OPTIMIZED METHOD TO PREDICT ENERGY IN A MICROGRID

Luc Dufour[1], Dominique Genoud[1], Bruno Ladevie[2], David Wannier[1]
Information System institute[1], CNRS UMR 5302[2]
TechnoPôle 3, 3960 Sierre, Suisse [1], Campus Jarlard,, Route de Teillet, 81000 Albi, France[2]
E-Mail: (luc.dufour, david.wannier,dominique.genoud)@hevs.ch [1] ; (bruno.ladevie, cendrine.gatumel)@mines-albi.fr[2]

1 SUMMARY

Key Words: Optimized Machine learning, Data mining, Energy prediction, Microgrid

This work presents a data-intensive solution to predict the different energy flux in a microgrid. The ability to predict locally the energy considering meteorological uncertainty can play a key role in the management of microgrid. Our approach is to provide an easy implemented and flexible solution to predict the consumption and the production at the building level based on the machine learning technique and tested on the real use cases in a residential and tertiary sector. A new evaluation of the consumption is realized: the point of view is energy and not only electrical. The energy consumption is decomposed between the heating, the hot water, the electrical devices, the lighting and the ventilation consumptions. A prediction every hour is provided to create scenarios to pilot devices. We provide an original optimized method to limit the input data number. A classification and a prediction for the electrical consumption in a tertiary sector are provided for the next hour with an accuracy over the 98% +/- 1%.

2 INTRODUCTION

PV is the fastest-growing energy technology since 2002 with an average increase of 48% [COC-2012]. The majority plants are grid-connected systems and a high penetration of PV (like in the case of islands) raises issues for the grid-operators. Subsequently, forecasting the power output of the PV plant is necessary to assure the grid stability. A large amount of research studies in the domain lay emphasis on predicting solar radiation which is a key data to improve the results. Neural networks are widely used to this purpose and manage to reach a MAPE around 7% on monthly and day-ahead solar radiation forecast [COC-2012]. We can cite the example of A. Mellit, a reference author in the subject who achieve a MAPE less than 6% for day-ahead solar radiation in Algeria [MEL-2006]. Regression trees are not widely used but show a MAPE of 33% for PV production prediction [Nom-2011]. The studies concern a wide choice of photovoltaic installations and data type, (study on 12 PV cells in the laboratory or 14,000 m2 of panels) with ARIMA models or neural networks [ZAM-2014] [GRA-2016] [GUL-2016].

The consumption prediction models are very important, time scales are multiples and the predictions are made by the hour, the day, the month, the year, and so on. [REB-2015] [MOU-2016] [CHE-2016]. In [ZHA-2016], the consumption time series is predicted by a hybrid method with a regression followed by a learning method and the SVM algorithm. A disaggregation's of the global load curves are also performed to identify the electrical appliances [PAR-2011]. Other studies create simplify models and the building is modeled by a single equivalent electrical circuit (RC, R2C2, R3C2, R6C2) [JAZ-2002] [ARB-2015].

3 METHODOLOGY

3.1 Energy problem Characterization

We begin by a decomposition of the global energy need for a system without the heating, the hot water, the lighting, the ventilation and the specific electrical named E_{USE} :

$$E\ t(i) = \left(E_{Heating} + E_{Hot\_Water} + E_{USE} + E_{Ventilation} + E_{Lighting}\right) t(i) \qquad \textbf{(eq. 1)}$$

The system can be connected at the decentralized production. With E_{ENRE} t(i) the photovoltaïque solar pannel production and E_{ENRT} t(i) the thermal solar pannel production, we have:

$$E\ t(i) = \left(E_{Heating} + E_{Hot\_Water} + E_{USE} + E_{Ventilation} + E_{Lighting} - E_{ENRE} - E_{ENRT}\right) t(i) \qquad \textbf{(eq. 2)}$$

A part of these energy can be controlled based on the different prediction. The system can have an electrical storage battery :

$$E\ t(i) = \left(+/-E_{Heating} +/- E_{Hot\_Water} +/- E_{USE} +/- E_{Ventilation} +/- E_{Lighting} +/- E_{Battery} - E_{ENRE} - E_{ENRT}\right) t(i) \qquad \textbf{(eq. 3)}$$

3.2 Transform Datas

The data collected depends of the sensor selected to measure the consumption and the production which influence the data process. In the case of the identification of the electrical devices, a study per second is required for the electrical devices with short cycle times. In this paper, we apply a median filter of size 60 to remove high frequency at the global load curve level. In a case of no stable electrical network, it may ne possible to complete by a data process based on the voltage [HAR-1992].

