

Detecting short-term anomalies in gas consumption using ARIMA and Robust Artificial Neural Networks

Marco De Nadai
Universita' degli studi di Trento
Master's student
me@marcodena.it

Maarten van Someren
Universiteit van Amsterdam
Science Park 107
Amsterdam, The Netherlands
m.w.vanSomeren@uva.nl

ABSTRACT

The focus of this paper is on predicting and finding anomalies on gas consumption that can successively be reported to the building managers who can identify wasteful apparatus and fix them. Even if predicting future consumption is very useful, the existing approaches separate it from the outlier detection task, typically identifying them with the Gaussian error theory. Moreover, previous examples of outliers (or cleaned data) are usually required.

In this work a 2-phase system is presented, where the first one forecasts the short-term (hourly) gas consumption, where the second phase detects outliers on the base of deviations from expected (forecast) value. Outliers are detected without possessing previously labelled examples, thanks to a very precise forecast with a RMSE error between 8 m^3 and 2.5 m^3 .

Keywords

Energy forecasting, Time series, Robust Artificial neural networks, ARIMA models, anomaly detection, outlier detection, gas consumption prediction, energy forecast

1. INTRODUCTION

Energy consumption in buildings is one of the fastest growing sectors. Approximately 41% of the total energy in Europe is consumed by buildings (households and services) [15]. Studies and states' directives about minimizing energy consumption and using renewable energy steadily increased with the reduction of fossil fuels, the border frictions with eastern countries like Russia, and the increase of various environmental problems. Under those circumstances, the European union, with a recent directive [14], 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 EU's energy efficiency. This means investments to re-qualify old buildings, new energy regulations and diagnosis, but also new efficiency systems for appliances.

Forecasting energy demands has become one of the major research field in the energy departments because it helps gas utilities, but also industries and families. Gas utilities buy gas from pipeline companies on a daily bases, consequently they need to know the needs in advance to be competitive. Similarly, Companies and families have the aim of reducing the energy consumption and increase efficiency. Lately, big companies like Google have shown their interest in this new market, developing thermostats which auto-

matically control the house climate basing the decisions on 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 [38]. Learning people behaviour can save a lot of money but system anomalies still raise and spoil the saving effort. It is showed [30, 48] that commercial buildings consume from 15% to 30% more energy than necessary due to poorly maintained, degraded, and improperly controlled equipment. These anomalies can become easy fixable problems with a reliable fault detection and diagnosis (FDD) system.

Gas consumption is very irregular and not easily predictable with classic methods. The outlier detection system presented is based on predictions made by a hybrid ARIMA-ANN which can model linear and non-linear behaviour of the data with very reliable results. Outliers are identified on the basis of deviations from forecast value. Thanks to this system days/hours with abnormally consumption can be reported to the building manager, who can further analyse and fix the HVAC system minimizing energy waste caused by the outliers.

1.1 What is an outlier

An outlier (or anomaly) by definition [21] *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 this paper, outliers are very closely related to the time-series forecasting problem, since outliers are declared on the basis of deviations from expected (forecast) value. Outliers can have distinct main reasons:

- 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.

Another important aspect to consider is the nature of an anomaly. Anomalies can be classified in three categories [7]:

Point Anomalies If an individual data-point is considered anomalous with respect to the rest of the data.

Contextual Anomalies If a data-point is anomalous in a specific context, but not otherwise. For example in Holland 35 degrees of temperature might be normal during summer but not during the winter.

Collective Anomalies If a group of points is anomalous with respect to the entire data set. Collective Anomalies firstly appear at a point affecting the values immediately next to it. After a while, this effect dies out, leaving the time-series to its normal behaviour.

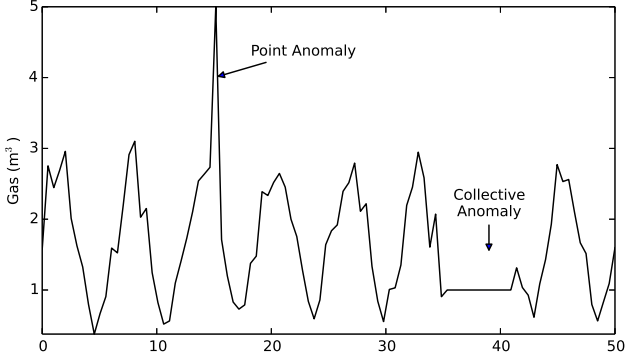


Figure 1: Different types of outliers

1.2 Organization

This paper is organized as follows. In section 2 the major related works are explained and divided in *supervised*, *semi-supervised* and *unsupervised* methods. In section 3 the solution proposed in this work is presented, describing also the possessed experimental data. 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 3.4 the outlier detection procedure is explained. In section 4 the system is completely evaluated. section 5 presents some future ideas for the readers, based on techniques that the author didn't have the time to apply. Finally, in section 6 the conclusions are presented.

2. RELATED WORK

Outlier detection system finds extensive use in a wide variety of applications such as fraud detection for credit cards, intrusion detection for cyber-security, performance analysis and fault detection in firms. Outlier detection can be *supervised*, *semi-supervised*, *unsupervised* and since the anomaly definition is highly application dependent, in this section only the energy consumption field methods will be reviewed.

2.1 Supervised anomaly detection

Techniques trained in supervised mode assume the presence of labelled training data, indicating previously known examples of anomalies. There are many algorithms in this category, but all of them rely on the classification of the points in outliers or normal data. To cite some of them, in [24] ANNs and Nearest neighbour classifiers are compared, while a Bayes classifier is used for fault detection. In [23] over 20 papers of ANN classifiers were compared.

2.2 Semi-supervised anomaly detection

Semi-supervised models are characterized by having a training set where there are only examples for normal data. There are no so many examples of this category in the energy field. In [17] this model is applied for spacecrafts fault detection.

2.3 Unsupervised anomaly detection

Unsupervised methods have no training data nor anomalies labelling. Typically most of the unsupervised outlier mechanisms use a measure of *outlierness* of a data point, such as sparsity of underlying region, nearest neighbour distance or the fit to underlying distribution [1]. In these cases a data-point is unusual due to one or more variables rather than a specific one (like in the supervised methods). Literature (in the energy field) is usually based on the *Gaussian error* theory. This suppose that the measurement follows the normal distribution, and outliers are the points falling outside three times the variance. One example above all is [16], who uses and improves this method to identify outliers with a rolling median window. Unsupervised methods are usually based clustering, where an algorithm tries to find similarities between points/trends and cluster them into groups, calculating the distance between them. A cluster is considered good when the intra-cluster distance is minimized and the intra-cluster distance is maximized. Popular methods in this group are *k*-means, one-class SVM and self-organizing maps. For example in [31] some clustering methods, like CART, *k*-means and DBScan, were applied to detect outliers in the office lighting energy consumption. The author showed different techniques applied along with the Generalized Extreme Studentized Deviate (GESD) and listed some irregularities found. He also stated that clustering methods were not able to detect faults strongly related to time variable.

