

# 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  
marco.denadai@studenti.unitn.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 building a robust outlier detection system, which relies on a hybrid Artificial neural network (ANN) and an Auto-regressive Integrated Moving Average (ARIMA) model. ARIMA is usually suitable for linear prediction and ANNs are suitable for non-linear prediction. It is proved that together they can model the complex non-linear relationship between weather forecast variables and the gas consumption. The system labels outliers, which can be reported to the building manager. He can further analyse and fix the HVAC system minimizing in this way the energy waste. The system is a 2-phase process where in the first one the short-term (hourly) gas consumption is forecasted thanks to a historical time-series, and in the second phase outliers are detected on the basis of deviations from expected (or forecast) value. Outliers are detected without the necessity of possessing the previously labelled examples, thanks to a very precise forecast with a RMSE error of  $10 m^3$ .

## 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) [17]. Studies and states' directives about minimizing energy consumption and using renewable energy increased steadily with the reduction of fossil fuels, the border frictions with eastern countries like Russia, and the increase of various environmental problems. With this in mind, the European union, with a recent directive [16], has 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 country laws, energy diagnosis, but also new efficiency systems from the used appliances.

Forecasting energy demands has become one of the major research field in the energy departments because it can help gas utilities but also companies and families. Gas utilities buy gas from pipeline companies on a daily bases, so they need to know the needs in advance to be competitive. Companies and families have the aim of reducing the energy

consumption and increase efficiency.

Lately, big companies like Google also have shown their interest in this new market, developing thermostats which automatically control the house climate basing the decisions on the schedule of the users. Nest, a company acquired by Google, declared that customers saved the 11.3% of AC-related energy usage without compromising comfort [39], thanks to the automatic learning implemented in their thermostats. If on one hand the automatic thermostat program setting based on the people behaviour, can help them to save money, anomaly detection can decrease even more the energy consumption. These anomalies usually have a large impact on the final consumption and if caused by deficiency, they represents a huge and easy fixable problem.

In this paper an automatic outlier detection system is proposed, where days/hours with abnormally high and low energy consumption are labelled and reported to the building manager. He can further analyse and fix the HVAC system minimizing the energy waste caused by the outliers. 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, and a comparison between the predicted value/trend and the actual one to find outliers.

## 2. BACKGROUND

### 2.1 What is an outlier

An outlier, by definition [23], 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.

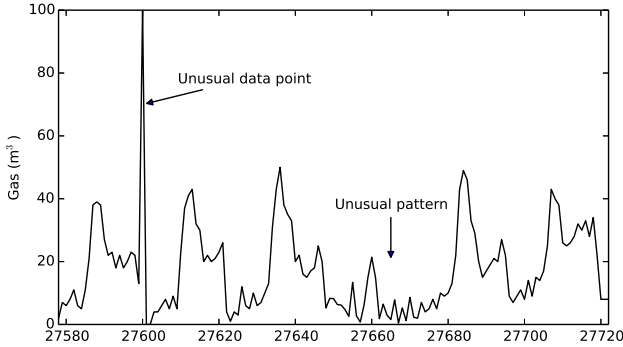
In this paper, outliers are very closed related to the problem of time-series forecasting, since outliers are declared on the basis of deviations from expected (or forecast) value. In this context a value is considered as outlier because of its relationship to its related data (*contextual* outlier [2] or conditional anomalies [49]). A sudden peak (fig. 1) in a time-series is a *contextual* outlier because its value is very different from the values of its adjacent items.

When a group of points are declared outliers, it is referred as *collective* anomaly or outlier [2]. It first appear at a point and then it affects the values immediately next to it. After a while, this effect dies out, leaving the time-series to the normal behaviour. This scenario is usually challenging to detect.

Outliers can have distinct main reasons:

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

In literature, outliers are also referred to as abnormalities, deviants, novelties or anomalies.



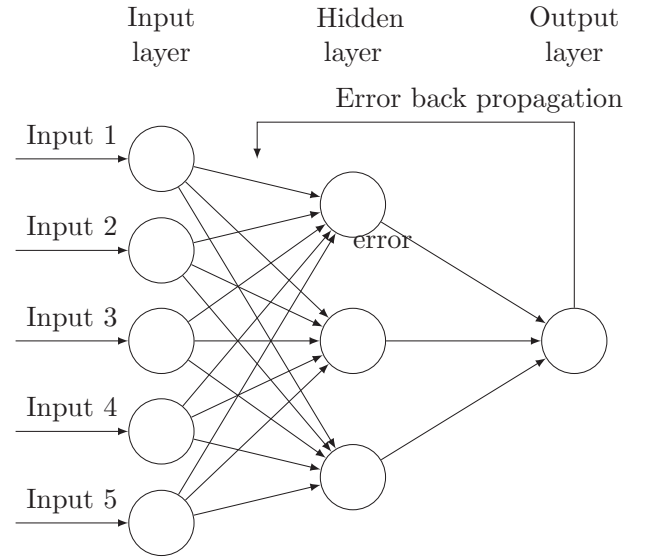
**Figure 1: Different types of outliers. On the left an unusual data point is presented, on the right an unusual pattern of changes can be recognized if compared to the other days shape.**