3.2.1 Events detection

The power thresholds identify the short variations during a predetermined period. These techniques enable identification of the electrical events. It is between 40 Watts and 50 Watts 15 VAR for reactive power [HART-1992] to identify all electrical devices. This events identification can be done from a sliding average over a window of time to be defined according to the problem.

$$Events\ t(i) = [Date, \Delta P\ t(i), \Delta Q\ t(i)] \qquad \textbf{(eq. 4)}$$

3.2.2 Statistics

The statistics patterns are computed for the powers and the events. The mean, the minimum, the maximum, the standard deviation are computed each hour.

3.2.3 Time characterization

The different variables from the time variable can be extracted to have statistics per period with the month ,m ranging from 1 to 12, the days of the year from 1 to 365, the days of the week from 1 to 7 (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday), the hour h from 0 to 23 and the days corresponding to cantonal school holidays characterized by a Boolean variable 0 or 1.

3.2.4 Production planning integration

the consumption prediction, In the tertiary and industrial sector, the arrival and departure times of staff and the days worked as input information help in the quality of the training set by creation of specific training set. The production planning enables to identify the people number adjusted with the production. For example, in the restaurant case, the restaurant may be open for the lunch and not for the dinner and open the weekday and not the week-end. These information's enable to create a specific training by period. In the residential case, a production planning associated with the heating and the hot water is added. pump is associated. This heat pump controls the heating and the hot water. A training model is created for days k from Monday to Friday (from 1 to 5) and another model for Saturday and Sunday (6 and 7).

3.2.5 Meteo Data integration

For the collection of meteorological data, we analyzed like World wea. ther online, Open Weather map, Meteoblue, Meteosuisse, Observations, Meteocentrale. The advantage of suppliers such as Meteoblue, Meteosuisse, Meteonews and Meteocentrale is the control of data by meteorologists which adjust them according to the changing meteorological context. The Meteosuisse is the solutions this project. the dataset contains historical real measures and forecast values of temperature and radiation. The forecasted weather values are a one day-ahead prediction in 2014.

3.2.6 Classification

For the hot water and the heating, a supervised training is computed. If the energy is above a threshold, it is labeled ON, otherwise OFF. The same learning is done for the hot water per hour.

In the solar production and the others energy flux, the number of classes has not been predetermined and we are talking about unsupervised learning. These techniques determine the classes to characterize the electrical consumption and the solar panel production. In this paper, the EM (expectation-maximization) algorithm is proposed to decompose the time series [VIC-2015] and create the classes represented by a Gaussian.

3.3 Train model

Our objective is to provide the prediction in the next hour to manage the microgrid:

$$E\,t(i+1) = \left(+/-E_{Heating}\ +/-\ E_{Hot\_Water}\ +/-\ E_{USE}\ +/-\ E_{Ventilation}\ +/-\ E_{Lighting}\ +/-E_{Battery}\ -\ E_{ENRE}\ -\ E_{ENRT}\right) t(i+1) \quad \textbf{(eq. 5)}$$

The solutions are provided by the software KNIME with an industrial deployment. A connection between the database and the software is created and the data are proceeding each hour. The techniques predictions described in Introduction are tested with the linear method (ARIMA model and linear regression) and the no linear method (MLP, PNN, SVM, Random Forest and the GBT, Boosted Tree gradient).

We must ensure an equitable distribution among the different classes, especially that which requires a particular interest, such as the peak power consumption or production. A step called "boostsrap" can be necessary if classes are underrepresented. This step is particularly adapted to our problem of predictions of sudden variations of the consumption or the production and that algorithms like the AdaBoost or the Gradient Boosting Tree. Finally, different training model are represented by the planning production described in 3.2.3.

