IDENTIFICATIONS OF MODELS BASED ON TIME SERIES FROM FIELD MEASUREMENTS FOR BUILDING SIMULATION 2013 CONFERENCE

Mathilde Grandjacques[1], Benoit Delinchant[1], Olivier Adrot[2], Clémentine Prieur[3], and Céline Helbert[3]
[1]G2ELab 38403 Saint Matrin d'Hères, France
[2]GSCOP 38031 Grenoble Cedex 1 avenue Félix Viallet, France
[3]LJK B.P. 53 38041 Grenoble cedex 9, France

ABSTRACT

To study the impact variations in input on output variation (Sobol index) of building model, a stochastic model (time series) of variables is introduced. To build the time series we used data measured in a building for each hour for a duration of one month.
The external temperature does not appear directly in the model but affects all the input. So after model it as Auto-Regressive process we propose to model input, for example the heating flux, by an ARMAX process.

INTRODUCTION

Energy consumption is becoming an important issue.
The building sector contributes up to 40% of total energy expenditure. An important part of waste comes from a poor design, the use of inappropriate technologies (e.g: choice of the heating system) and from the users' unreasoned behaviour.
Different models reflecting the design parameters and taking into account different inputs (external temperature, users' behaviour) are used today to improve the performance energy of new buildings(Clarke et al. (2002)). The uncertainty of the inputs must also be taken into account to forecast the consumption and predict some comfort.

To study the impact variations in inputs on variations in the output one of the tools is Sobol's index (Saltelli et al. (2004) and Da Veiga et al. (2009)) based on variance that allows to classify inputs X following of their influence on the output Y.

$$S_1 = \frac{V(E(Y|X))}{V(Y)} \in [0,1] \tag{1}$$

There is no existing model for all this inputs. The first step is to introduce a stochastic model in order to generate useful data for this analysis that will present in an other paper.

We address an accurate model for temperature and heat flux hour per hour in order to make forecasts (various model temperature day per day exist (Semenov and Barrow (1997)).

In the first part we describe the model and the data that we have used. In the second we briefly introduce the time series model. The third part explains the different choices of model and presents some practical results.

MODEL AND DATA DESCRIPTION

A building is composed of a thermal envelope, electrical equipments which can consume or store energy and of users.
Building can be modelled using state space representations:

$$\begin{cases} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{cases} \tag{2}$$

u: operative vector (temperatures, heat flux, occupancy,. . .)
x: state vector: e.g. wall temperature
y: output vector: internal temperature
A, B, C, D are matrices containing the physical characteristics of the building.

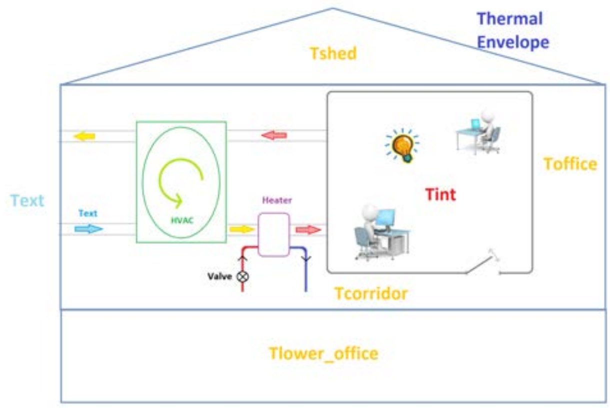

Figure 1: Building model

The following data of 1 are measured on the system with a sampling of one hour:

- the external temperature: T_{ext_t}
- the heat produced by the occupant: N_t
- temperatures of adjacent rooms: corridor, next office, lower office, shed.
- the heating flux: K_t
- intern temperature of the room of interest: Y_t

The heat flux is calculated using the specific heat of water fixed at $C_e = 4000J/kg.°C$, the water flux maintained constant by a regulation loop $Q_e(kg/s)$, and the difference upstream and downstream of the water in the heater.

$$K_t = Q_e * \Delta T * C_e$$

Unfortunately, the data contains a large amount of drop-outs. We decide to work only on one month.

From the study of the data, different trend observations can be drawn to choose the types of models constructed for different inputs.