3. PROPOSED SOLUTION

The available dataset does not indicate if any instance is *normal* or *anomalous*, and labelling the dataset would be prohibitively expensive. For this reason an unsupervised method (section 2.3) needs to be applied. Clustering methods are very difficult to be applied in time-series (eq. (1)) data with decent results. The solution would be to apply a Gaussian error method, instead this project wants to be more exhaustive, and develop a new method which also provide a energy use forecast, useful for future building plans. In this work outliers are declared on the basis of deviations from expected (forecast) value.

A time-series is a sequence of data-points typically measured at successive points of a uniform time interval t .

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

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

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

where s is the step size. A *multivariate* time-series is a $(n \times 1)$ vector of n time-series variables.

It can be seen that in academic and industry research, linear regression-based systems represent the standard “de facto” of energy forecasting. These techniques are usually applied to electric consumption, which is much more easily predictable than the gas one. Moreover, they have been applied in short-term forecast, since Auto-regressive methods are potentially useless in long ahead data-points. In recent

Building name	Date interval	Number of rows
HvA 740 - NTH	01/2008 - 03/2014	54.725
Hva 761 - KMH	01/2009 - 09/2013	40.407
Hva 882 - WBW	01/2008 - 03/2014	54.647

Table 1: Buildings used.

works auto/linear regressive methods were also combined with weather forecast data with energy use, but this relationship is clearly non-linear [23]. Consequently, even if some papers have acceptable results with measured datasets, these systems cannot adequately capture the relationship in all the situations and data.

Since ANNs are the state of the art technique of many machine learning problems where there are complex non-linear hypothesis, the proposed solution is based on them. Reviews of ANNs based forecasting system have concluded that they are promising but “a significant portion of the ANN research in forecasting and prediction, lacks validity”. Much work (as this one) needs to be done before they are accepted as forecasters.

Zhang et al [49] stated that ANN models really have advantages in dealing with a large amount of historical load data with non-linear characteristic, but they neglected the linear relations including the data. 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 Box-Jenkins approach.

3.1 Experimental data

The energy consumption datasets used were collected by Ebatech in various buildings of the Hogeschool van Amsterdam and provided by Universiteit van Amsterdam. The buildings are located in Amsterdam, the capital city of The Netherlands. This city has 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.

Ebatech collected different features in each building, with different granularity. This project used three buildings where the features overlapped (see section 3.1). All these buildings were heated only by gas, which was used only for this purpose.

The weather data was collected by KNMI¹ in Schipol, the Amsterdam airport 16 km far from the tested buildings. The dataset, findable in their website, consists in over 21 variables hourly collected. The proposed solution only uses few of them, as explained in the section 3.3, and the measured weather conditions at time t are used at time $t - 1$ as simulated forecast conditions. It is necessary to note that this will introduce an error in the final model, due to the effect of the weather forecast uncertainty [12, 42]. Another error is caused by using the conditions collected in a different location from the buildings positions.

3.1.1 Data analysis

¹Koninklijk Nederlands Meteorologisch Instituut <http://www.knmi.nl>

The energy consumption dataset covers a very large period ranging more than 5 years, allowing to see similar patterns even with different yearly/seasonal behaviour (one year could be different from another one for external factors like weather or building use). The dataset is very rich (with more than 50 variables) but also sparse because the monitored variables are not the same in all the buildings.

The gas consumption data is highly seasonal: daily and weekly cycles are quite perceptible, as it can be seen from fig. 3 and fig. 2. From the latter the weekly behaviour is clear: the last two days of the week (Saturday and Sunday) are completely different from the others. Moreover Monday seems a bit different from the rest of the days. The system reveals us that after a couple of hours, it decreases the consumption again. Furthermore, around 4:00-5:00 AM the system seems to react turning on the heating system, where in the previous hours of the night it seems only keeping a minimum temperature.

fig. 3 shows that the temperature has a clear daily/hourly relation with the gas consumption, while in fig. 4 the electric consumption is shown to be very smoothed and more regular than the gas one.

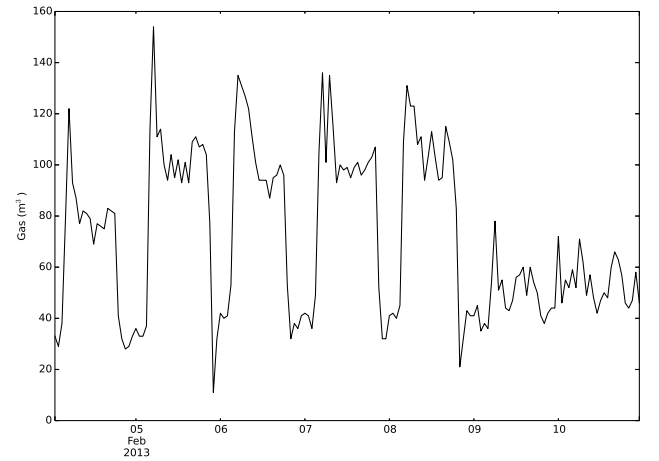


Figure 2: Typical weekly and daily gas consumption behaviour in building 740-NTH. The weekly pattern can be noticed by observing that the last two days of the week (Saturday and Sunday) have a completely different shape than the others. During the week, the daily behaviour is very similar, with one peak around 4:00-5:00 AM.

3.1.2 Data preparation

The data presented some irregularities like repeated and missing data-points. Repeated data-points were deleted, while missing data-points were reconstructed by linear interpolation because they could cause problems to ARIMA. Cubic and spline interpolation were also considered but the performance was heavily affected by these methods.