3.4 Evaluation

In our binary problem, the evaluation is realized by the Receiver Operating Characteristic (ROC) curve. It is a measure of the performance of a binary classifier. In no binary problem, the evaluation is realized by statistics computed. the coefficient of determination (R^2), the Mean Square Error (MSE), the RMSE (root mean square deviation), the Mean Absolute Error (MAE), the mean squared difference (MSD) [BER-2010].

3.5 Optimization

3.5.1 Optimized the weather data predicted

Meteorological data are essential in particular for the solar and the heating predictions. The prediction values provided by MeteoSwiss included extreme luminosity and temperature errors corresponding to local phenomena. An error is defined between the real and the predicted data for the temperature predictions and an error for the luminosity. A training model of the error associated with the time identification described in 3.2.2 is characterized by a non-parametric method.

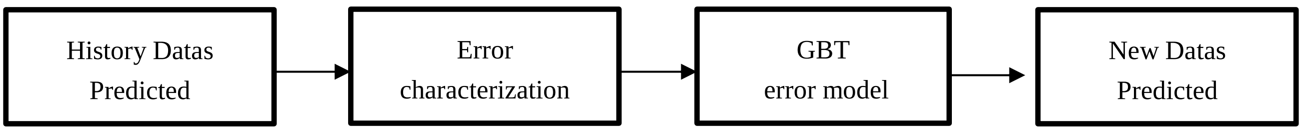

Figure 1: Method to optimize the weather data predicted

3.5.2 Variables number impact

The aim is to define a minimal inputs variable without decrease the prediction scored. Different algorithms are tested and evaluated in 2.4. At the output of the different implementations of the decision trees, the statistics provide the construction of the tree and enable a hierarchy in the input variables. The sum of statistics is computed for the construction of the three first levels. By iteration the number of the input variables is reduced based on their importance in the construction of trees. At each iteration, the scored of the prediction or the classification is computed describe in 3.4.

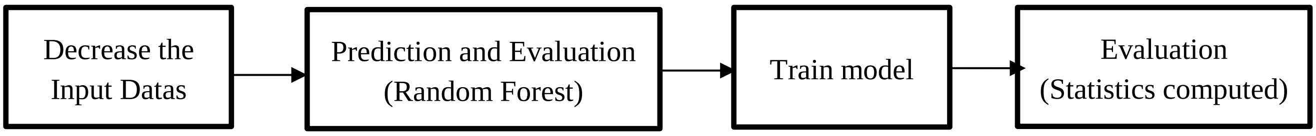

Figure 2: Method to optimize the weather data predicted

3.5.3 Frequency impact

The difference the prediction results is computed between the data at the second aggregate per minute and one data per hour to predict the next hour.

4 USE CASES

Techno-Pôles microgrid is the use case contextualized in the I-BAT Swiss Project and ENTROPY European project. The data are displayed in real-time through: http://www.technopole-vert.ch. An Advanced Metering Infrastructure (AMI) based on the Internet of Things (IoT) has been implemented. This deployment provides energy-related parameters such as the overall building load curves and a wireless network of IoT-based smart meters to measure and control appliances. The Techno-Pôle of Sierre, the sunniest city in Switzerland as a 203 kWp PV plant that represents 1200 m2 of the roof surface. A weather station has been recently installed in 2015 and will provide more accurate weather data for the microgrids energy management. The site gathers 500 people working for 50 companies including private service providers as well as research institutes like HES-SO which carried out the microgrid project. The microgrid can also operate as an energy storage management demonstrator as batteries of 25 kWh with a remote control of charge/discharge have been installed.