External temperature is modelled by an AR(Auto-Regressive) process which is detail in the second part. A fast graphical study shows a strong seasonality of 24-hours (figure: 3 and 4).
The external temperature is not properly included in the model 2 presented above, but affects all of the inputs (except the heat produced by the occupants).The temperatures of adjacent rooms are closer to a classical regression. We have chosen to create a second model linking the temperatures of adjacent rooms to T_{ext} by Spline regression (figure: 2)(Eubank (1999)).

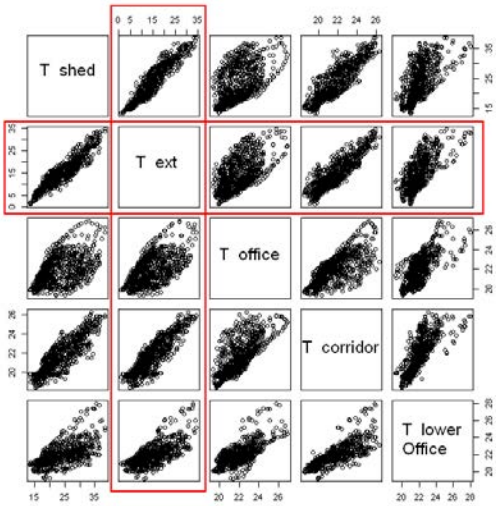

Figure 2: Temperature scatterplot

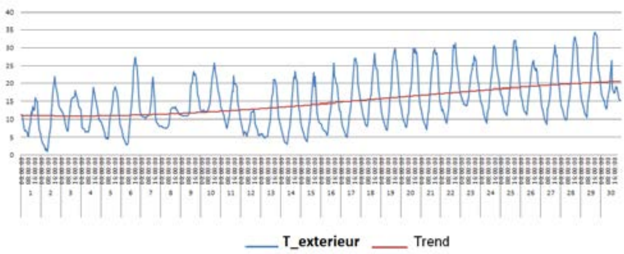

Figure 3: External temperature in April 2012: measured every hour for a month

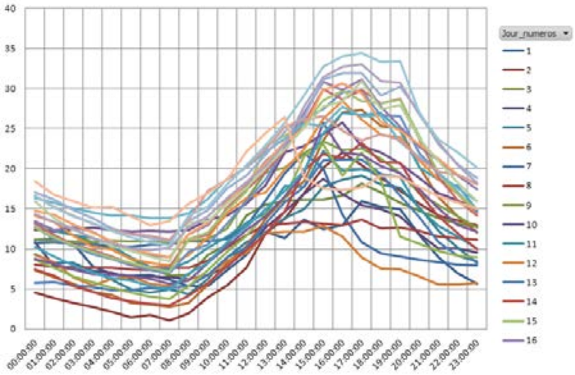

Figure 4: External temperature in April 2012: measured every hour of a day

As before the first graphical analysis shows a 24-hours seasonality (figure:7) and also a 7-day (figure:6) seasonality.

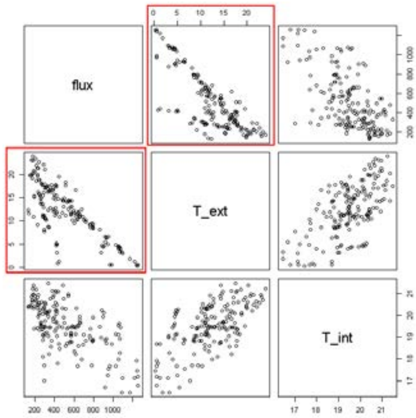

Figure 5: Scatterplot of working days

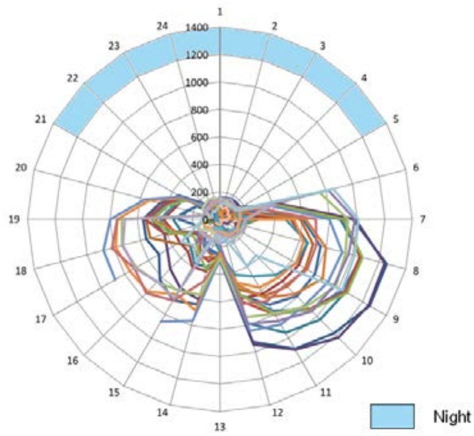

Figure 7: flux in November 2012: measured every hour of a day. Each ray of the circle represents a time of day

