Estimation of the water flow rate and energy consumption of a central heating system in an office building using system identification

Daniele Antonucci – Institute for Renewable Energy, EURAC research – Faculty of Science and Technology, Free University of Bolzano, daniele.antonucci@unibz.natec.it
Federico Noris – Institute for Renewable Energy, EURAC research – federico.noris@eurac.edu
Ulrich Filippi Oberegger – Institute for Renewable Energy, EURAC research – ulrich.filippi@eurac.edu
Andrea Gasparella – Free University of Bolzano – andrea.gasparella@unibz.it

Abstract

This study focuses on the application of system identification to estimate the flow rate and the heating consumption in a central heating system in an office building in San Michele all' Adige, Italy. The study was done within the European ICT PSP project Smart Build with the aim to reduce the energy consumption of existing public buildings through the application of ICT systems. The installed monitoring system, designed after an early energy auditing, foresaw the use of permanent temperature sensors coupled with a portable flow rate meter, to estimate thermal consumption. The flow rate varied with time depending on several factors therefore making it challenging to monitor it for only a short period. To overcome this obstacle we monitored the flow rate for four workdays and one weekend to identify a mathematical model which could explain the real system behaviour. A statistical analysis was performed at first to assess the reliability of the monitored data. Subsequently, we focused on the identification of autoregressive models with exogenous inputs with varying complexity and level of confidence. Different tests were performed on the hourly data to assess the reliability of the models. The best model was subsequently used to estimate the time-varying heating consumption for available monitoring periods with the aim to understand the impact of the ICT system as a function of the floor of the building and in order to show how system identification can reliably estimate the water flow rate in a central heating system and consequently be used to calculate the heating consumption. The calculated monthly average deviated from the bills by 20 %, attributable to the heating consumption of the non-monitored ground floor.

1. Introduction

In the context of the EU project SmartBuild (http://www.smartbuild.eu) we wanted to calculate the heating consumption in a monitored old office building in San Michele all' Adige constructed in 1874 and renovated in 2000 (more building detailed are summarized in Tab.1). The projects aims to reach 30 % of energy saving both in thermal and electrical consumption with the application of an ICT system. This latter has been installed in two of three building floors (1st and 2nd floor), but during the evaluation of the impact of ICT on energy consumption a lack of flow rate data impeded to calculate the heating consumption of each floor. This issue was overcome by monitoring the flow rate for a period of five days. These monitored data were then used to identify a series of Auto-Regressive with Exogenous Input (ARX) models capable to estimate the system behaviour. The ARX model structure has been found to be the simplest one to find analytical solutions with excellent performance and accurate description of the physical model (Ljung, 1999). It is also widely applied in different research fields. For example, in (Ismail et al., 2011) an ARX model was identified to describe the partial input-output data of a heating process in a steam distillation essential oil extraction system. In (Yun., 2012) an ARX time and temperature indexed model was generated to predict the building thermal load.
In this paper, a method to identify the water flow rate of a heating system in order to estimate its heating consumption is shown. System

identification deals with the problem of describing a physical system with mathematical equations derived from measured data. In this context, the equations that describe the physical system are also referred to as "black box model".

Important steps to identify a black box model are (Rabbani et al., 2013):

- Analysis of the monitored data;
- Selection of a family of models;
- Identification of a model that belongs to that family;
- Validation of the model.

Each of these stages have a high impact on the success of the identification process.

The identification of the models has been done using both the R programming language and the System Identification Toolbox of Matlab in order to minimize the root mean square error between the measured and predicted values, analyse and evaluate the models.

The crucial step has been the model validation. We have indeed carried out a series of tests in order to check that the identified models are suitable and to select the best one among them.

Tests included computing the coefficient of determination, autocorrelation of residuals, cross-correlation of residuals with input parameters, comparing the coefficient of determination over the training period with that over the validation period in order to detect overfitting.

Concerning the case study, the building has three main floors with offices and laboratories and an unheated underground floor used for laboratories and the thermal power station. After an energy audit, only the first and second floor (offices) were equipped with a monitoring system leaving the ground floor (laboratories) out of the analysis.