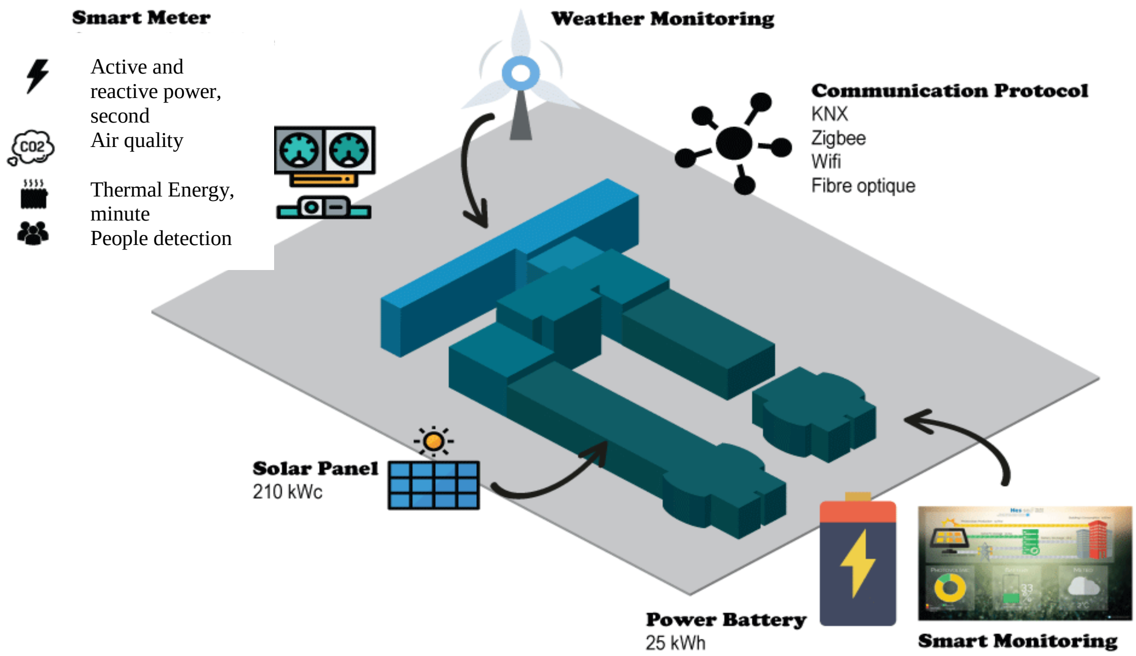

Figure 3: Statistical results at the output of the different algorithms for the prediction of overall electricity

The metering infrastructures provide at the one hertz frequency parameters (load curves from the photovoltaic plant provided by ELKO, and the grid consumption provided by Sierre-energy) and high frequency parameters (devices measures from the Ecowizz Zigbee smart meters). The information system contains the elements necessary for the storage of data via NO-SQL as the data is formatted in JSON which connected with Knime software for the prediction.

In this paper, we present our result through the restaurant of this technopole. In the tertiary sector, days and hours outside the production schedule are removed from our training and test data set. For the restaurant, the weekend is removed and only the hours between 6am and 5pm are represented in our training dataset corresponding to the production planning. For the restaurant, the electrical consumption includes the ventilation, the lighting, and the specific electrical devices like the oven (eq. 6). The heating and the hot water are provided by the radial panel.

$$E\ t(i+1) = \left((+/-\ E_{USE}\ +/-\ E_{Ventilation}\ +/-\ E_{Lighting}) - E_{ENRE}\right) t(i+1)\ +/- E_{Battery} \qquad \textbf{(eq. 6)}$$

The dataset contains the data from September 1st 2013 to October 31st 2014 and the October 2014 is the test set. In output of the classification, each hour is labeled by a class c. A historic of each input variables for the last two days is created to identify one hour.

$$\begin{aligned}
&\mathrm{E}\quad \mathrm{t(i)} = [\mathrm{Date}\ ;\ \mathrm{Statistics\ Power}_{\mathrm{Phase\ 1}}\ \mathrm{t(i)}, \ldots, \mathrm{Statistics\ Power}_{\mathrm{Phase\ 1}}\ \mathrm{t(i-48)},\\
&\mathrm{Statistics\ Events}\ \mathrm{t(i)}, \ldots, \mathrm{Statistiques\ Events}\ \mathrm{t(i-48)}\\
&\mathrm{Weather}_{\mathrm{Real_Datas}}\ _\mathrm{City}_{Sierre}\ \mathrm{t(i)}, \ldots., \mathrm{Weather}_{\mathrm{Real_Datas}}\ _\mathrm{City}_{Sierre}\ \mathrm{t(i)}\ \mathrm{t(i-48)},\\
&\mathrm{Weather}_{\mathrm{New_Predicted_Datas}}\ _\mathrm{City}_{Sierre}\ \mathrm{t(i+1)}, \ldots, \mathrm{Weather}_{\mathrm{New_Predicted_Datas}}\ _\mathrm{City}_{Sierre}\ \mathrm{t(i-48)}\\
&\mathrm{Year}_{\mathrm{j}}, \mathrm{Month}_{\mathrm{m}}, \mathrm{Day}_{\mathrm{j}}, \mathrm{WeekDay}_{\mathrm{js}}, \mathrm{Hour}_{\mathrm{h}}, \text{Jour fériés}_{0.1}, \text{Vacances scolaires}_{0.1}, \mathrm{Classes}_{\mathrm{c}}\ ]
\end{aligned} \qquad \textbf{(eq. 7)}$$