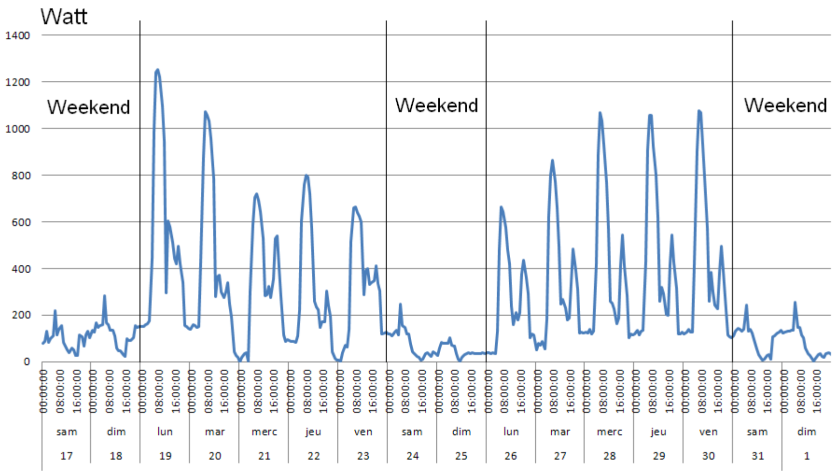

Figure 6: flux in November 2012: measured every hour for a month

The variance of the heat flux during working days is much superior to that of weekends and nights (figure:6 and figure:7) so we choose to study three different models:

- Nights: from 9 pm to 4 am
- Working days from 5 am to 8 pm
- Weekends

Next we will detail only the model of working days. The work is the same for the other two models.

TIME SERIES

The most often, the time series X_t is considered as an additive model (Brockwell and Davis (2009)):

$$X_t = m_t + s_t + Z_t \tag{3}$$

- m_t is the trend component which represents a slowly changing function
- s_t the seasonal component, a function with a known period: d
- Z_t is a "random noise component" which is stationary and general not white.

A time series is a set of observations x_t, each one being recorded at a specified time t.
The specificity of time series analysis, which distinguishes it from other statistic analyses, is precisely the emphasis on the order in which the observations are made. Classical statistical methods often require variables which are *stochastically independent* and *observed several times*. In time series analysis the main source of information is the temporal dependence between variables.
The study will consist in supposing that the observations, after some possible transformations, are stationary. The structure of the phenomenon is thus reflected by the correlation between the variables. The classical models used are ARMA (Autoregressive moving average) models.

Our first aim is to estimate and extract m_t and s_t in the hope that Z_t will turn out to be a stationary random process. We can then use the theory of such processes to find a satisfactory probabilistic model for $\{Z_t\}$, to analyse its properties, and use it in conjunction with m_t and s_t for purposes of prediction and control of $\{X_t\}$.
An alternative approach, developed extensively by Box and Jenkins, is to apply difference operators repeatedly to the data $\{x_t\}$ until the differenced observations resemble a realization of some stationary process $\{W_t\}$. We can then use the theory of stationary processes for the modelling, the analysis and prediction of $\{W_t\}$ and hence of the original process.

1.1 Graphical description

Before starting to study the time series, we need to check some points:

- regularity of observations
- stability of structures conditioning the phenomenon studied: The analysis technique seeks to determine the slow evolution of the phenomenon and its seasonal variations. This implies some kind of stability. When not checked, it can be obtained by decomposing observed time series in several series for instant working day and week-end.
- permanence of the definition of the variable studied (e.g.the rate air flux change can change)

So the first step in the analysis of any time series is to plot the data. Graphs allow to detect discontinuities and outlying observations and to detect a seasonality or see a trend. The inspection of a graph is also useful in the choice of the model.

Estimation and Elimination of trend and seasonal components

To estimate and eliminate the trend and the seasonal components we can use one of those three methods:

1. Small trend method (when the trend is small): We will estimate the mean trend on one period, then create a pattern through a period of seasonality after removing the trend.