The heating and cooling facilities are:

- A district heating system with heat exchanger and heat tank, supplying hot water to the building;
- An external refrigeration plan and storage system to ensure the supply of cold water during the cooling season;
- 2 pipe fan-coils with manual temperature and fan control in all offices;
- radiators and ceiling fans in all laboratories;
- One Air Handling Unit (AHU) providing the labs with sensible and latent heat.

The thermal parameters measured were divided in three categories:

- Indoor Environmental Quality (IEQ) related parameters; indoor temperature, humidity, CO_2, occupancy and illuminance, measured in five offices (two on the first floor and three on the second floor);
- Climate parameters; outside temperature, global radiation, wind speed and humidity, from a weather station already installed on the building;
- Heating and cooling system parameters; water supply and return temperatures for first floor, second floor and bathrooms, water supply and return temperatures of the hot and cold AHU coils, temperature and humidity of the air supply duct.

The flow rate was monitored using a portable flow rate device (Type: Dynasonic® ultrasonic flow meter) during the period from 5/12/2013 to 10/12/2013.

Table 1 Characteristics of the building

PARAMETER	QUANTITY	MEAS. UNIT.
Conditioned Volume	3825	m^3
Surface of each floor	425	m^2
Number of floor	3 conditioned floors + 1 underground floor not conditioned	
Intended use	Laboratory/Office	
Type of wall	Normal brick without insulation	

2. Simulation

2.1 Data set

The monitored data are summarized in Table 1.

Table 2. INPUT/OUTPUT used for model estimation

NAME	SYMBOL	MEAS. UNIT.	Uncert.	INPUT/ OUTPUT
Flow_rate 1Floor	FR_1F	m^3/h	±2 %	OUTPUT
Flow_rate 2Floor	FR_2F	m^3/h	±2 %	OUTPUT/ INPUT
T_supply 1Floor	Tsupp_1F	°C	±0.3 °C	INPUT
T_return 1Floor	Tret_1F	°C	±0.3 °C	INPUT
T_supply 2Floor	Tsupp_2F	°C	±0.3 °C	INPUT
T_return 2Floor	Tret_2F	°C	±0.3 °C	INPUT
T_supply Bath	Tsupp_B	°C	±0.3 °C	INPUT
T_return Bath	Tret_B	°C	±0.3 °C	INPUT
External T	T_ext	°C	-*	INPUT
Global radiation	GR	W/m^2	-*	INPUT
Average Occup._BUI	Occ_BUI	[-]	-	INPUT
Average Temp._BUI	T_BUI	°C	±0.3 °C	INPUT

* Measurements taken by the weather station situated at "Fondazione Edmund Mach"

2.2 Reliability of the measured data

As first step the quality of the data measured by the sensors was assessed. Each physical quantity has a true value that is not observable. During the measurement of a physical quantity it is common that errors occur. These can be of two types:

- Systematic errors, such as device accuracy, device construction defects and wrong sensor usage;
- Random errors due to uncontrollable factors, such as device precision, and variation of internal and external environmental conditions.

As ARX models are based on the assumption that the behaviour of a physical system varies about a stationary (not changing with time) working point, the first step is to verify the stationarity of the time series (Andrews, 2013). Indeed, a stationary time series is a series of successive measurements for which mean and variance are constant over time and the auto-correlation function depends on the lag alone. Therefore, by stationarizing the time series one is able to obtain meaningful sample statistics such as mean, variance, auto-correlation, and cross-correlation with other variables. Such statistics are useful as descriptors of future behaviour only if the series is stationary.
To assess the stationarity of the flow rate time series, the Augmented Dickey-Fuller (ADF) t-statistic test was carried out.
This test uses the following regression model:

$$y'_t = \theta y_{t-1} + \beta_1 y'_{t-1} + \beta_2 y'_{t-2} + \cdots + \beta_k y'_{t-k} \qquad (1)$$

Where:
y'_t: is the differenced series ($y'_t = y_t - y_{t-1}$)
k: is the maximum lag
If the null hypothesis $H_0: \theta = 0$ of the ADF test is rejected, the data is stationary and does not need to be differenced.