5 RESULTS

5.1 Evaluation of the classification

At the end of unsupervised learning, the electricity consumption for the restaurant is represented by six classes (Table 38). The Class 6 represents the peak power of the restaurant during the considered period. The Gradient Boosted Tree best classifies the six classes representing power consumption (Table 1) with 98% +/- 0.6 accuracy.

Table 1 : Electricity consumption described by gaussian

Energy	Clusters	Mean (kWh)	Standard deviation	Repartition %
Electrical Energy consumption, tertiary Sector	0	2.304	0.535	10
	1	4.448	1.29	24
	2	8.961	2.392	18
	3	16.622	2.539	18
	4	22.982	2.449	9
	5	28.246	1.927	35
	6 : Power peak	31.866	4.367	2

Table 2 : Classification results for the electricity load curve

Models	Classification results	Standard deviation
PNN	84.6 +/- 1.2	1.6
MLP	88.8 +/- 1.3	1.5
SVM	91.8 +/- 1.1	1.6
Random Forest	96.1 +/- 0.7	1.5
Gradient Boosted Tree	98.0 +/- 0.6	0.8

5.2 Evaluation of the prediction

For the electricity consumption, the decision tree and the Gradient Boosted Tree provides the best results. The average difference in absolute value is 1,162 kWh for the GBT (Table 1). The precision of the models is explained by the repetition at specific times of stages to prepare the meals. A priori data collected by the restaurant as the number of customers planned or special orders (eg company parties) can improve the accuracy of our mathematical models.

Tablle 3 : Comparison of the Statistical results per algorithms for the Electrical consumption prediction for the restaurant

Models	MAE	MSE	RMSE	MSD	R^2
MLP	1.9	5.4	2.1	0.2	0.64
PNN	1.9	5.4	2.1	0.2	0.64
Linear Regression	2.3	5.4	2.1	0.2	0.64
ARIMA	2.2	5.4	2.2	0.2	0.64
RF	1.5	4.5	2.1	0.1	0.95
GBT	**1.2**	**4.3**	**1.9**	**0.1**	**0.98**

