., B. Lasternas, K. P. Lam, R. Yun, and V. Loftness (2014, October). Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. *Energy and Buildings Volume 82*, Pages 341–355.