3.2 Background

3.2.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) were originally developed to mimic the brain functionality. The complexity of a brain neuron is abstracted by artificial neurons, also called

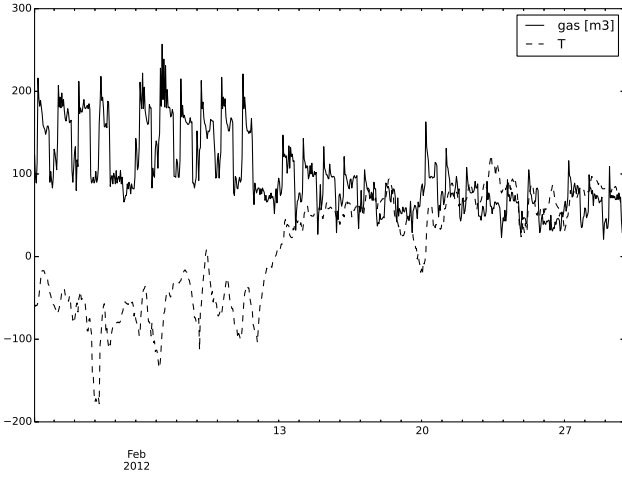


Figure 3: Typical monthly gas consumption behaviour in building 740-NTH and its relation with the temperature in building 740-NTH.

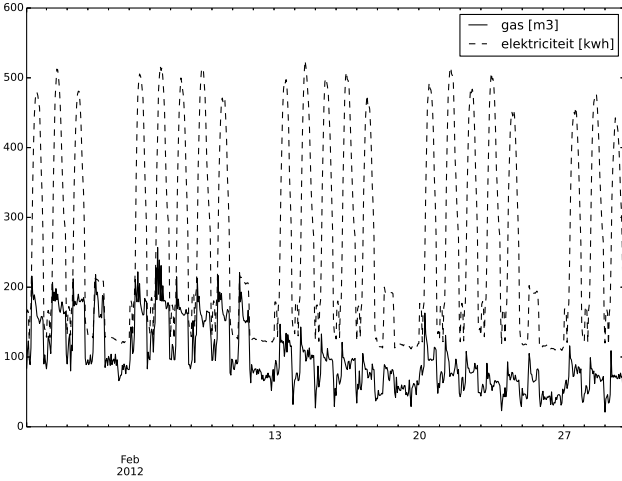


Figure 4: Relation between electric and gas consumption in building 740-NTH

units. These consist of *weighted inputs* computed by a mathematical function which determines the *activation* of a neuron. The *output* is the result of this function. Neuron activations are usually represented with linear, hyperbolic tangent or sigmoid function. Recent findings argued that Rectifier units seems to be more biologically plausible [18].

The most famous architecture is the one shown in fig. 5. Neurons are usually organized in layers, where the layer between input and output is called *hidden*. Neurons are usually fully connected with the neurons of the next layer. Learning is done by the Stochastic Gradient Descent (SGD) algorithm, which tries to find the best weight configuration to optimize the output (reducing a cost/error function). It is usually combined with the *Backpropagation* algorithm, which back-propagates the output error to the inner levels.

More information about ANNs can be found in [3].

3.2.2 Auto-regressive models

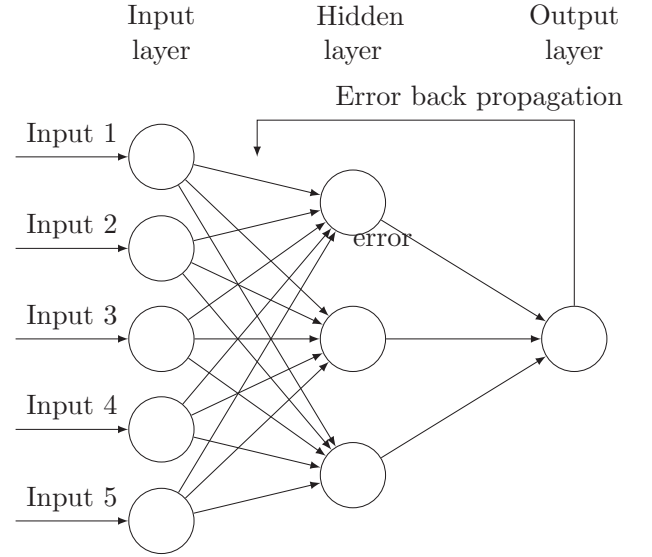


Figure 5: ANN with *Backpropagation*, 5 inputs, 3 hidden neuron in the hidden layer, and one output (in this paper, the gas consumption value).

Let X_1, X_2, \dots, X_t be the values in an univariate time-series. In the *Auto-Regressive Moving Average* (ARMA) model, the value of X_t is defined in terms of the values in the last windows of length p and q 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 $X_{t-1}, X_{t-2}, \dots, X_{t-p}$, the right-hand part is called *moving average* because the error at time t is the linear combination of the previous errors $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$.

These methods are applied to *stationary* time-series, so called when the mean, variance and autocorrelation structure do not change over time. However many time-series have seasonal effects or trends. In particular, random walks, which characterise many types of series, are non-stationary. Differencing the data-points can often transform a non-stationary time-series into a *stationary* one. Based on the Box-Jenkins models of the 1970s, ARIMA models differentiate series with deterministic trends if necessary, then applies 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 [43].

3.3 Artificial Neural Network forecast

Since section 2 is focused on outliers detection works, the next subsection will try to review energy consumption forecast. After this, the features and the ANN architecture are described.

3.3.1 Related work

Gas forecasting is divided in three macro categories and each of them have their own techniques: short-term, medium-term and long-term forecasting. The first one has to deal with irregularities and sudden changes of the values because it usually refers to prediction with a horizon of hours or days, the second refers to weeks. Long-term forecasting usually deals with data that rarely presents significant distortions and irregularities, so they have a small effect on the overall value. It usually refers to monthly or annual horizon. There are essentially four types of prediction models, extensively reviewed in [51] and [23]:

Engineering methods use physical principles to calculate thermal dynamics and energy behaviour of the building.

Statistical methods build empirical models to apply a regression to time series of values.

Machine learning methods based on ANNs and Support Vector Machines (SVMs) that try to predict energy usage with a learning algorithm.

Grey models apply a mixture of the models.

[29] used a *back-propagated* ANN to predict the required heating load of 225 buildings, Ekici and Aksoy used the same model to predict building heating loads in three-buildings. Nizami and Al-Garni [40] tried a simple *feed-forward* ANN to relate the electric energy consumption to weather data and population, Taylor and Buizza [47] used weather data (51 variables) to predict the 10 days ahead electric load. Gonzales [19] predicted hourly electric consumption. Some researchers tried to specialize the ANNs: Neto and Fiorelli [39] compared generic ANNs with working days ANNs and week-end ANNs, Lazzerini and Rosario [9] specialized them to predict electric lighting with weather data.