2. Trend and seasonality are parametrized and we adjust each parameter by the least square method.

3. Differencing at lag d:
We introduce the lag-d difference operator ∇_d defined by:
$$\nabla_d X_t = X_t - X_{t-d} \tag{4}$$
To suppress a linear trend $m_t = at + b$ we just apply the operator $\nabla = \nabla_1$: $\nabla m_t = a$. In the same way any polynomial trend of degree k can be reduced to a constant by application of the operator $\nabla^k = \nabla \circ \cdots \circ \nabla$.
If $X_t = m_t + Z_t$ with $m_t = \Sigma_{j=0}^{k} a_j t^j$ and Z_t is stationnary with zero mean then:
$$\nabla^k X_t = k!a_k + \nabla^k Z_t \tag{5}$$
$\nabla^k X$ is a stationnary process with mean $k!a_k$.

Now $X_t = s_t + m_t + Z_t$ where $s_t$is the seasonal component of period-d, that means $s_{t+d} = s_t$.
To remove s_t we will apply the ∇_d. We obtain:
$$\nabla_d X_t = (m_t - m_{t-d}) + \nabla_d Z_t \tag{6}$$
Then, to remove the new trend $(m_t - m_{t-d})$ we can use ∇^k.

Modeling a stationnary process: Z_t

After removing trend and seasonality components we have to model the stationnary process Z_t.
When dealing with a finite number of random variables, it is often useful to compute the covariance matrix in order to gain insight into the dependence between them. For a time series $\{X_t, t \in T\}$ we need to extend the concept of covariance matrix to deal with infinite collections of random variables. The autocovariance function provides us with the required extensions.

Definition 1.1.1 *Autocovariance function:*
If $\{X_t, t \in T\}$ is a process such that $Var(X_t) < \infty$ for each $t \in T$, then the autocovariance function $\gamma_X(.,.)$ of $\{X_t\}$ is defined by:

$$\begin{aligned} \gamma_X(r,s) &= Cov(X_r, X_s) \qquad r,s \in T \\ &= E[(X_r - E(X_r))(X_s - E(X_s))] \end{aligned} \tag{7}$$

Definition 1.1.2 *AutoCorrelation Function ACF:*
The autocovariance function of a stationary process is the function of just one variable defined by:

$$\gamma_X(h) \equiv \gamma_X(h,0) = Cov(X_{t+h}, X_t) \quad \forall t,h \in T \tag{8}$$

The autocorrelation function (ACF) of $\{X_t\}$ is the function whose value at lag h *is*

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_{t+h}, X_t) \forall t,h \in T \tag{9}$$

Definition 1.1.3 *Partial AutoCorrelation Function PACF:*
The partial autocovariance function $\alpha(.)$ of a stationary time series is defined by:

$$\begin{aligned} &\alpha_X(1) = \rho(1) \\ &\alpha_X(k) = Corr(X_1 - P_{X_2,\ldots,X_k}(X_1), \\ &X_{k+1} - P_{X_2,\ldots,X_k}(X_{k+1})) \end{aligned} \tag{10}$$

where $P_{X_2,\ldots,X_k}(X_1)$ is the projection of X_1 on the space $(X_2,\ldots,X_k)$. It can be interpreted as the best explanation of X_1 by the linear function of $(X_2,\ldots,X_k)$. (ditto for $P_{X_2,\ldots,X_k}(X_{k+1})$)

This coefficient expresses the dependence between X_1 and X_{k+1} which is not due to the other $X_2,\ldots,X_k$

AutoRegressive process of order p**:** $AR(p)$

Definition 1.1.4

$$X_t = \Sigma_{j=1}^{p} a_j X_{t-j} + \epsilon_t \qquad \{\epsilon_t\} \sim WN^1(0,\sigma^2) \tag{11}$$

The PACF will be zero for all $k > p$ and the ACF will decrease to zero when k tends to infinity.

Moving Average of order q**:** $MA(q)$

Definition 1.1.5

$$X_t = \Sigma_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \qquad \{\epsilon_t\} \sim WN^1(0,\sigma^2) \tag{12}$$

The ACF will be zero for all $k > q$ and the PACF will decrease to zero when k tends to infinity.

AutoRegressive Moving Average process of order (p,q)**:** $ARMA(p,q)$

Definition 1.1.6

$$X_t = \Sigma_{k=1}^{p} a_j X_{t-j} + \Sigma_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \quad \{\epsilon_t\} \sim WN^1(0,\sigma^2) \tag{13}$$

The ACF and PACF will decrease to zero when for all $k > q$ tends to infinity.

