Short-term anomaly detection in gas consumption through ARIMA and Artificial Neural Network forecast

Marco De Nadai
University of Trento
Email: me@marcodena.it

Maarten van Someren
University of Amsterdam
Email: M.W.vanSomeren@uva.nl

***Abstract*—The focus of this paper is on the discovery of anomalies in gas consumption that can be reported to the building managers who can identify wasteful equipment and settings and repair these. Our approach is to treat this task similar to prediction, using historical data. Firstly, the short-term (hourly) gas consumption is predicted. Then the outliers are detected on the base of deviations from the predicted value. Anomalies are thus detected without previously labelled examples. We proved that a simple approach by design allows the detection of potential flaws in the energy consumption in a fast and straightforward way.**

I. Introduction

Approximately 41% of the total energy in Europe is consumed by buildings (households and services) [1]. Studies and states' directives about minimizing energy consumption and using renewable energy steadily increased with the reduction of fossil fuels, political conflicts and the increase of various environmental problems. Under those circumstances, the European Union has set the target to raise EU energy consumption produced from renewable resources to 20%, to reduce by 20% the EU greenhouse gas emissions and to improve by 20% the energy efficiency [2]. This means investments to re-qualify old buildings, new energy regulations and diagnosis, but also new efficiency systems for appliances.

Moreover, forecasting energy demands is one of the major research field in the energy departments because it helps gas utilities, but also industries and families. In fact, companies and families can leverage with it to reduce their energy consumption and increase efficiency. For this reason, big companies like Google have shown their interest in this new market, developing thermostats which automatically control the house climate basing their decisions on the users' schedule. Nest, a company acquired by Google, declared that thanks to its automatically learning thermostat, customers saved 11.3% of AC-related energy usage without compromising comfort [3]. Thus, learning to predict consumer behaviour can save a lot of money because it enables buying energy more cheaply. However, an additional opportunity is to use the predictions to detect system anomalies. Commercial buildings consume from 15% to 30% more energy than necessary due to poorly maintained, degraded, and improperly controlled installations [4]. Such anomalies can be automatically detected by a fault detection and diagnosis (FDD) system, which can then send an alarm to a human engineer to analyse the situation further.

In this project we focus on gas consumption, which shows an irregular behaviour which is not easily predictable with classic methods. The algorithm presented in the following sections is based on predictions made by a system which can model linear and non-linear behaviour of the data with very reliable results. Anomalies are identified on the basis of deviations from the predicted value.

A. What is an anomaly?

An anomaly (or outlier) by definition [5] *is an observation which deviates significantly from other observations so that it creates suspicion that it was created by different dynamics*. Despite this general definition, the more appropriate way of defining outliers is highly application-dependent, because even same scenarios may require different determinations of outliers.
In our project, anomalies are very closely related to the time-series forecasting problem, since anomalies are detected on the basis of deviations from the predicted value. Anomalies can have distinct main causes:

- Defective system (e.g. a defective heater in a room).
- Bad human behaviour (e.g. people who leave open the window in a room while the system is trying to heat it).
- Defective monitoring system, where the system monitors different values from the real one, due to a malfunction, computing process errors or a recording negligence.

B. Organization

This paper is organized as follows. section II reviews the major related work. In section III the solution proposed in this work is presented, with the data that are used. Since this work uses Artificial Neural Networks and Auto-regressive methods, in the same section these are briefly explained, along with the gas forecaster. Finally, in section III-C the anomaly detection procedure is explained. In section IV the system is evaluated. section V discusses the conclusions of this project.

II. Related work

Anomaly detection systems are used extensively in a wide variety of applications such as fraud detection, intrusion detection for cyber-security, performance analysis and fault detection. In this section we review anomaly detection systems for energy consumption.

Anomaly detection is often viewed as an unsupervised learning problem because it is unknown in which variable(s) an anomaly will occur. Unsupervised algorithms typically use a distance or probability measure to identify anomalous data-points [6]. One of the most popular approaches is based on the *Gaussian error* theory, which supposes Gaussian distributed data and identifies anomalies observing the points with a low probability to lie in this distribution. For example Ferdowsi et al. [7] identified anomalies with this method through a rolling median window of points.