## 2.2 Artificial Neural Networks

Artificial Neural Networks (ANNs) were originally developed to mimic the brain functionality. There is not a widely accepted definition but by [4]: "A neural network is a circuit composed of a very large number of simple processing elements that are neurally based. Each element operates only on local information. Furthermore each element operates asynchronously; thus there is no overall system clock." From , it is possible to see a fully connected ANN with five inputs, 3 neurons on the hidden layer (so called because the ANN is like a black-box) and one output. The inputs are also called features and they represent the characteristics to describe the output. The hidden neurons will map this relation. Each connection has an activation which represents the importance (weight) of the connected neuron. ANNs with at least 1-hidden layer can map non-linear relations.

The hidden neurons are usually represented with logistic sigmoid units, which internally calculates the sigmoid function of the inputs, but also with hyperbolic tangent. Recent findings argued that rectifier units seems to be more biologically plausible [20] and they also seem to perform better in ANNs.

The ANNs are commonly applied with the *Stochastic Gradient Descent* algorithm, which tries to find the right weights of each connections to have the right output value. It is usually combined with the *Backpropagation* algorithm which calculates the error in the output layer and then back-propagates it to the previous layers in order to adjust the weights [46]. More information can be found in [5].



**Figure 2: Artificial neural network with *Backpropagation*. Example with 5 inputs, 3 hidden neuron in the first hidden layer, and one output (in the case of this paper, the gas consumption value).**

Nowadays, ANNs are the state of the art technique of many applications.

## 2.3 Autoregressive models

Let  $X_1, X_2, \dots, X_t$  be the values in an univariate time-series. In the *Auto-Regressive Moving Average* model, the value of  $X_t$  is defined in terms of the values of the last window 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. Sadly, 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 where series with deterministic trends should be differenced first, then an ARMA model applied. 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.

$$ARIMA(p, d, q)(P, D, Q)_m$$

where  $m$  is the number of periods per season. The upper-case notation 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 [45].

### 3. RELATED WORK

Since in this paper outliers are declared on the basis of deviations from expected (or forecast) value, this section is divided in related work in forecasting and outlier detection.

#### 3.1 Outlier detection

Outlier detection system is a wide area from introduction detection systems to fraud detection systems, from law enforcement systems to earth science anomaly detection systems.

Outlier detection can be supervised, when available data is labelled indicating previously known examples of anomalies, semi-supervised, where only example of normal data or anomalies are available, or unsupervised, where previous examples of interesting anomalies are not available. 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 [2]. In these cases a data-point is unusual due to one or more variables rather than a specific one (like in the supervised methods).

In energy consumption outlier detections, literature is usually based on the Gaussian error theory, stating that when the measurement accord with normal distribution, the probability that the residual falls in three times the variance is more than 99.7%. Therefore the residuals falling outside it, can be considered outliers. In [18] the author further improved this system considering a rolling window median which seems to improve the results when the distribution is not fixed. Supervised methods are usually based on classifications using trees, ANNs and other different algorithms, thanks to the possession of previous examples of anomalies. In energy consumption field 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. Despite this can be true, they are very difficult to apply in time-series data and the results are not usually excellent. For this reason in this paper a prediction algorithm will be built, where outliers are declared on the basis of deviations from expected (or forecast) value. The more the predictor will be accurate, the more it will detect abnormal data points.

#### 3.2 Forecasting

Traditionally, several techniques have been used for energy use forecasting, but it is needed to differentiate between short-term, medium-term and long-term energy forecasting. The former usually refers to prediction with a horizon of hours or days, the second refers to weeks, the latter refers to monthly or annual horizon. Long-term forecasting usually deals with data that rarely presents significant distortions and irregularities, so they have a small effect on the overall value. On the contrary, short-term forecasting has to deal with irregularities and sudden changes of the values (due to weather changes, human behaviour, etc.).

There are essentially five types of prediction models [57]: Engineering methods, Statistical methods, Artificial Neural networks, Support Vector Machines and Grey models. 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, Neural networks try to predict energy using an artificial intelligence network of interconnected neurons, Support vector machines are based in a machine learning algorithm and Grey models apply a mixture of the models. All the principal methods are extensively reviewed in [57] and [25].

Several techniques have been traditionally applied for energy use forecasting, and among the statistical methods Kalman filtering and ARIMA/ARMAX time-series techniques are the most famous.