[1] White Noise

AutoRegressive Moving Average with eXogenous inputs process of order (p, q): $ARMAX(p, q)$

Definition 1.1.7

$$X_t = \Sigma_{k=1}^{p} a_j X_{t-j} + \Sigma_{j=0}^{q} \theta_j \epsilon_{t-j} + \Sigma_{i=0}^{b} \eta_i D_i \quad \{\epsilon_t\} \sim WN \qquad (14)$$

with $\{D_t\}$ an external time series

Choice of the model order

The study of the PACF and the ACF can suggest the nature of the model. Once the model chosen and the coefficient estimated some criteria allow to choose the best the model combining complexity and the precision.

- AIC criteria :

  $$AIC = -2log(L(\theta)) + 2\nu \qquad (15)$$

  where $L(.)$ the likelihood function, θ model parameters and ν numbers of model parameters

- BIC criteria :

  $$BIC = -2log(L(\theta)) + log(n)\nu \qquad (16)$$

  where n is the number of observations of the series.

We will choose the model that minimize this criteria. Parameters' model are estimated by the Yule Walker algorithm (Brockwell and Davis (2009))

MODELS CONSTRUCTION:

Temperature model:

First we have to remove the trend (m_t) and the seasonal component (s_t).
We chose for this example the second method because we can easily estimate the trend and the seasonal component.

1. The trend is set so as polynomial. We choose the best polynomial that minimizes the AIC criteria and the most simple. We obtain a polynomial of third order:

   $$m_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

   Regression coefficients are obtained by a least square method.

2. After removing the trend, the seasonal component is identified by a Fourier series of period 24 hours (regression coefficients are obtained by a least square method). We can observe in figures 4 and 3 there is a seasonal form sinusoidal of period 24 hours.

3. The seasonal component is subtracted. The new process is called Zt.

To choose the best model process, we plot the ACF and PACF functions.

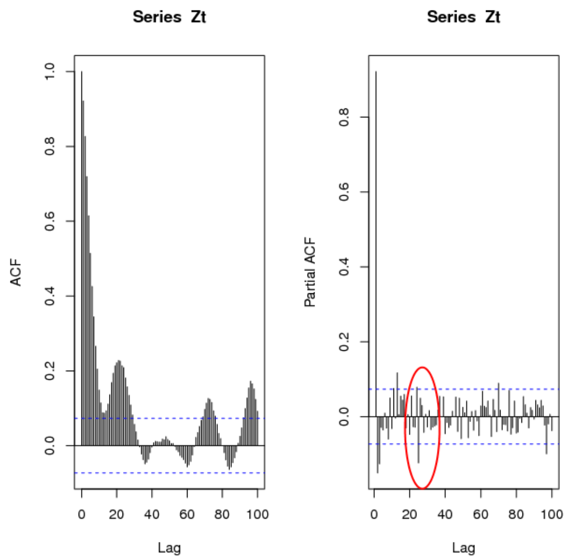

Figure 8: ACF and PACF of Zt

There is a strong correlation at 24 hours. The PACF function tends to zero for all lag superior to 24. The ACF tends to zero when lag tends to infinity. It suggests an $AR(p)$ process. The best $AR(p)$ identified under the AIC criterion is $AR(25)$ (AIC=2304)

The work of Cao and Wei (1998) and Dischel (1999) suggests to integrate a periodic variance for temperature model.
The time series decomposition can be written:

$$X_t = m_t + s_t + v_t * Z_t \qquad (17)$$

with v_t a periodic function and m_t the trend of the variance.

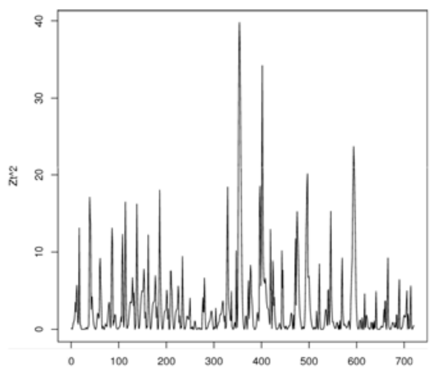

Figure 9: Z_t-variance as a function of time

The seasonal component is identified as an average daily pattern of length 24 v_t repeated 30 times (30 days) (figure 10).