Unsupervised methods are usually based on a clustering algorithm which tries to group similar points/trends. In this case, the least similar points to these groups are anomalies. For example Khan et al. [8] applied k-means and DBScan to detect anomalies in office lighting energy consumption. They correctly discovered some anomalies but they encountered some problems on detecting faults strongly related to time.

Anomaly detection can also be approached as a form of supervised learning when labeled data are available. Hence, an anomaly is a data-point for which the value is very different from the predicted value. Although the problem is usually related to classification methods, in this project we also want to predict energy consumption. To the best of our knowledge, no other projects applied this combined method. This is probably because it is simpler to split the problem in two systems. Forecasting is not a trivial task, especially if short-term predictions of irregular time-series are necessary. In fact, precise energy consumptions predictions with a hourly time horizon have to deal with irregularities and sudden changes of the values, making the task very hard. In literature we find many ANN models for heating and electric load. For example, Kalogirou et al. [9] used a *back-propagated* ANN to predict the required heating load of 225 buildings simply analysing the historical load pattern. Differently, some researchers tried to add meaningful features in order to better model the load. For example, Taylor et al. [10] used weather data (51 variables) to predict the 10 days ahead electric load, while Gonzales et al. [11] predicted hourly electric consumption with 7 weather variables, showing that good results can also be achieved with lower complexity.
Since the behaviour of weekends and business days is completely different, researchers such as Neto et al. [12] build two models for the two different typology of day and achieved an error lower than 10%. Similarly to what we try to apply in this project, Zhang et al. [13] applied a hybrid ARIMA and ANN model to three datasets from different field, proving that a hybrid methodology that has both linear and non-linear modelling capabilities, can be a good strategy for practical use.

Finally, Hippert et al. [14] extensively reviewed many ANN based electric forecasting systems and concluded that most of them seem misspecified. In fact, models had been incompletely tested (no standard benchmarks, no synthetic data, etc.), proving that the results of many models could be improved (or misleading).

Since literature is more focused on electric load forecasting, we will try to apply some of these methods to forecast the more irregular gas consumption.

III. Proposed solution

The data in our possession are not labeled as *normal* or *anomalous*, and labelling the dataset would be prohibitively expensive. For this reason, an unsupervised learning method should be applied. However, our interest is in one variable: energy consumption. Therefore, our approach is based on supervised learning: we predict the energy consumption at each moment and detect points in time when this is much higher (or lower) than predicted.

Not all algorithms are universally good, in fact each algorithm works especially well with one type of data. Linear regression and auto-regression are the most common applied models in energy consumption. These methods find a linear function that predicts energy consumption as a linear function of variables like energy consumption of the previous period and outside temperature. They are applied using energy consumption at previous time instants but also including weather data and the type of use in the building. However, the relation between these variables and energy consumption is often non-linear [14]. Consequently, these systems cannot adequately capture the relationship in all the situations and data. Similarly, non-linear models like ANNs might have problems modeling linear relations [15]. For this reason a hybrid approach is proposed, where the ANN is helped in linear forecasting by the popular method ARIMA (auto-regressive integrated moving averages), commonly known as the BoxJenkins approach.

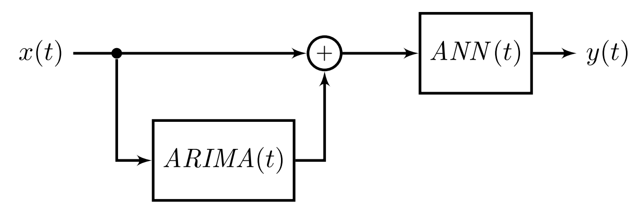


Fig. 1: The hybrid approach proposed in this project: the ANN is helped in linear forecasting by the popular method ARIMA (auto-regressive integrated moving averages).