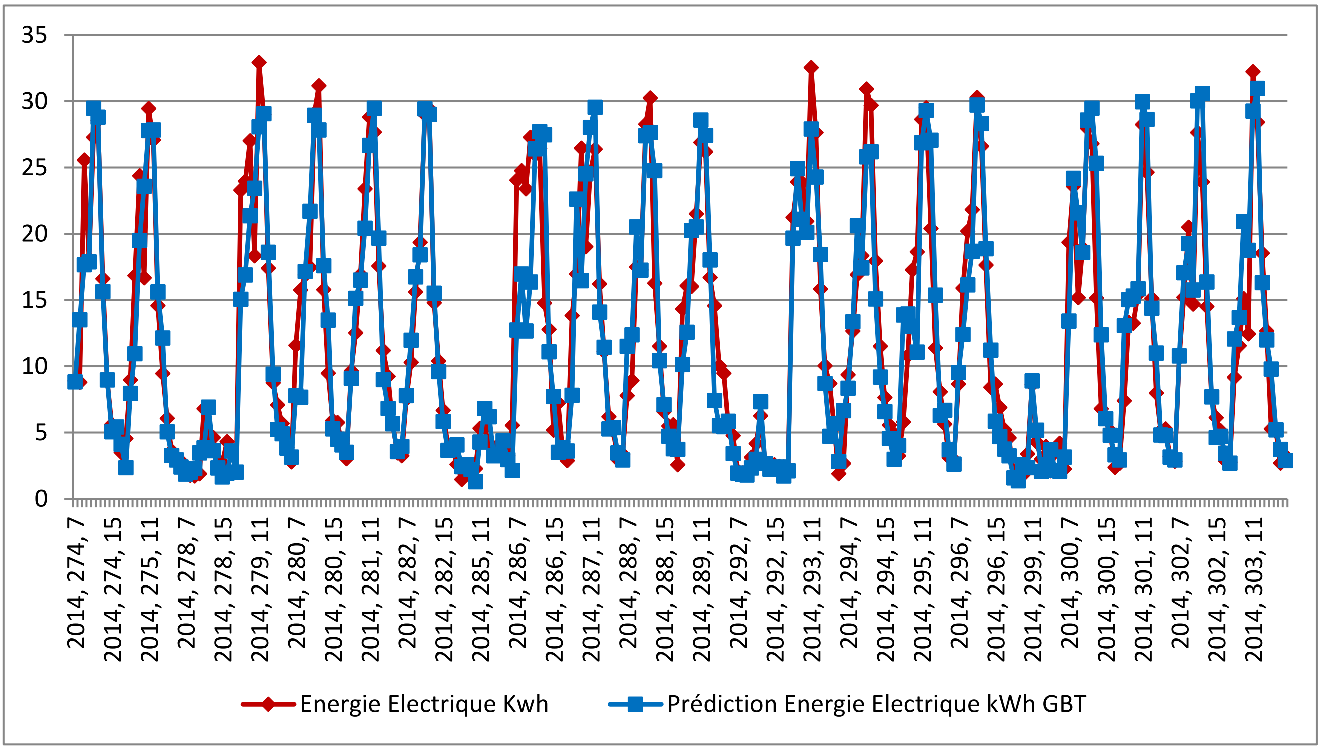

Figure 4: Electrical consumption prediction for the restaurant with the Gradient Boosting Tree (Year, Day, hour)

5.3 Optimization results

5.3.1 Optimization meteo data

To characterize the error, a history based on the actual and predicted data is created. This history represents the input data of our mathematical model. The training data correspond to measurements from January 1, 2010 to August 31, 2012. Test data are from September 1, 2012 to March 31, 2014.

Table 5 : Characterization of the errors for the luminosity before and after our method

Variables number	Min	Mean	Median	Max	Std. Dev
History error	-910	5.5	0.1	1116	304
New error	-550	7.1	0	510.	203

5.3.2 Variables number optimization

In this test, our objective is to determine the impact of the quantity of our input data in the variation of the result. The 15 best variables to predict the electrical consumption prediction for the restaurant is the standard deviation t (-1), the standard deviation t (-2), the minimum power t (-1), the median power t (-1), the standard deviation Median Active power t (-1), the maximum events t (-1), the maximum Variations active power t (-2), the standard deviation events t (-1), the Minimum events t (-1), the Standard deviation Median events t (-1).

Table 6 : Variables number impact for the Electrical consumption prediction for the restaurant

Variables number	MAE	MSE	RMSE	MSD	R^2
4160	1.4	4.4	2.1	0.1	0.94
300	2.0	8.6	2.9	0.1	0.89
150	1.9	7.5	2.7	-0.1	0.90
50	2.0	8.2	2.8	-0.3	0.89
15	**2.22**	**10.**	**3.2**	**-0.1**	**0.86**

5.3.3 Frequency optimisation

We also test the difference in our results between the data at the second aggregate per minute and one data per hour to predict the next hour. The average error of the deviations increases by 0.5 kWh (Table 7). The daily average electricity consumption curve follows the same pattern. In the tertiary sector, the companies follow a precise production schedule during the day, contrary in the residential sector [LUC-2017].