Other researchers tried also to apply a hybrid model to increase the ANNs performance. One example above all is [50] which applied a hybrid ARIMA and an ANN model to forecast electricity use, another one is [32] who improved the previous one. This paper is based also on his work.

As can be seen the focus of the majority of research is on electric forecasting. There are only few of them are about natural gas forecasting: Brown et al.[5] built one of the first predictor for natural gas consumption and Khotanzad et al [33] developed a two stage system ANN with very good results.

Finally [23] reviewed many ANN based electric forecasting systems and concluded that most of the papers seems misspecified models that had been incompletely tested (no standard benchmarks, no synthetic data, etc.). This work will try to give the readers the most exhaustive and proved work in this field.

3.3.2 Feature engineering

Feature Engineering is the process of transforming the raw data that better represent the underlying problem, and result in a better model performance. This process was done almost iteratively, cumulatively introducing, analysing, comparing and removing features from the model.

Time-series are characterized by more or less complex dependencies:

Known dependencies like date-time dependencies.

Hidden dependencies like the behaviour of the HVAC system (when it starts, when it raises the temperature etc).

Short/long-term dependencies between variables.

Data scientist and experts are focused in known dependencies while the proposed ANN is focused on the hidden ones. The short/long-term dependency was 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 $\bar{x}_i(t) = x(t - i + 1)$. The memorized variables are listed later in this section.

Since the value to predict is time dependent, the first feature was time. Energy consumption depends on the hour of the day but also on the day of the week and the seasonality of the year (month and day of the year) (as explained in section 3.1.1, fig. 3 and fig. 2). The weekday was represented as a number from 0 to 6, where 0 is Monday and 6 is Sunday. Since the behaviour of the holidays were considered similar to the weekends (particularly similar to Sundays), a function encoded all the holidays as weekend days. The day of the year, a number between 1 and 366, was added as well. All these date-time features are encoded by means of their sine and cosine values as usual reported in literature [41, 11, 19]. This transformed the time component in a cyclic feature that spans a fixed length (a single day for the hour) and it is bounded in $[-1, 1]$.

Another added feature was the current system load, which is the energy consumption at the k state when the load at $k + 1$ needs to be predicted.

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 construction: transparency ratio, orientation, etc [13], unfortunately these variables were not available in the dataset. 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. Even if they were available in the dataset, only the temperature and the wind speed were found to be significant.

The system consumption was believed to be related to the difference of the outdoor temperature between two instants (eq. (3)), representing a positive/negative change of the external environmental conditions.

$$\Delta T_{k+1} = T_{k+1} - T_k \quad (3)$$

where T_{k+1} is the predicted temperature for the period $k + 1$ and T_k is the value measured in the instant k . It needs to be noted that the real behaviour of the system was unknown, consequently it was not possible to know if this change would

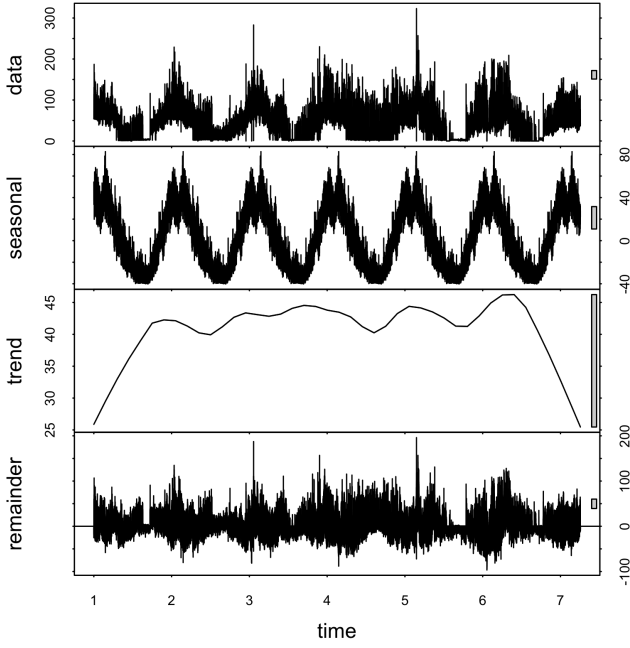


Figure 6: Yearly STL decomposition by LOESS in building 740-NTH.

have an immediate effect on the HVAC system and/or its reaction time.

Gas usage has a clear daily cycle but there also a weekly and annual cycle that the ANN may not be able to capture. Gas usage $u(t)$ 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 *remainder*. The time-series was analysed by the STL decomposition by LOESS [8](fig. 6), which decomposes a time-series into seasonal, trend and irregular components by an additive method. Since the ANN is interested in the *remainder* the daily, weekly and yearly remainders were added as feature. For the same reason the temperature, the wind speed and the electric consumption was decomposed by the STL method, resulting in stationary time-series added as feature.

Moreover, the time-series was processed iteratively with a 21 days long moving window to fit an ARIMA model. After this, the fitted model was able to forecast the next 24 values (the gas consumption of the following day). The seasonal ARIMA fitting was finding best ARIMA $(p, d, q)(P, D, Q)_m$ parameters by comparing the *Akaike information criterion* (AIC) of the tested models. Just for the sake of the reader curiosity, the most fitted model was ARIMA $(3, 0, 3)(2, 0, 1)_{24}$.

As previously stated, the ANN needed to model the time-series time dependency, consequently some moving windows could help the model knowing the shape of the time-series. This somehow simulated the Recurrent Neural Networks (RNNs) behaviour. Two moving windows were created for the gas consumption, and other two were created for the STL yearly residuals, memorizing sum and peak values of the previous 5 hours.

All the features that were not between the limits $[-1, 1]$

Variable	Data
Electricity load	$E(t)$
Hour	$\sin(2\pi(h)/24); \cos(2\pi(h)/24)$
Week day	$\sin(2\pi(wDay)/6); \cos(2\pi(wDay)/6)$
Month	$\sin(2\pi(mon)/12); \cos(2\pi(mon)/12)$
Year day	$\sin(2\pi(d)/366); \cos(2\pi(d)/366)$
Temperature	$T(t)$
Gas peak'	$\max_{1 \leq k \leq 5} G(t - k)$
Gas sum'	$\sum_{i=1}^5 G(t - i)$
Gas mean'	$\frac{1}{288} \sum_{i=1}^{288} G(t - i)$
Gas peak''	$\max_{1 \leq k \leq 24} G(t - k)$
Gas sum''	$\sum_{i=1}^{24} G(t - i)$
Electricity peak''	$\max_{1 \leq k \leq 5} E(t - k)$
Electricity sum''	$\sum_{i=1}^5 E(t - i)$
Temp peak	$\max_{1 \leq k \leq 5} T(t - k)$
Temp sum	$\sum_{i=1}^5 T(t - i)$
Wind speed	$FH(t)$
ΔT_{k+1}	$T_{k+1} - T_k$
ARIMA forecast	$forecast(ARIMA(3, 0, 3)(2, 0, 1)_{24})$
STL year res.	$YearRes(t)$
STL day res.	$DayRes(t)$
STL E res.	$Res(E(t))$
ARIMA peak'	$\max_{1 \leq k \leq 5} ARIMA(t - k)$
ARIMA sum'	$\sum_{i=1}^5 ARIMA(t - i)$