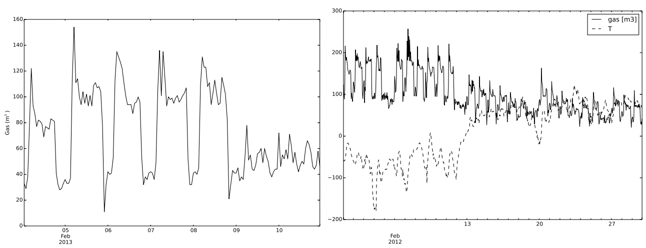
A. Data

The energy consumption datasets are about some buildings of the Hogeschool van Amsterdam. They are located in Amsterdam, a city with a maritime climate similar to England, strongly influenced by the North Sea. Winters are fairly cold and summers are rarely hot for the European standards. Amsterdam is characterized by the common presence of rain and wind and the weather conditions vary frequently.
This project used three buildings that were all heated by gas. In addition to the hourly aggregated gas consumption, data about the weather and about the use of the building were collected. The weather data that we used was collected by KNMI[1] at Schipol Airport, about 15 km from the tested buildings. The dataset consists of over 21 hourly collected variables from 2008 to 2013.

The gas consumption shows a strong seasonal component with daily and weekly cycles (see fig. 2a). The weekly pattern is composed by low consumption weekends and a higher one

[1] Koninklijk Nederlands Meteorologisch Instituut http://www.knmi.nl

during the business days, when schools are open. Unexpectedly, Monday shows lower consumptions than the rest of the business days.
The daily pattern is composed by a peak around 4:00-5:00 AM, when the system starts heating the building, and an approximately stable consumption until evening, when it drops to the minimum. This pattern is caused by the expected use of the building and by the outside temperature which has a clear daily/hourly relation with the gas consumption (see fig. 2b).



(a) (b)

Fig. 2: (a) Typical weekly and daily gas consumption behaviour in one building (740-NTH). (b) Typical monthly gas consumption behaviour in building 740-NTH and its relation with the temperature in building 740-NTH.

1) Data preparation: The data presented some irregularities like repeated and missing data-points caused by registration errors. The formers were deleted, while the latter were reconstructed by linear interpolation.

B. Method

Our solution uses a system composed by an Artificial Neural Network and an Auto-regression model. The ANN has a standard architecture with one hidden layer, a *Rectifier* function and a *back-propagation* Stochastic Gradient Descent SGD algorithm (see [16]). The number of nodes on the hidden layer and the learning rate were optimized using a split of the data.

Both models are applied to time-series which are a sequence of data-points typically measured at successive points of a uniform time interval t.

$$\{x(t_0), x(t_1), \dots x(t_i), x(t_{i+1}) \dots\} \tag{1}$$

where x is the value and t the time. Time-series forecasting is about predicting future values given past data (see eq. (2)).

$$\hat{x}(t+s) = f\left(x(t), x(t-1) \dots\right) \tag{2}$$

where s is the step size.

In the *Auto-Regressive Moving Average* (ARMA) model, the value of $x(t)$ is defined in terms of the last previous values in an interval of length p and q, also called moving average terms.

$$x(t) = \sum_{i=1}^{p} \varphi_i x(t-i) + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + c + \varepsilon_t.$$

The left-hand part is called *auto-regressive* part because it depends on the previous (lagged) values $x(t-1), x(t-2), \dots x(t-p)$, where φ_i is an input parameter. Similarly, the right-hand part is called *moving average* and it considers the error at time t as a linear combination of the previous errors $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$, where θ_i is an input parameter.
ARMA models are applied to *stationary* time-series, where the mean, variance and auto-correlation structure do not change over time. However many time-series have seasonal effects or trends defined as parameters of random walks, which characterise many types of series that are non-stationary. Taking the difference between consecutive data-points can often transform a non-stationary time-series into a *stationary* one. Consequently, ARIMA models differentiate series with deterministic trends if necessary, then apply an ARMA model. ARIMA models are usually mentioned as ARIMA (p, d, q), to show the ARMA parameters and the d order of difference. ARIMA models are also capable of modelling a wide range of seasonal data. $\text{ARIMA}(p, d, q)(P, D, Q)_m$, where m is the number of periods per season. The upper-case notation (P, D, Q) is used for the seasonal parts of the model, and the lower-case notation for the non-seasonal parts of the model.