Table 3 shows the ADF test for flow rates of the 1st and 2nd floor:

Table 3 Augmented Dickey-Fuller (ADF) test for Fl_2rate1f and Fl_rate2f

Data	ADF test	Lag order	p-value
Fl_rate1f	-4.2143	4	0.01
Fl_rate2f	-4.0124	4	0.01

According to the t-statistic, the p-value is lower than the selected significance level of 5 % for both variables. The two time series can therefore be considered stationary and do not need to be differenced.
The mean can then be subtract from the data:

$$Z = x - \mu \qquad (2)$$

Where x is the samples variable and μ is the mean. Indeed, for steady state data it is reasonable to assume that the mean corresponds to a physical equilibrium and the aim to build linear models is to describe deviations from this equilibrium, which are responses to excitations of the physical system (Ljung, 1999).

2.3 Methodology

The flow rates were measured every hour for five days, for a total of 120 samples. The data were subsequently divided in two parts. The first part was used for the model calibration and the second one for the validation. We varied the amount of data used for calibration. The best models were obtained using 70 % of the data for calibration and the remaining 30 % for validation (Mourad et al., 2005). The estimation of the ARX model was done through the application of a MATLAB script that tries out and evaluates different input combinations and polynomial orders. The best model among all identified models was then chosen as follows. For each model, we computed three performance indicators: i) the coefficient of determination over the calibration period R^2_{cal}, ii) the coefficient of determination over the validation period R^2_{val}, and iii) the absolute difference between the two coefficients of determination $\Delta R^2 := |R^2_{val} - R^2_{cal}|$. In addition, all residuals autocorrelation values up to lag 20 had to be inside the 95 % confidence interval. Next, we considered only those models as valid that satisfied the following inequalities:

- $R^2_{val} > 0.7$;
- $\Delta_{R^2} < 0.1$.

At the end, the achieved list was sorted by descending R^2_{val}.
The first three models in the list were investigated further as follows:

- Computation of the autocorrelation to lag 20. If all values were inside the 95 %-confidence interval, we could assume randomness;
- Visual inspection of a scatter plot of residuals against fitted values. The points should be randomly distributed around the zero line and form a horizontal band, showing that variances of the residuals are equal and that there are no outliers;
- Computation of the cross-correlations between residuals and inputs, to see if a specific input generates a pattern in the residuals;
- Visual inspection of the distribution of the residuals with a QQ-plot and a histogram;
- Multicollinearity analysis through Besley collinearity diagnostics.

2.4 Results and discussion

The first model identified is the flow rate of the second floor (FR_2F). The output of this model is subsequently used to identify the flow rate model of the first floor (FR_1F). This choice was done because the two circuits are directly connected to the same collector and because the flow rate profile of the first floor follows the profile of the second floor.
The inputs used are shown in Table 4.

Table 4 Second floor model inputs

Model	**Inputs**
Mod_1.2F	Tsupp_2F; Tret_2F; Tret_B; T_ext; T_BUI;
Mod_2.2F	Tsupp_2F; Tret_2F; Tret_B; T_ext; T_BUI; Occ_BUI
Mod_3.2F	Tsupp_2F; Tret_B; T_ext; T_BUI;

The performance indicators for the best three models are shown in Table 4.

Table 5 Coefficients of determination of 2nd floor flow rate models

Model	R^2_{cal}	$R^2_{Adj\_cal}$	R^2_{val}	$R^2_{Adj\_val}$	Δ_{R^2}
Mod_1.2F	0.903	0.898	0.855	0.847	0.048
Mod_2.2F	0.904	0.898	0.847	0.839	0.057
Mod_3.2F	0.916	0.911	0.847	0.838	0.057

The low values of Δ_{R^2} in Table 5 show that models do not overfit the data. Indeed, the coefficients of determination and validation are very similar, and the model can predict in the best way the data as shown in Figure 12.

The first model in Table 5 has the best $R^2_{Adj\_val}$ and Δ_{R^2}. However, the other two models have values very close to the first one. Therefore, the best model is chosen depending on the residual analysis.