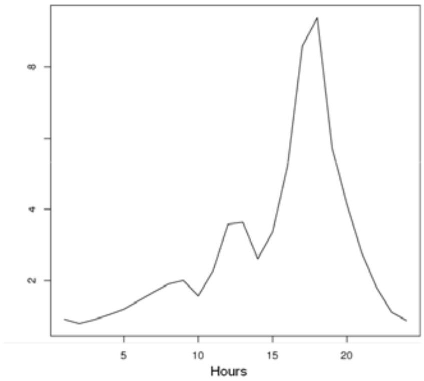

Figure 10: Daily patern variance: v_t

The new process studied is the centred standard process:

$$\eta_t = \frac{X_t - m_t - s_t}{v_t} = \frac{Z_t}{v_t} \tag{18}$$

The study of the η_t's ACF and PACF, shows that η_t is an $AR(p)$ process. There is always a strong correlation: at 24 hours. The best $AR(p)$ identified under the AIC criterion is $AR(25)$ but this time AIC criterion is equal to 542 (a decrease of 76%). Taking into account the periodicity of the variance we improved our model (the criteria of second model is smaller). We choose the second $AR(25)$.

Before conclude we will test the residual part. The hypothesis says that the residuals must be independent. One way is to plot the ACF function of the residual part or the p-value of the test of Ljung-Box which test the non correlation perform on them.

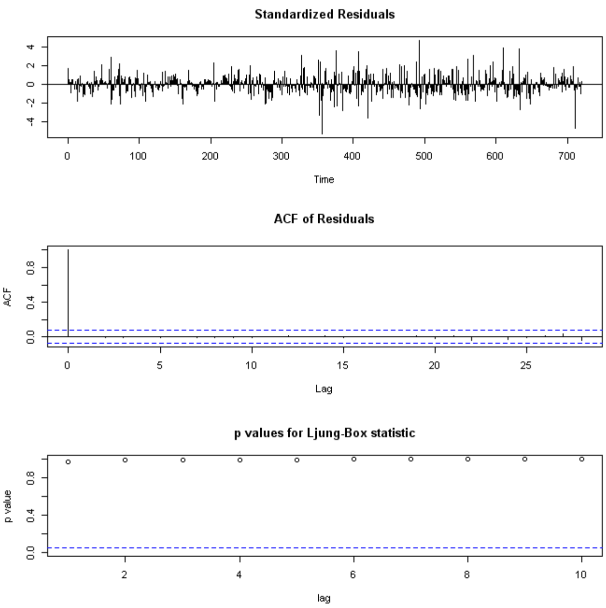

Figure 11: study of the residual parts ϵ_t

Here the residual are independent: the p-value is around 0.9 so the hypothesis that there are independent is accept more over the coefficient of the ACF are near to zero since the first lag.
To conclude we present the results of simulations.

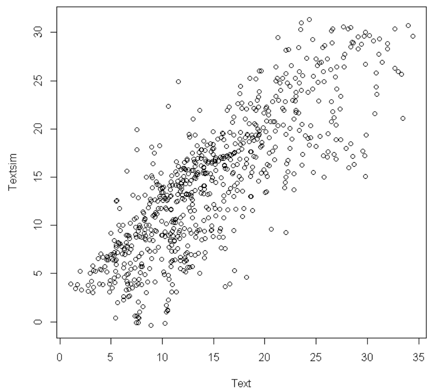

Figure 12: Simulated outdoor temperature as a function of real outdoor temperature

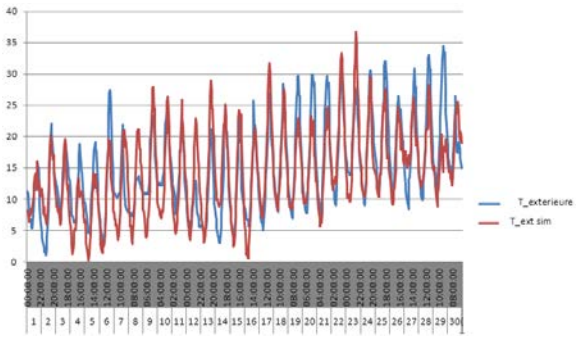

Figure 13: Simulated outdoor temperature and real outdoor temperature as a function of time

In figure 12, the curve representing the simulated outdoor temperature as a function of real outdoor temperature is almost a straight line through the origin with slope 1.
The figure 13, shows that the simulated outdoor temperature has the same behaviour than the real outdoor temperature with an aleatory part.