The choice of the parameters p, d, q is highly application dependent and it relies on theory that is beyond the scope of this paper. More information can be found in [17]. At each time t, the ANN receives in input the ARIMA forecasted value of the gas consumption. This is done by a model ARIMA $(3, 0, 3)(2, 0, 1)_{24}$, fitted with the previous 21×24 historical values.

1) Feature engineering: Since the value to predict is time dependent, the most important feature is time. In fact, energy consumption depends on multiple factors such as the yearly season, the day of the week and the hour. Thus, days are represented by the weekday number in $[0, 6]$ (where 0 is Monday) and day of the year, in $[0, 366]$.
Holidays are special business days when schools are supposed to be closed, thus the gas consumption should be as low as weekends. For this reason we encoded all the holidays as Sundays.

There are many factors that affect the energy needs of buildings. They can be divided in three main groups namely *physical environmental*, *artificial designing parameters* and *human thermal discomfort*. The first one is composed of weather related parameters like outdoor temperature, wind speed, solar radiation, etc. The *artificial designing parameters* are related to the building: transparency, orientation, etc [18] (unfortunately not available in the datasets). The *human sensation of thermal discomfort* is correlated not only to the temperature, but also to other variables such as relative humidity, irradiation and wind speed. For this reason the temperature and the wind speed are included in the model. Contrarily, humidity and irradiation were found to be useless.
The system consumption is related to the outdoor temperature $T(t)$ at time t but also to the change in outdoor temperature from the previous hour $t-1$: $\Delta T(t) = T(t) - T(t-1)$, representing a positive/negative change of the external environmental conditions.

Gas usage $u(t)$ has a clear daily cycle but there also a weekly and annual cycle that the model may not be able to capture. It is defined as:

$$u(t) = s(t) + f(t) + r(t)$$

where $s(t)$ is the seasonality at time t, $f(t)$ is the trend and $r(t)$ is called *residual*. The time-series was analysed by the STL decomposition by LOESS [19], which decomposes a time-series into seasonal, trend and residuals components by an additive method. *residuals* are important for the model because they help it on understanding the gas behaviour without the seasonal component. Thus, gas consumption, wind speed and electricity *residuals* were added in the model as features.

Since the ANN needs to model the time-series time dependency, some moving windows help the model knowing the time-series "shape". The short/long-term dependency is represented by a moving window containing a "memory" of the previous states for some interesting variables. These memories formed a new set of states $\{\bar{x}_1(t), \bar{x}_2(t), \dots \bar{x}_n(t)\}$ from the original states $\{x(1), x(2) \dots x(n)\}$ where $\bar{x}_i(t) = x(t-i+1)$ (see table II).

Finally, the forecasted value from the ARIMA model was added.

All features (see table I and table II) were Z-score scaled to have a faster convergence (cf. [20]). The process to find the right parameters for the ANN was done iteratively, optimizing and testing them every time.

Variable	Data
Electricity load	$E(t)$
Hour	$\{0, \dots, 23\}$
Week day	W(t)=$\{0, \dots, 6\}$
Month	$\{1, \dots, 12\}$
Year day	$\{1, \dots, 366\}$
Next day	W(t+1)
Temperature	$T(t)$
Wind speed	$FH(t)$
ΔT_{k+1}	$T_{k+1} - T_k$
ARIMA forecast	$forecast(\text{ARIMA}(3,0,3)(2,0,1)_{24})$
STL year res.	$YearRes(t)$
STL day res.	$DayRes(t)$
Gas $(t-1)$	$G(t-1)$

TABLE I: ANN features.