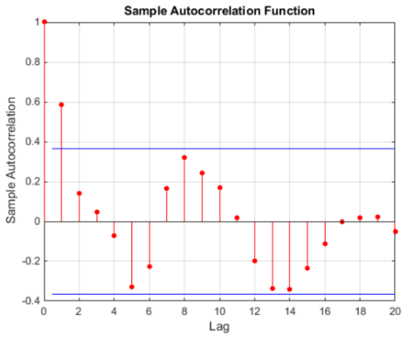

Figure 1 ACF Mod_2.2F

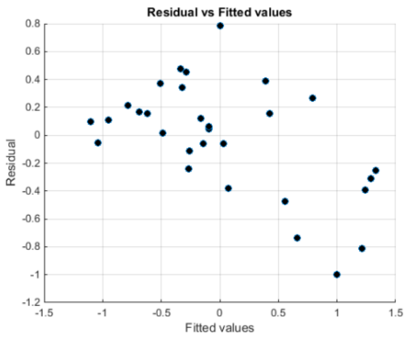

Figure 2 Residuals vs fitted values Mod_2.2F

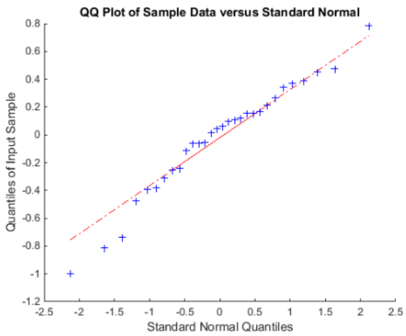

Figure 3 QQ plot residuals Mod_2.2F

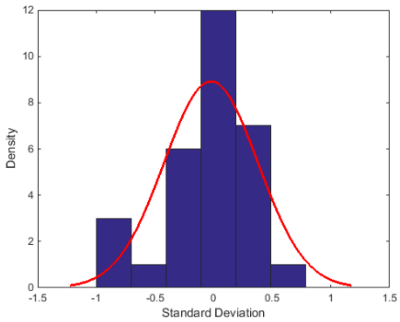

Figure 4 Normal distribution residuals of Mod_2.2F

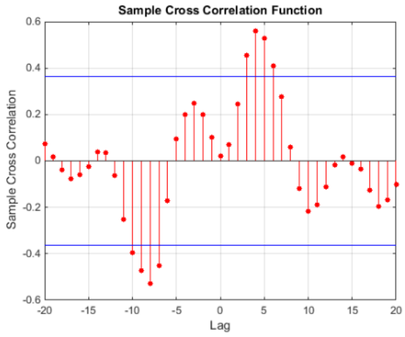

Figure 5 CRF Residuals vs Occ_BUI for Mod_2.2F

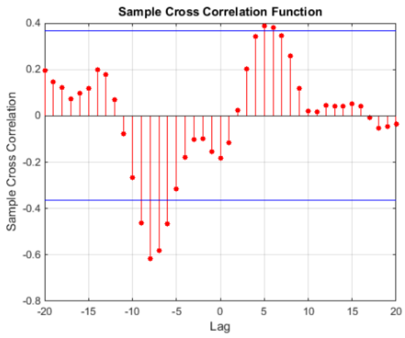

Figure 6 Residuals vs T_BUI for Mod_2.2F

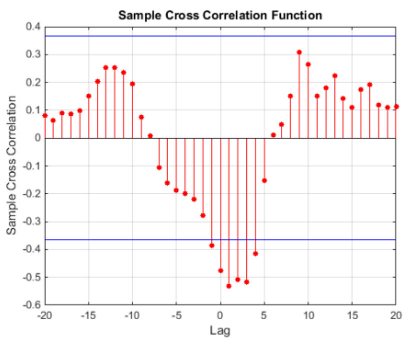

Figure 7 Residuals vs Tret_B for Mod_2.2F

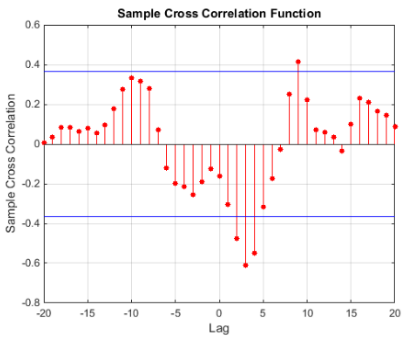

Figure 8 Residuals vs Tret_2F for Mod_2.2F

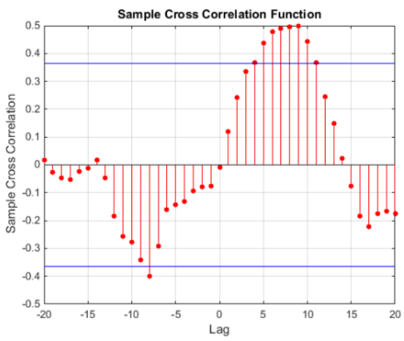

Figure 9 Residuals vs T_ext for Mod_2.2F

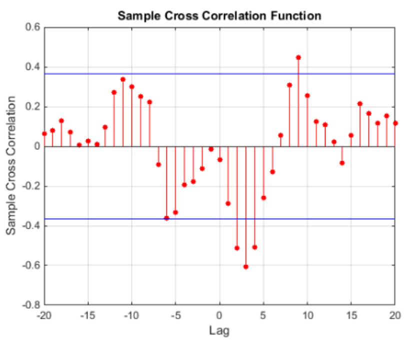

Figure 10 Residuals vs Tsupp_2F for Mod_2.2F

Table 6 Belsley multicollinearity test

CI	S-value	Tsupp_2F	Tret_2F	Tret_B	T_ext	T_BUI	Occ_BUI
1	1.956	0.004	0.002	0.015	0.000	0.012	0.008
2	1.130	0.001	0.000	0.024	0.349	0.049	0.002
3	0.707	0.002	0.000	0.255	0.437	0.162	0.012
4	0.486	0.106	0.010	0.569	0.064	0.201	0.002
5	0.371	0.054	0.000	0.018	0.003	0.561	0.600
13	0.146	0.833	0.887	0.117	0.146	0.015	0.374