Table 2: ANN features.

were scaled (eq. (4)) to have a faster convergence [35] of the *Stochastic Gradient Descent*.

$$x'_i = \frac{x_i - \frac{\max(x) + \min(x)}{2}}{\frac{\max(x) - \min(x)}{2}} \quad (4)$$

where x_i is the original value and x'_i is the scaled one. Many practical tricks like element shuffling, the normalization and initialization were taken from [35, 4].

3.3.3 Architecture

ANNs were trained trying to minimize a cost function of the form

$$E = \frac{1}{N} \sum_{i=1}^n p(r_i)^2$$

where p the cost function is symmetric and continuous, $r_i = Y_i - \hat{Y}_i$ is the residual between the actual value and the forecast one, and N is the number of training patterns.

The most used cost function is based on the Mean Squared Error (MSE), commonly known in data modelling as Least Mean Squares (LMS) method. The basic idea of LMS is to optimize the fit of a model with respect to the training data by minimizing the square of residuals

$$p(r) = \frac{1}{2} r^2$$

but it is greatly influenced by outliers [36]. For this reason, the ANN was instead focused to minimize the Mean Log Squared Error (MLSE). This is based on the Least Mean Log Squares (LMLS) method (eq. (5)), presented by [36].

$$p(r) = \log(1 + \frac{1}{2} r^2) \quad (5)$$

The applied ANN was a 1-hidden-layer *Multilayer Feed-forward* ANN with a feedback structure, called *Backpropa-*

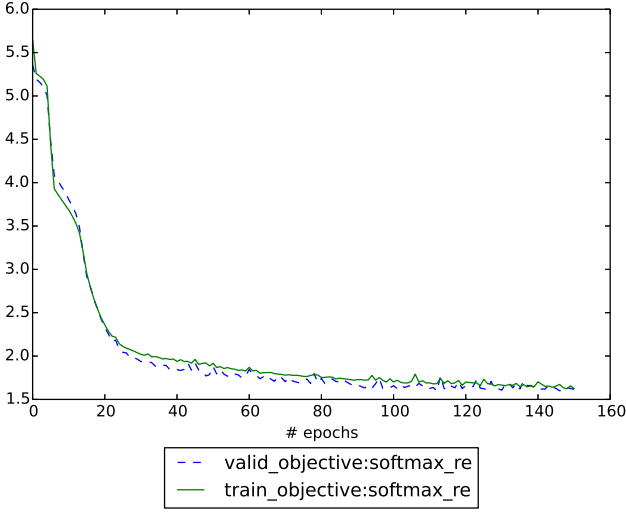


Figure 7: Training curve of the hybrid model with 80 hidden neurons

gation. Choosing the number of hidden units for the ANN is always a tricky task because it might lead to *overfitting*. As stated by [34, 44], using *early stopping* in a oversized *Backpropagation* ANN, where the number of hidden neurons is higher than the number of the features, makes easier to find global optimum and it avoids bad local optima. For this reason the number of hidden units was chosen to be greater than $2 \times |\text{features}|$ and then test-driven and the training was early stopped to prevent *overfitting*. The ANN was composed by *Rectifier* units and one output linear node. Training was done by the *Stochastic Gradient Descent* algorithm with 10 batch size and was characterized by a learning rate of 0.003 and fixed by a Momentum of 0.05, which could help to increase the speed, avoiding local minima.

This project used Python and *pandas* for the data analysis, *Pylearn2* [20] to construct and test the ANN. The *Forecast* [26] R package was used for the ARIMA forecast.

3.4 Outlier detection

According to Chebschev's theorem [2], almost all the observations in a data set of system states falls into the interval $[\mu - 3\sigma, \mu + 3\sigma]$, where μ and σ are respectively the mean and standard deviation of the data set, and the data points outside this interval are declared outliers. In this paper the ANN is used to predict the gas consumption, for this reason a point will be considered outlier if it will fall outside the 95% confidence interval² expressed for the RMSE. If it is assumed that the difference between the actual values x_i and the predicted value \hat{x}_i have:

$$\hat{x}_i - x_i \sim \mathcal{N}(0, \sigma^2) \quad (6)$$

- mean zero.
- follow a Normal distribution (it is assumed that it holds for the large amount of data utilized).
- and all have the same standard deviation σ .

$$\hat{x}_i - x_i \sim \mathcal{N}(0, \sigma^2) \quad (7)$$

²For the following description user *fabee* of *CrossValidated* needs to be mentioned: <http://tinyurl.com/19gvz65>.

it is possible to say that eq. (7) follows a χ_n^2 distribution with n degrees of freedom. Which means:

$$P\left(\chi_{\frac{\alpha}{2}, n}^2 \leq \frac{n\text{RMSE}^2}{\sigma^2} \leq \chi_{1-\frac{\alpha}{2}, n}^2\right) = 1 - \alpha \quad (8)$$

$$\Leftrightarrow P\left(\frac{n\text{RMSE}^2}{\chi_{1-\frac{\alpha}{2}, n}^2} \leq \sigma^2 \leq \frac{n\text{RMSE}^2}{\chi_{\frac{\alpha}{2}, n}^2}\right) = 1 - \alpha \quad (9)$$

$$\Leftrightarrow P\left(\sqrt{\frac{n}{\chi_{1-\frac{\alpha}{2}, n}^2}}\text{RMSE} \leq \sigma \leq \sqrt{\frac{n}{\chi_{\frac{\alpha}{2}, n}^2}}\text{RMSE}\right) = 1 - \alpha. \quad (10)$$

Therefore

$$\left[\sqrt{\frac{n}{\chi_{1-\frac{\alpha}{2}, n}^2}}\text{RMSE}, \sqrt{\frac{n}{\chi_{\frac{\alpha}{2}, n}^2}}\text{RMSE}\right] \quad (11)$$

4. EXPERIMENTAL EVALUATION

The ANN has been trained with *early stopping*, with a fixed number of training epochs (phases). All the results showed are obtained from a *k-fold Cross-Validation* technique, where the network was trained k times, each time leaving out a subset of data from training in order to test the ANN. The results of the k tests, were divided by k . The network was always trained with 70% of the data, 15% was used for validation and the other 15% to test.