Table 7 : Electrical consumption prediction results with different period for the restaurant

Models	Period	MAE	MSE	RMSE	MSD	R^2
Random Forest	**1 min**	**1.488**	4.494	2.12	0.124	0.946
Random Forest	**1 hour**	**2.133**	11.36	3.371	-0.273	0.863

6 CONCLUSION

Several methods and algorithms have been tested and we can conclude that the Combinatorial Random Forest and Gradient Boosting Tree is particularly adapted to our problem with predictions around 98% +/- 0.6. These techniques with the boosting are very useful in the context of the energy prediction in particular for the consumption and the production peaks which represent only a few % in the training set.

From an information system point of view, the Combinatorial Random Forest and Gradient Boosting Tree analysis allows us to limit the number of data, create a simple model and guarantee the same level of precision, It is possible to interpret the prediction by analysis of the statistics in output of the random Forest algorithm.

The identification of the human presence is an essential parameter both for the piloting and and the prediction of energy consumption. To create a non-intrusive system, it is easier to understand this problem in the residential and tertiary sector by integration of a production planning. This allows us to target model training on what interests us here, predict overall power consumption during business hours. We think that we should first focus on the microgrids in the tertiary and industrial sector, where the human variable is more predictable by the definition of business occupation schedules.

7 REFERENCES

[ARB-2015] Thomas Berthou, Développement de modèles de bâtiment pour la prévision de charge de climatisation et l'élaboration de stratégies d'optimisation énergétique et d'effacement, Ecole Nationale Supérieure des Mines de Paris, 2013

[BER-2010] Guide to Intelligent Data Analysis,How to Intelligently Make Sense of Real Data, Berthold, M.R., Borgelt, C., Höppner, F., Klawonn, F, 2010

[CHE-2016] Extended modeling procedure based on the projected sample for forecasting short-term electricity consumption, Che-Jung Chang, Jan-Yan Lin, Meng-Jen Chang, 2016

[COC-2012] M. Cococcioni, E. DAndrea, B. Lazzerini, One day ahead forecastingof energy production in solar photovoltaic installations: An empiricalstudy, Intelligent Decision Technologies 6 , 2012

[JAZ-2002] S. JAZOULI et S. BARROIS : Calage d'un modèle thermique de bâtiment "r3c2" à partir de la courbe de charge et de la météo. Rapport technique, EDF, 2002

[HART-1992] G.W. Hart. Nonintrusive appliance load monitoring. Proceedings of the IEEE, vol. 80,no. 12, pages 1870 –1891, 1992

[LUC-2017] Luc Dufour. Contribution à la mise au point d'un pilotage énergétique décentralisé par prédiction. Génie civil. Ecole des Mines d'Albi-Carmaux, 2017.

[MEL-2006] A. Melli, M. Benghanem, S. A. Kalogirou. An adaptive wavelet-network model for forecasting daily total solar-radiation. Applied Energy, vol. 83, no. 7, p. 705-722, 2006.

[MOU-2016] Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review. Moulay Larbi Chalal , Medjdoub Benachir,Michael White,Raid Shrahily, 2016

[NOM-2011] F. Nomiyama, J. Asai, J Murata, A study on Global Solar Radiation Forecasting Using Weather Forecast Data, IEEE, 2011

[PAR-2011] Le perceptron multicouche et son algorithme de rétropropagation des erreurs Marc Parizeau Département de génie électrique et de génie informatique Université Laval, 10 septembre 2004

[REB-2015] Modeling and forecasting energy consumption for residential buildings in Algeria using bottom-up approach." Rébha Ghedamsi, Noureddine Settou, Abderrahmane Gouareh, Adem Khamouli, Nadia Saifi, Bakhta Recioui, Boubekker Dokkar, 2015

[VIC-2015] Victoria Moreno, Luc Dufour, Antonio F Gómez Skarmeta, Antonio J Jara, Dominique Genoud, Bruno Ladevie and Jean-Jacques Bezian, Big data: the key to energy efficiency in smart buildings, Soft computing manuscript, 2015

[ZAM-2013] "A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production." M. Zamo, O. Mestre, P. Arbogast, O. Pannekoucke, 2013

[ZHA-2016] Time series forecasting for building energy consumption usingweighted Support Vector Regression with differential evolution optimization technique." Fan Zhanga, Chirag Debb, Siew Eang Leeb, Junjing Yangb, Kwok Wei Shah, 2016

8 CONFERENCE TOPIC

Smart building