The multicollinearity test shows that, although the variance decomposition proportions are high in the last line of Table 6 for Tsupp_2F and Tret_2F, the S-value is never below 0.1 and the Condition Indices (CI) are low. Indeed, the latter have a significance only if higher than 30, as specified in Belsley et al., 1980.

Based on the residual analysis results, the model that best represents the flow rates of the 2nd floor is Mod_2.2F. The estimated model equation is:

$$FR2F\ (t) - 0.718\ FR2F\ (t-1) + 0.1643\ FR2F\ (t-2) - 0.06596\ FR2F\ (t-3) + 0.1032\ FR2F\ (t-4) - 0.1105\ FR2F\ (t-5) + 0.06817\ FR2F\ (t-6) + 0.007497\ FR2F\ (t-7) = -67.98\ Tsupp2F\ (t) - 13.73\ Tsupp2F\ (t-1) + 83.56\ Tret2F(t) - 30.42\ Tret2F(t-1) + 3.5820\ TretB\ (t) - 9.6160\ TretB\ (t-1) - 67.98\ Text\ (t) - 13.73\ Text\ (t-1) - 23.14\ TBUI\ (t) - 11.1200\ TBUI\ (t-1) + 99.98\ OccBUI(t) - 28.59\ OccBUI(t-1) \tag{3}$$

Table 7 Distribution of model inputs

Quantile	Tsupp_2F	Tret_2F	Tret_B	T_ext	T_BUI	Occ_BUI
Min	23.15	24.31	16.01	1.40	17.20	0.00
25%	27.27	28.55	18.55	5.65	18.21	0.00
Median	47.52	39.31	37.33	8.00	19.30	0.00
75%	63.54	57.54	57.14	9.10	20.09	0.20
Max	69.41	63.38	61.96	12.50	21.22	0.78