Variable	Hours	Data
Gas peak$'$	5	$\max_{1 \le k \le 5} G(t-k)$
Gas sum$'$	5	$\sum_{i=1}^{5} G(t-i)$
Gas peak$''$	24	$\max_{1 \le k \le 24} G(t-k)$
Gas sum$''$	24	$\sum_{i=1}^{24} G(t-i)$
Electricity peak$''$	5	$\max_{1 \le k \le 5} E(t-k)$
Electricity sum$''$	5	$\sum_{i=1}^{5} E(t-i)$
Temp peak	5	$\max_{1 \le k \le 5} T(t-k)$
Temp sum	5	$\sum_{i=1}^{5} T(t-i)$
ARIMA peak$'$	5	$\max_{1 \le k \le 5} \text{ARIMA}(t-k)$
ARIMA sum$'$	5	$\sum_{i=1}^{5} \text{ARIMA}(t-i)$

TABLE II: ANN's moving windows.

C. Anomaly detection

We wanted to develop a simple and efficient anomaly detection algorithm based on the detection of deviations from the predicted value. For this reason, we simply compared at each step the Euclidean distance between the time-series and defined an anomaly comparing the distance with a threshold defined by the user: $d(i,p) >= \delta$.

$$d(p,q) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2} \quad (3)$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \quad (4)$$

This means that the developer has to defined an arbitrary parameter which is not easy to choose. Despite this, the method is so simple that allows to decrease (increase) the threshold δ iteratively, in order to detect anomalies of different magnitude.

IV. Experimental Evaluation

In this section, we evaluate the method on synthetic and real data. All the described results are obtained by a *k-fold cross-validation*, meaning that the network is trained $k = 5$ times, each time leaving out a subset of data from training in order to test the ANN. The average result over the k tests is saved. Data is randomly sampled, taking as validation data a random (contiguous) 15% of the original dataset, splitting the rest of the dataset in test (15%) and training data (70%). The test and training data are randomly shuffled before each cross-validation training phase.

Although the Mean Absolute Percentage Error (MAPE) is considered a standard for examining the quality of the models prediction of energy load, it represents an adequate error measure only when the loss function is linear. Recent studies demonstrated that this is not the case [9], [21]. Moreover the percentage error is infinite if there are zero values on the series or frequent intermittent data, and it puts a heavier penalty on positive errors than on negative errors [22]. Because of these disadvantages, this paper only considered the minimization the Mean Absolute Error (MAE).

$$\text{MAPE}^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{Y}_i - Y_i \right|$$

Since Linear regression is one of the most common forecast model in the energy field, all the forecast results are compared to it. This linear regressor was tested with the same previously described *Cross-Validation* parameters, and fitted on 2 variables: Electricity and Humidity (found the best for this problem).

A. Experiments with artificial data

First we did a naive experiment with synthetic data to see how the method works for simple clear examples and to test its correctness. Two anomalous days were generated by different algorithms. In the first one the real consumption was modified by a random value, simulating a system measurement/control malfunction which makes the consumption bouncing up and down (see section IV-A). The second synthetically created day was created adding $50m^3$ of gas consumption to the real one,

[2] MAPE errors are calculated only on the non-zero values, to avoid the problems described before.

Model	neurons	epochs	MAPE	MAE
Linear R.	-	-	149.94	15.27
ARIMA[3]	-	-	117.27	22.52
ANN	60	15	41.78	9.52
Hybrid	60	35	31.06	7.33

TABLE III: Best selected results in building 740-NTH, to compare the ARIMA, ANN and hybrid model.

creating a pattern which simulates a strange behaviour and/or a malfunction of the heating system (see eq. (5)). $G(t) = G(t) + v * 30$ where $v \sim \mathcal{N}(0, \sigma^2)$ and:

$$G'(t) = G'(t) + 50 \tag{5}$$

These two synthetically generated outliers were correctly identified.

B. Experiments with real data

The robustness of the design was proven with several buildings: Hva 740 -NTH, Hva 761 - KMH and Hva 882 - WBW. This system was able to identify anomalies with different energy consumption patterns without any further building specific configuration.

Some interesting behaviours were found through this work, and immediately reported to the facility management. Thus, it is possible to comment some unusual behaviours:

Holidays Regardless of whenever the school was operational or not, the first tests showed that the HVAC system was heating certain buildings like any other working day. For example, Tuesday 25th of December 2012 (Christmas) was heated like a normal business Tuesday even if the building was certainly closed. This causes an avoidable waste.