Heat flux:Working days

Introduced the ARMAX model for K_t the heat flux.

$$\begin{array}{rcl} K_t & = & s_t^K + m_t^K + \gamma_t^K \nu_t^K \phi_t \\ X_t & = & m_t + s_t + v_t * \eta_t \\ \phi_t & = & \alpha\eta_t + \lambda_t + \epsilon_t \end{array} \tag{19}$$

where λ_t is an ARMA process.

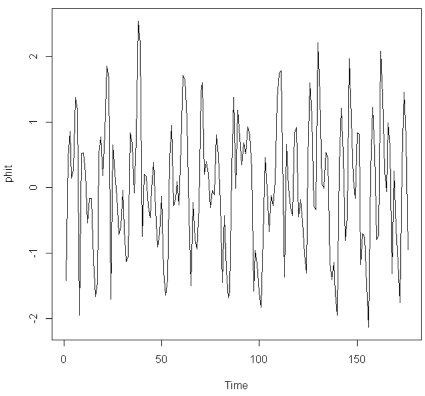

Figure 14: ϕ_t per hours

We calculate the trend and seasonal (m_t^K,s_t^K, γ_t^K, ν_t^K) component and remove them (figure:14) .
After remove the link with the external temperature, we study the process λ_t (figure:15).

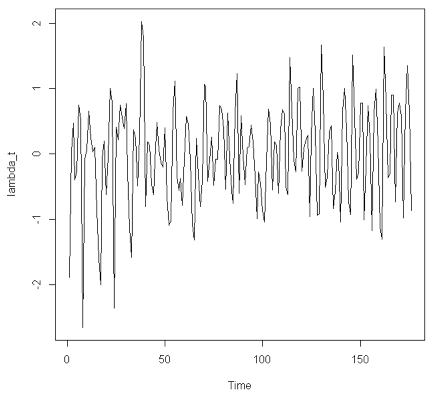

Figure 15: λ_t per hours

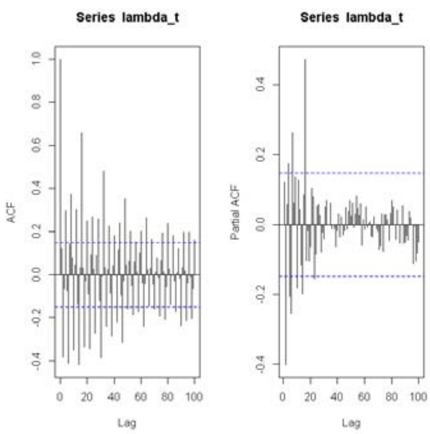

Figure 16: ACF and PACF of λ_t

There remains a strong correlation at lag 16. We will model after minimisation of AIC criterion by a $AR(16)$ process.
To choose definitely this model after plot the ACF and the p-value of the Ljung-Box test of the residuals part(figure:17):

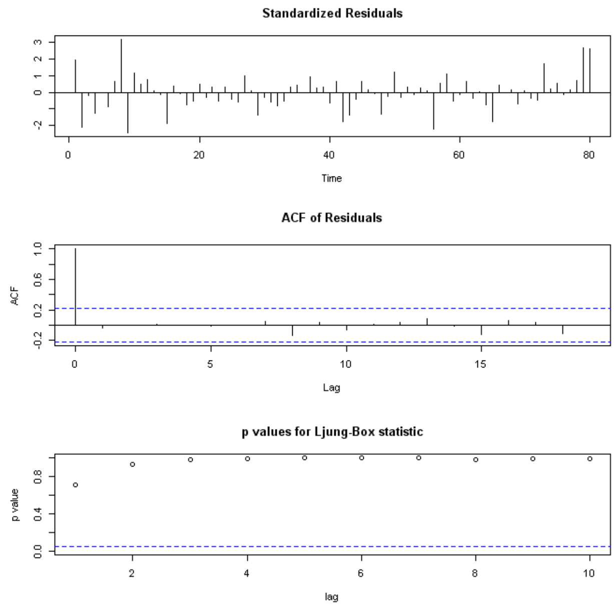

Figure 17: ACF and the p-value of the Ljung-Box test

All the coefficient of the ACF are close to zeros since the first lag and the p-value is around 0.9: the residuals are independent.
We model the heat flux by an ARMAX(16,0).
In figure 18, the curve representing the simulated heat flux as a function of real heat flux is almost a straight line through the origin with slope 1.