Although the Mean Absolute Percentage Error (MAPE) is considered a standard for examining the quality of the models prediction of energy load, it represent an adequate error measure only when the loss function is linear. Recent studies demonstrated it is not [29][28]. 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 [25]. Because of these disadvantages, this paper only considered the minimization of the Root Mean Square Error (RMSE), which penalize large errors. For every experiment the Mean Absolute Error (MAE) was calculated as well.

$$\text{MAPE}^3 = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100 \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2} \quad (13)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i| \quad (14)$$

The system was tested with synthetic and measured experiments. The first one refers to author generated data, while the latter used the Ebatech dataset.

4.1 Synthetic experiments

Two days outliers were synthetically generated by different algorithms. In the first one the real consumption was

³MAPE errors will be calculated only on the non-zero values, to avoid the problems described before.

modified by a random value, simulating the system measurement/control malfunction which makes the consumption bouncing up and down (see eq. (15)). The second synthetically created day was created adding $50m^3$ of gas consumption to the real one, creating a pattern which simulates a strange behaviour and/or a malfunction of the heating system (see eq. (16)).

$$G(t) = G(t) + v * 30 \quad (15)$$

where

$$v \sim \mathcal{N}(0, \sigma^2)$$

$$G'(t) = G'(t) + 50 \quad (16)$$

The two outliers were correctly detected, as it can be seen in fig. 8.

4.2 Measured data experiments

The first test was done by placing a week-end gas consumption in a business day, in order to simulate an anomaly with real data. In fig. 9 it can be seen that the outlier mechanism works perfectly when the Sunday gas consumption is placed in a weekday.

The outlier was correctly detected, as it can be seen in fig. 9. The robustness of the design was proved with different building, listed in section 3.1.

Apart this little experiment, some interesting behaviours were found though this work:

Holidays Regardless of whenever the school was operational or not, the first tests showed that the Ebatech system was heating the 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. 10 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). Is it a test? In building 761-KMH, irregular peaks during April 2013 were found every day, probably when the heating system turns on.

August with heaters In building 740-NTH, during August 2009 and August 2011 the heaters were active even without an apparently cold summer.

Outliers Some other outliers are found but they need to be confirmed by the managers, hopefully after the verification of the previously mentioned behaviours.

All these problems will be immediately reported to the company.

An ANN with the standard cost function MSE was also trained, apparently resulting in a smaller RMSE error (table 3). However, the Hybrid MLSE model was more precise,

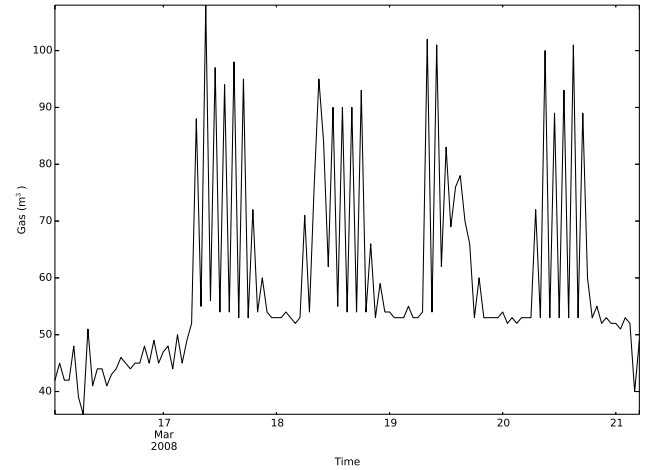


Figure 10: Strange zig-zag behaviour found by the algorithm.

Model	neurons	epochs	RMSE	MAPE	MAE
ARIMA ⁴	-	-	88.50	117.27	22.52
ANN	80	15	11.95	34.78	8.52
HyMSE	80	70	9.4	27.66	6.90
HyMLSE	150	140	10.02	30.05	7.26

Table 3: Best selected results in building 740-NTH, to compare the ARIMA, ANN and hybrid model. HyMSE is the Hybrid model with MSE cost function, while HyMLSE is the same model with MLSE cost function.

detecting better the possible outliers. They contributed the most to the error.

In table 4 the results in the different buildings can be read.

5. FUTURE WORK

In this project a holidays list was used but the outlier detection can be improved asking a time-table for the buildings, indicating when these are closed.

ARIMA models can't detect more than one seasonality but it can be helped with Fourier terms and ARIMA *dummy variables* to produce reasonable forecasts. When a multi-seasonality is present, a non-parametric model like TBATS[10] could overpass ARIMA and help further the ANN. This recent algorithm is extremely slow, but it is probable that the performance will be improved soon.

ANNs are sensitive to missing values and irregularities, but it was not possible to contact the buildings managers in order to possibly confirm/identify previously known outliers. For this reason the ANN training was done with not entirely clean data, and this probably affected the performance. It

⁴Calculated iteratively as described in section 3.3

Model	neurons	epochs	RMSE	MAPE	MAE
740-NTH	150	140	10.02	30.05	7.26
761-KMH	150	140	2.49	18.30	1.00

Table 4: Best selected results for all the buildings.