Consumption bounces In fig. 3a a strange zig-zag behaviour can be seen for building 740-NTH. It seems that the system is wasting energy and this shape is totally different from the usual one (fig. 2). This lasts for months and it is clear that also the ANN training is affected by this outlier-like behaviour.

Peaks Around the initial days of September there is a huge bounce of the consumption (up to three times more than the maximal consumption of the year). In building 761-KMH, irregular peaks during April 2013 were found every day, probably when the heating system turns on (see fig. 3c).

August with heaters In building 740-NTH, during August 2009 and August 2011 the heaters were active even without an apparently cold summer (similar to what happens in fig. 3f).

Outliers Some other outliers are found but they need to be confirmed by the managers, hopefully after the verification of the previously mentioned behaviours. For example fig. 3e shows a big bounce of consumption during late evening of May 4th, when the following day is a holiday, but also when the builing should be already closed. Similarly, fig. 3f) shows a zig-zag behaviour on April 9th and a completely different behaviour on the 10th of April.

It has to be noted that ANN was trained with non-cleaned data (data which can have anomalies), thus the model can be improved after the exclusion of these.

3 Calculated iteratively

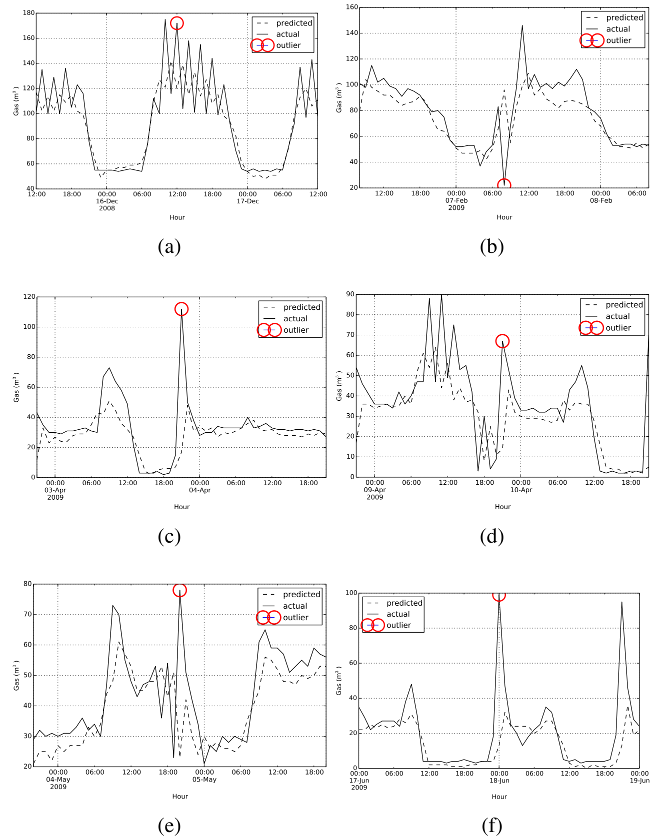


Fig. 3: (a) Some anomalies found in building 740-NTH. An anomaly is identified with a circle, even if the anomaly is collective (continues after the identified point).

V. Conclusion

Any model can't accurately treat all the situations for a large amount of historical load data. The irregular fluctuation of the gas consumption was hardly predictable, consequently the ANN model was helped with the well known ARIMA model. Although literature presented similar models to forecast electric consumption, the hybrid model presented here is almost unique because it focuses on short-term gas consumption forecast, which are very irregular and not easily predictable with classic methods. Since the prediction is very accurate (with MAE from $\sim 7\ m^3$ in building 740-NTH, to MAE $2.5\ m^3$ in building 761-KMH), the outlier mechanism is able to easily detect strange behaviours without the need to possess previous examples of outliers. Although this is a good achievement, we play to take advantage on the recent development in ANNs [23] in order to further refine the model and improve this prediction method.