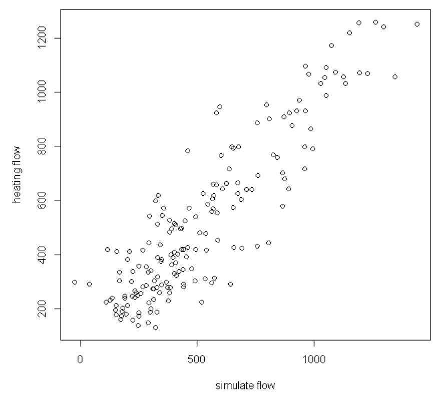

Figure 18: Real heat flux as a function of simulated heat flux

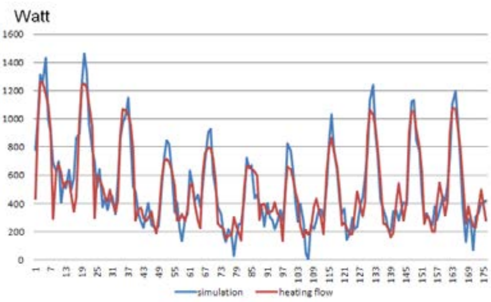

Figure 19: Simulated heat flux and real heat flux as a function of time

The figure 19, shows that the simulated heat flux has the same behaviour than the real heat flux with an aleatory part.

CONCLUSION

Time series provide a good representation of the temperature and heat flux.
These models allow us to take into account the temporal dependence between different times, which is essential for our application: computation of Sobol index.
It can be observed that seasonality plays an important role in the modelling of these time series. It's important to take into account the seasonality of the variance.
External temperature is an exciting variable of the model. All the inputs have to be model as an ARMAX process. After it will be necessary to integrate the users has the modeling process.
All the inputs can be considered as an ARMAX process of the external temperature and the number of occupants.
The result of this work will be to compute Sobol index for each input, in order to see which input is the most influent over time.

ACKNOWLEDGEMENT

A special thank is dedicated to the French National Research Agency (ANR) for its support through the FIABILITE project that was an occasion to gather many partners specialized in different buildings domains: thermal, electrical, computer science and energy control.

REFERENCES

Brockwell, P. J. and Davis, R. A. 2009. *Time Series: Theory and Methods*. Springer, 2nd ed. 1991. 2nd printing 2009 edition.

Cao, M. and Wei, J. 1998. Pricing Weather Derivative: an Equilibrium Approach.

Clarke, J., Cockroft, J., Conner, S., Hand, J., Kelly, N., Moore, R., OBrien, T., and Strachan, P. 2002. Simulation-assisted control in building energy management systems. *Energy and buildings*, 34(9):933–940.

Da Veiga, S., Wahl, F., and Gamboa, F. 2009. Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4):452–463.

Dischel, R. 1999. At last : a model for weather risk. risk, energy and power risk management.

Eubank, R. L. 1999. *Nonparametric Regression and Spline Smoothing, Second Edition*. CRC Press, 2 edition.

Gaaloul, S., Delinchant, B., Wurtz, F., Verdière, F., et al. 2011. Software components for dynamic building simulation. In *Building simulation conference, Sidney, Australie*.

Haldi, F. and Robinson, D. 2009. Interactions with window openings by office occupants. *Building and Environment*, 44(12):2378–2395.

Saltelli, A., Chan, K., and Scott, E. M. 2008. *Sensitivity Analysis*. Wiley, 1 edition.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. 2004. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons Ltd.

Semenov, M. A. and Barrow, E. M. 1997. Use of a stochastic weather generator in the development of climate change scenarios. *Climatic change*, 35(4):397–414.

Taylor, J. W. 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8):799–805.