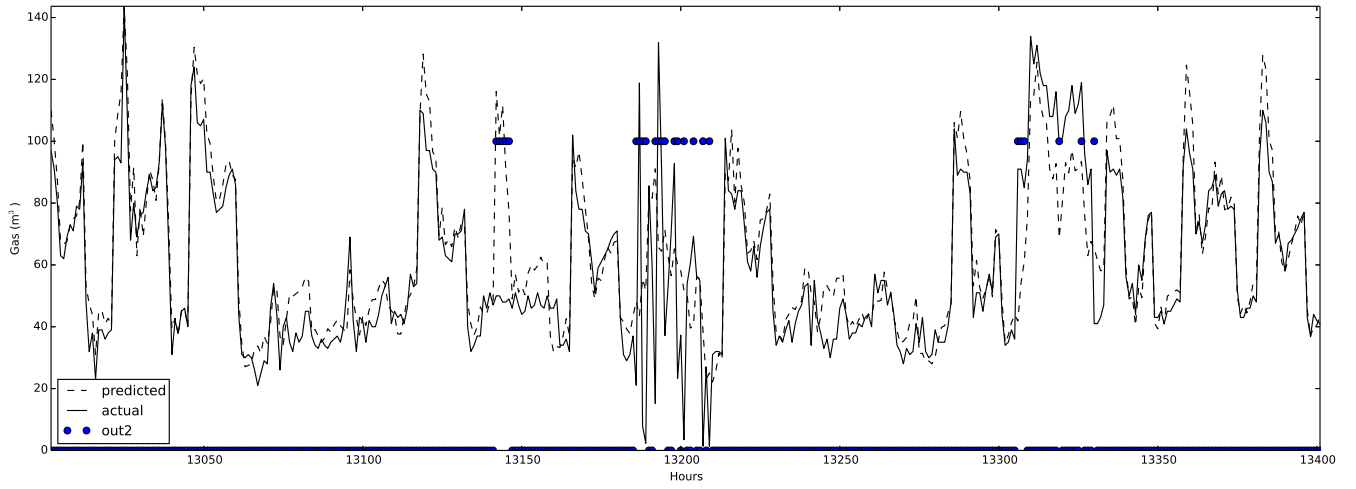


Figure 8: Outlier detection with synthetically generated data. The circle represents the hours where an outlier is detected.

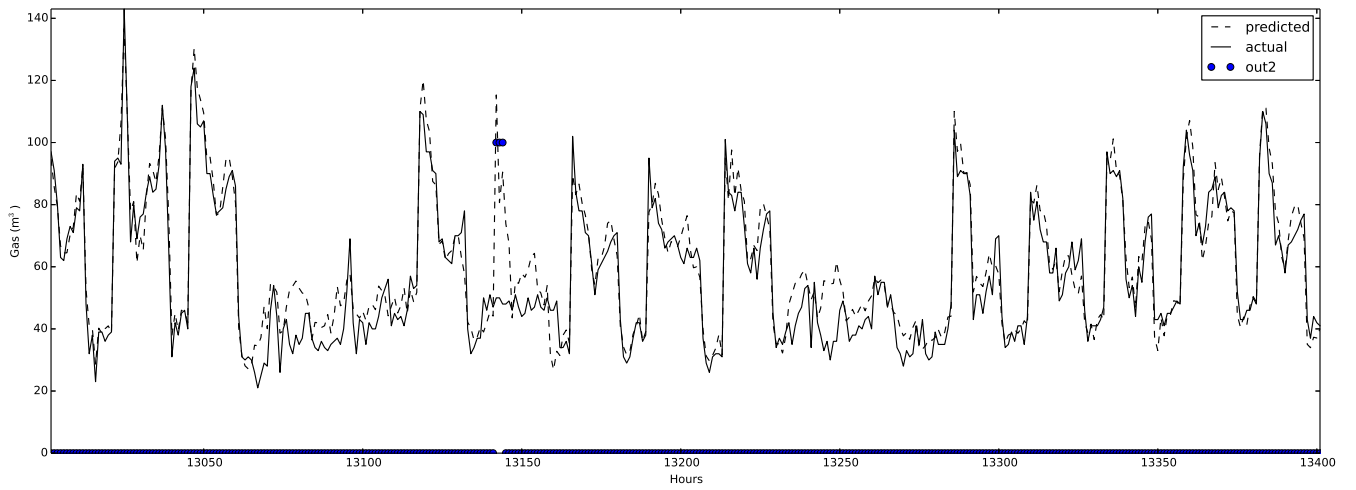


Figure 9: Outlier detection where the gas consumption of a weekday was replaced by a Sunday one. The (three) circles represent the outliers detected by the system.

is necessary to contact these building managers to further help the training of this algorithm.

The features were scaled standardizing them to the range $[-1, 1]$. It is also possible to normalize them to have mean 0 and standard deviation 1. In this case Robust estimates of location and scale are desirable if the inputs contain outliers. [27] and the recent [37] can be the base of a future refine of the ANN inputs.

In a future work it is also possible to analyse the difference between the MSE and the MLSE cost, not only in the prediction, but also in the outlier detection.

Before 2006, ANNs were almost always associated with the *Backpropagation* algorithm and with the 1-hidden layer architecture. The problem with these architectures is that they get stuck in poor local-optima. In 2006 there was a huge breakthrough, mainly started by [22], which is called *Deep learning* and it represent the new fashion of ANNs, based on multi hidden layers and new algorithms. Future improvements can be based on Recurrent Neural Networks (RNNs) and Restricted Boltzmann Machines (RBMs) which were recently proved to be interesting in time-series forecasting [6, 46, 45].

6. 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 robust cost function and with the well known ARIMA model. Although other papers 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 RMSE from $8 m^3$ in building 740-NTH, to RMSE $2.5 m^3$ in building 761-KMH), the outlier mechanism is able to easily detect strange behaviours without the need to possess previous examples of outliers. The scope of this paper was forecasting the highly irregular gas consumption time-series, but it is believed that similar results could be also obtained with the more regular electric consumption time-series.

It is hoped that this work would lead to new analysis of the energy consumption in public buildings.

7. ACKNOWLEDGMENTS

This internship was economically supported by Università degli Studi di Trento. I also thank Jesse Eisses who helped me to find some interesting ideas for this paper.