Our model is of course limited by the absence of a ground truth of the managers and by the absence of a public available dataset to test and compare the model with other researchers. Moreover, the model is simple by design thus it could not detect some complex anomalies in the underlying data.
Despite this, we believe that the obtained results are compelling evidence that a well configured and simple prediction system can monitor a HVAC system and help public and

private administrations to save energy that would be otherwise wasted. One of the biggest advantage of this type of approach is that managers doesn't have to choose between detecting anomalies and predicting energy consumptions, having both the results at the same time. Even though this project was focused on forecast the highly irregular gas consumption time-series, it is believed that similar results could be also obtained with the more regular electric consumption time-series. The system architecture is voluntarily general and can be applied in various fields.

ACKNOWLEDGMENT

This project was supported by Universita' degli Studi di Trento. The authors wish to thank Jesse Eisses for discussions and help with acquiring the data.

REPLICABILITY

The sources of this project are available at `https://github.com/denadai2/energyUva`.

REFERENCES

[1] E. C. Eurostat, *Energy balance sheets - 2010-2011 - 2013 edition*. Luxembourg: Publications Office of the European Union, 2013. [Online]. Available: http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-EN-13-001

[2] European Parliament and Council of the European Union, "Directive 2009/28/EC," Brussels, 2009.

[3] Nest, "Energy savings from nest white paper preview," https://nest.com/downloads/press/documents/efficiency-simulation-white-paper.pdf, 2014.

[4] S. Katipamula and M. R. Brambley, "Review article: methods for fault detection, diagnostics, and prognostics for building systemsa review, part i," *HVAC&R Research*, vol. 11, no. 1, pp. 3–25, 2005.

[5] D. Hawkins, "Identification of outliers," *London: Chap*, 1980.

[6] C. C. Aggarwal, *Outlier analysis*. Springer, 2013.

[7] H. Ferdowsi, S. Jagannathan, and M. Zawodniok, "A neural network based outlier identification and removal scheme," in *Prognostics and Health Management (PHM), 2013 IEEE Conference on*. IEEE, 2013, pp. 1–6.

[8] I. Khan, A. Capozzoli, S. P. Corgnati, and T. Cerquitelli, "Fault detection analysis of building energy consumption using data mining techniques," *Energy Procedia*, vol. 42, pp. 557–566, 2013.

[9] S. A. Kalogirou, "Artificial neural networks in energy applications in buildings," *International Journal of Low-Carbon Technologies*, vol. 1, no. 3, pp. 201–216, 2006.

[10] J. W. Taylor and R. Buizza, "Neural network load forecasting with weather ensemble predictions," *Power Systems, IEEE Transactions on*, vol. 17, no. 3, pp. 626–632, 2002.

[11] P. A. González and J. M. Zamarreño, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," *Energy and Buildings*, vol. 37, no. 6, pp. 595–601, 2005.

[12] A. H. Neto and F. A. S. Fiorelli, "Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption," *Energy and Buildings*, vol. 40, no. 12, pp. 2169–2176, 2008.

[13] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[14] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *Power Systems, IEEE Transactions on*, vol. 16, no. 1, pp. 44–55, 2001.

[15] G. Zhang, B. Eddy Patuwo, and M. Y Hu, "Forecasting with artificial neural networks:: The state of the art," *International journal of forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[16] C. M. Bishop *et al.*, *Pattern recognition and machine learning*. springer New York, 2006, vol. 1.

[17] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John Wiley & Sons, 2005, vol. 589.

[18] B. B. Ekici and U. T. Aksoy, "Prediction of building energy consumption by using artificial neural networks," *Advances in Engineering Software*, vol. 40, no. 5, pp. 356–362, 2009.

[19] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[20] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.

[21] S. Kajl, M. Roberge, L. Lamarche, and P. Malinowski, "Evaluation of building energy consumption based on fuzzy logic and neural networks applications," in *Proc of CLIMA*, 2000, p. 264.

[22] R. J. Hyndman, "Another look at forecast-accuracy metrics for intermittent demand," *Foresight: The International Journal of Applied Forecasting*, vol. 4, no. 4, pp. 43–46, 2006.

[23] I. Mizera and C. H. Müller, "Location–scale depth," *Journal of the American Statistical Association*, vol. 99, no. 468, pp. 949–966, 2004.