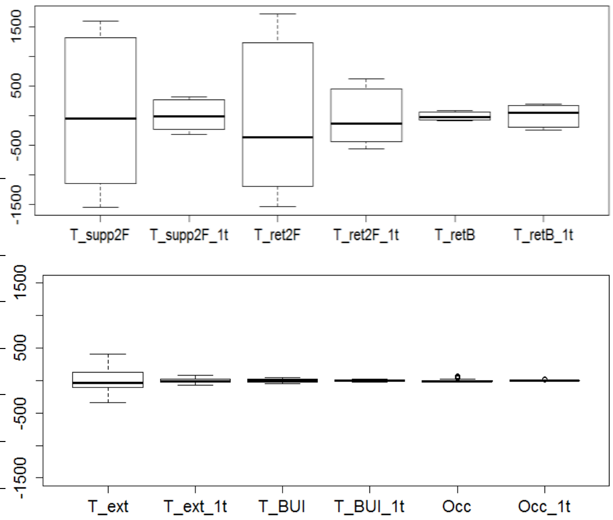

Figure 11 Impact of each input term in model equation (3)

The boxplot (Figure 11) shows that the most important parameters influencing the flow rate are the supply and return temperature of the 2nd floor, the return temperature of the bathrooms and the external temperature. This assertion is confirmed by the fact that the system is controlled by an external temperature probe.

The model calibration and validation plots are shown hereafter.

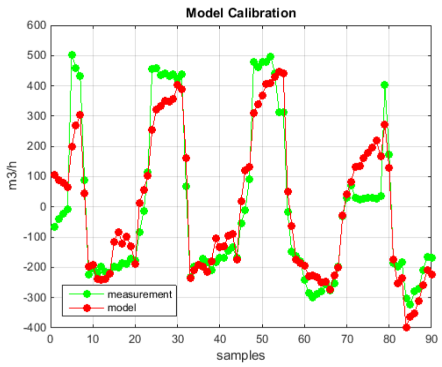

Figure 11 Calibration of Mod_2.2F

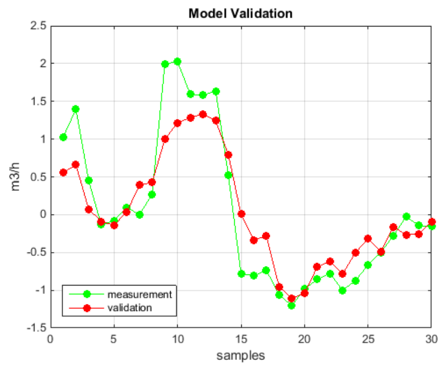

Figure 12 Validation of MOD_2.2F

An analogous process was carried out to identify a model for the first floor, with similar results.

Table 7 1st floor model inputs

Model	Inputs
Mod_1.1F	Tret_B; GR; FR_2F;
Mod_2.1F	Tret_B; FR_2F;
Mod_3.1F	Tret_2F; Tret_B; GR; FR_2F;

Table 8 Determination coefficients of first floor model

Model	R^2_{cal}	$R^2_{Adj\_cal}$	R^2_{val}	$R^2_{Adj\_va}$	Δ_{R^2}
Mod_1.1F	0.981	0.980	0.989	0.989	-0.008
Mod_2.1F	0.980	0.979	0.988	0.988	-0.008
Mod_3.1F	0.982	0.981	0.988	0.987	-0.005

Based on the results of Table 8 and on the residuals analysis, the best model chosen to evaluate the flow rate for the 1st floor is Mod_2.1F. Its model equation is:

$$FR1F(t) - 0.1492\ FR1F\ (t-1) + 0.0084\ FR1F\ (t-2) - 0.0503\ FR1F\ (t-3) + 0.0595\ FR1F\ (t-4) - 0.0387\ FR1F\ (t-5) - 0.0245\ FR1F\ (t-6) - 0.04106\ FR1F\ (t-7) = 33.6900\ FR2F\ (t) + 0.8957\ TretB\ (t) \qquad (4)$$

The following figures show the model calibration and validation for the FR_1F.