8. REFERENCES

- [1] C. C. Aggarwal. *Outlier analysis*. Springer, 2013.
- [2] B. G. Amidan, T. A. Ferryman, and S. K. Cooley. Data outlier detection using the chebyshev theorem. In *Aerospace Conference, 2005 IEEE*, pages 3814–3819. IEEE, 2005.
- [3] C. M. Bishop et al. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.
- [4] L. Bottou. Stochastic gradient descent tricks. In *Neural Networks: Tricks of the Trade*, pages 421–436. Springer, 2012.
- [5] R. H. Brown and I. Matin. Development of artificial neural network models to predict daily gas consumption. In *Industrial Electronics, Control, and Instrumentation, 1995., Proceedings of the 1995 IEEE IECON 21st International Conference on*, volume 2, pages 1389–1394. IEEE, 1995.
- [6] E. Busseti, I. Osband, and S. Wong. Deep learning for time series modeling. Technical report, Technical report, Stanford University, 2012.
- [7] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.
- [8] 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*, 6(1):3–73, 1990.
- [9] E. D'Andrea, B. Lazzerini, and S. L. del Rosario. Neural network-based forecasting of energy consumption due to electric lighting in office buildings. In *Sustainable Internet and ICT for Sustainability (SustainIT), 2012*, pages 1–5. IEEE, 2012.
- [10] A. M. De Livera, R. J. Hyndman, and R. D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, 2011.
- [11] R. H. Dodier and G. P. Henze. Statistical analysis of neural networks as applied to building energy prediction. *Journal of solar energy engineering*, 126(1):592–600, 2004.
- [12] A. P. Douglas, A. M. Breipohl, F. N. Lee, and R. Adapa. The impacts of temperature forecast uncertainty on bayesian load forecasting. *Power Systems, IEEE Transactions on*, 13(4):1507–1513, 1998.
- [13] B. B. Ekici and U. T. Aksoy. Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, 40(5):356–362, 2009.
- [14] European Parliament and Council of the European Union. Directive 2009/28/EC, 2009.
- [15] E. C. Eurostat. *Energy balance sheets - 2010-2011 - 2013 edition*. Publications Office of the European Union, Luxembourg, 2013.
- [16] 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*, pages 1–6. IEEE, 2013.
- [17] R. Fujimaki, T. Yairi, and K. Machida. An approach to spacecraft anomaly detection problem using kernel feature space. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 401–410. ACM, 2005.
- [18] X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. JMLR W&CP Volume*, volume 15, pages 315–323, 2011.
- [19] P. A. González and J. M. Zamarreño. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, 37(6):595–601, 2005.
- [20] I. J. Goodfellow, D. Warde-Farley, P. Lamblin, V. Dumoulin, M. Mirza, R. Pascanu, J. Bergstra,

- F. Bastien, and Y. Bengio. Pylearn2: a machine learning research library. *arXiv preprint arXiv:1308.4214*, 2013.
- [21] D. Hawkins. Identification of outliers. *London: Chap*, 1980.
- [22] G. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [23] 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*, 16(1):44–55, 2001.
- [24] J. M. House, W. Y. Lee, and D. R. Shin. Classification techniques for fault detection and diagnosis of an air-handling unit. *TRANSACTIONS-AMERICAN SOCIETY OF HEATING REFRIGERATING AND AIR CONDITIONING ENGINEERS*, 105:1087–1100, 1999.
- [25] R. J. Hyndman. Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, 4(4):43–46, 2006.
- [26] R. J. Hyndman and Y. Khandakar. Automatic time series for forecasting: the forecast package for r. Technical report, Monash University, Department of Econometrics and Business Statistics, 2007.
- [27] B. Iglewicz. *Robust scale estimators and confidence intervals for location*. New York: Wiley, 1983.
- [28] 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*, page 264, 2000.
- [29] S. A. Kalogirou. Artificial neural networks in energy applications in buildings. *International Journal of Low-Carbon Technologies*, 1(3):201–216, 2006.
- [30] S. Katipamula and M. R. Brambley. Review article: methods for fault detection, diagnostics, and prognostics for building systemsa review, part i. *HVAC&R Research*, 11(1):3–25, 2005.
- [31] I. Khan, A. Capozzoli, S. P. Corgnati, and T. Cerquitelli. Fault detection analysis of building energy consumption using data mining techniques. *Energy Procedia*, 42:557–566, 2013.
- [32] M. Khashei and M. Bijari. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, 37:479–489, 2010.
- [33] A. Khotanzad, H. Elragal, and T.-L. Lu. Combination of artificial neural-network forecasters for prediction of natural gas consumption. *Neural Networks, IEEE Transactions on*, 11(2):464–473, 2000.
- [34] S. Lawrence, C. L. Giles, and A. C. Tsoi. What size neural network gives optimal generalization? convergence properties of backpropagation. 1998.
- [35] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.
- [36] K. Liano. Robust error measure for supervised neural network learning with outliers. *Neural Networks, IEEE Transactions on*, 7(1):246–250, 1996.
- [37] I. Mizera and C. H. Müller. Location-scale depth. *Journal of the American Statistical Association*, 99(468):949–966, 2004.
- [38] Nest. Energy savings from nest white paper preview. <https://nest.com/downloads/press/documents/efficiency-simulation-white-paper.pdf>, 2014.
- [39] 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*, 40(12):2169–2176, 2008.
- [40] S. J. Nizami and A. Z. Al-Garni. Forecasting electric energy consumption using neural networks. *Energy Policy*, 23(12):1097 – 1104, 1995.
- [41] M. B. Ohlsson, C. O. Peterson, H. Pi, T. S. Rognvaldsson, and B. P. Soderberg. Predicting system loads with artificial neural networks—methods and results from” the great energy predictor shootout”. *ASHRAE Transactions-American Society of Heating Refrigerating Airconditioning Engin*, 100(2):1063–1074, 1994.
- [42] D. K. Ranaweera, G. G. Karady, and R. G. Farmer. Effect of probabilistic inputs on neural network-based electric load forecasting. *Neural Networks, IEEE Transactions on*, 7(6):1528–1532, 1996.
- [43] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*, volume 589. John Wiley & Sons, 2005.
- [44] W. S. Sarle. Stopped training and other remedies for overfitting. In *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics (.’ .\.* pp. 352–360. *Interface Foundation of North America, Fairfax Station. VA, USA*, 1995.
- [45] I. Sutskever. *Training recurrent neural networks*. PhD thesis, University of Toronto, 2013.
- [46] G. W. Taylor and G. E. Hinton. Factored conditional restricted boltzmann machines for modeling motion style. In *Proceedings of the 26th annual international conference on machine learning*, pages 1025–1032. ACM, 2009.
- [47] J. W. Taylor and R. Buizza. Neural network load forecasting with weather ensemble predictions. *Power Systems, IEEE Transactions on*, 17(3):626–632, 2002.
- [48] S. Wu and J.-Q. Sun. Cross-level fault detection and diagnosis of building hvac systems. *Building and Environment*, 46(8):1558–1566, 2011.
- [49] G. Zhang, B. Eddy Patuwo, and M. Y Hu. Forecasting with artificial neural networks: The state of the art. *International journal of forecasting*, 14(1):35–62, 1998.
- [50] G. P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.
- [51] H.-x. Zhao and F. Magoulès. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6):3586–3592, 2012.

APPENDIX

A. SOURCES

“Very often scientific studies rely on complex textual explanations of what has been done to analyse the data that can overwhelm the reader that has to accept them as an

act of faith”⁵. To avoid this, the author provides all the sources⁶, while the raw data are protected by privacy of the Hogeschool van Amsterdam. The Universiteit van Amsterdam is working to make them available to achieve total reproducibility.

⁵<http://labs.biblioteca.uoc.edu/blog/?p=4332>

⁶<https://github.com/denadai2/energyUva>