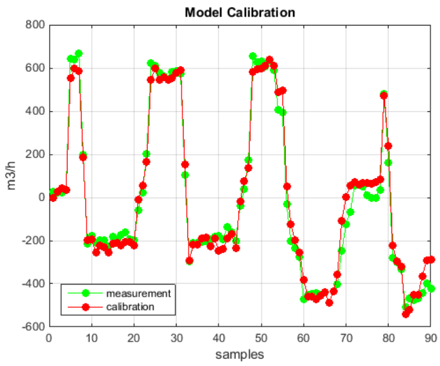

Figure 13 Calibration of Mod_2.1F

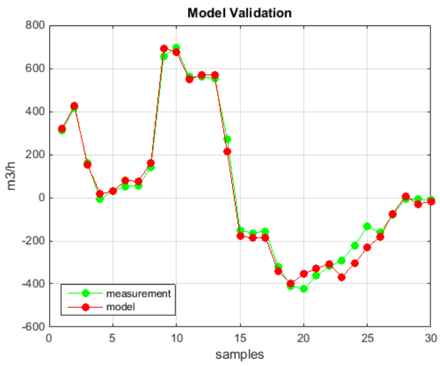

Figure 14 Validation of Mod_2.1F

The estimated models for both flow rates were used to find the heating consumption of the San Michele building for February, March, April, October and December 2013 ("heating season" in Table 9). In the other months, it was not possible to use the model due to a modification of the system setting. In November, billing data is not available.
Table 9 shows the comparison between the utility meter (normalized by the area of all 3 floors: 1425 m²) and the modelling values (normalized by the area of the 1st and the 2nd floor: 950 m²).

Table 9 Real and model values of building heating consumption

Season	Real value [kWh/m²]	Modelling value [kWh/m²]
Heating season	77.02	85.32
Average	15.40	17.06

The estimated average heating consumption differs from the real consumption by less than 10 %. This value is influenced also by the fact that the ground floor was not monitored, therefore reducing the prediction accuracy. Additionally, due to the presence of only few offices on the ground floor it is expected to have a small heating requirement compared with the first and second floor.

3. Conclusion

We showed how system identification can help to estimate the water flow rates in a central heating system (and the relationship with influencing variables) and subsequently the heating consumption. We defined and demonstrated a methodology aimed at identifying the best model among the many generated. The results show that the identified model reliably estimates the measured heating consumption. However, there is room for improvement. An error model could increase the quality/fitness of the model and reduce the cross correlation of residuals with inputs.

References

[1] ICT PSP Smart-Build project http://www.smartbuild.eu/ 10/11/2014

[2] L. Ljung, 1999. "System identification: Theory for the User, 2nd ed. New Jersey: Prentice Hall PTR

[3] N. Ismail, M.H.F. Rahiman, M. N. Taib. 2011. "Investigation of ARX Model on Partial Input-Output Data in Heating Process" IEEE Control System Graduate Research Colloquium 978-1-4577-0339

[4] K. Yun, R. Luck,P.J.Mago, H.Cho .2012. "Building hourly thermal load prediction using an indexed ARX model" Energy and Buildings 54(2012):225-233.

[5] M.J.Rabbani, K.Hussain, A. jhan, A. Ali. 2013. "Model Identification and Validation for a Heating System using MATLAB system Identification Toolbox " ICSICCST:Material Science and Engineering 51 (2013). doi :10.1088/1757-899X/51/1/012022

[6] B.H.Andrews,M.D. Dean, R.Swain, C.Cole. 2013."Builidng ARIMA and ARIMAX Models for predicting Long-term Disability Benefit Application Rates in the Public/Private Sectors". Society of Actuaries Health Section

[7] R.Nau, Fuqua School of Business Duke University, "stationarity and differencing" http://people.duke.edu/~rnau/411diff.htm 10/11/2014

[8] OTexts, online open access textbooks, "stationarity and differencing" https://www.otexts.org/fpp/8/1 10/11/2014

[9] M. Mourad, J.Bertrand, L.Krajewski, G. Chebbo. 2005. "Calibration and validation of multiple regression models for stormwater quality prediction: data partitioning, effect of dataset size and characteristics". Water Science & Technology 52(2005): 45

[10] D. A. Belsley,K. Edwin,E.W.Roy 1980. "Regression Diagnostics: Identifying Influential Data and Sources of Collinearity". New York: John Wiley and Sons.