The first reports about applications of Artificial Neural Networks (ANNs) were published in the early 1990's [10]. Since then the number of publications increased steadily. Kalogirou et al. [30] used back propagation neural networks 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 [41] tried a simple feed-forward NN and related the electric energy consumption to weather data and population, Taylor and Buizza [52] used an ANN with weather data (51 variables) to predict load of 10 days ahead. Gonzales [21] built an ANN to predict hourly energy consumption. Some researchers tried to specialize the ANNs: Neto and Fiorelli [40] compared generic ANNs with working days ANNs and week-end ANNs, Lazzerini and Rosario [11] specialized them to predict electric lighting with weather data.

Some researchers tried also to apply a hybrid model to increase the performance of the ANN. One example above all is [56] which applied a hybrid ARIMA and neural network model to forecast electricity use, another one is [31] who improved the previous one. This paper is based also on his work.

Until now only electric forecasting was presented, because the majority of the existing forecasters are related to electric forecasting. There are only few of them are about natural gas forecasting: Brown et al. [7] built one of the first predictor for natural gas consumption and Khotanzad et al [32] developed a two stage system ANN with very good results.

Even if ANNs might outperform traditional methods, the researchers are not still convinced about the results of the ANNs in this field. Nevertheless it is also stated that "a significant portion of the ANN research in forecasting and prediction, lacks validity" [1] and that most of the papers seems misspecified models that had been incompletely tested (no standard benchmarks, no synthetic data, etc.) [25]. This paper will try to avoid these mistakes.

It needs also to be pointed that ANNs are *multi-step ahead* forecasters, while Auto-Regressive methods are potentially useless in long ahead data-points.

### 4. PROPOSED SOLUTION

As stated before, in this paper an regression algorithm is proposed where outliers are declared on the basis of deviations from expected (or forecast) value.

A time-series is a sequence of data-points typically mea-

**Table 1: Buildings used**

Building name	Date interval	Number of rows
HvA 740 - NTH	01/2008 - 03/2014	54.725
Hva 761 - KMH	01/2008 - 03/2014	54.744
Hva 882 - WBW	01/2008 - 03/2014	54.647

sured at successive points of a uniform time interval  $t$  (eq. (1)).

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

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 are the standard “de facto” of energy forecasting and in recent works this problem is treated by combining weather forecast data. However, this relationship is clearly non-linear [25]. 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 composed by a *multilayer feed-forward* neural network with *backpropagation*.

## 4.1 Experimental data

The energy consumption datasets used are collected by Ebatech in different buildings of the Hogeschool van Amsterdam. These 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 types of features in each buildings, with different granularity. For this project three buildings are used: *HvA 740 - NTH*, *Hva 882 - WBW* and *Hva 761 - KMH*. In these buildings the company collected the energy consumptions, the gas consumptions as other variables.

The weather data was collected by KNMI<sup>1</sup> in Schipol, the Amsterdam airport 16 km far from the tested buildings. The available data consists in over 21 variables hourly collected. The proposed solution only uses few of them, as explained in the section 4.2, and they are used as forecast values: the measured weather conditions are linked to the previous hour of energy consumptions. It is necessary to consider that there will be an error in the built model, since the weather data is collected in a different location from the buildings positions and in practice the error will be larger than those obtained in this simulation, due to the effect of the weather forecast uncertainty [14, 44]. The advice is to keep it in mind before applying the methods contained in this paper with days forecasting.

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

### 4.1.1 Data analysis

The available dataset is conform to the nine characteristics of a good dataset as outlined by Few [19]:

**Volume** As demonstrated in table 1, the dataset seems adequately large to train an ANN and test the results.

**Historical** The dataset covers a very large period ranging more than 5 years, allowing to see similar patterns even with different yearly/seasons behaviour (one year could be different from another one for external factors like weather or building use).

**Consistent** The dataset of the selected buildings and for the variables that will be used, is consistent.

**Multivariate** In this paper a *multivariate* time-series problem is considered and the dataset is composed by more than 50 variables, measuring the different behaviours of the buildings and the weather.

**Clean-Atomic** The dataset is clean and atomic after the cleaning process (section 4.1.2).

**Clear** The dataset of the selected buildings and for the variables that will be used, is clear. The only problems were found on the not used variables: although they could be very useful, it was not possible to contact the buildings energy managers in order to understand them, for this reason they were excluded.

**Richy Segmented** This principle is not considered in this paper, but the dataset can be segmented in several meaningful segments like gas consumption, electric consumption, water consumption, number of people using the buildings, etc.

The gas consumption data is highly seasonal: daily and weekly cycles are quite perceptible, as it can be seen from fig. 4 and fig. 3. From the latter the weekly behaviour is clear: the last two days of the week (Saturday and Sunday) are completely different from the others and Monday seems a bit different from the rest of the days. Everyday, 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. The system reveals us that after a couple of hours, it decreases the consumption again. In fig. 4 the Temperature has a clear daily/hourly relation with the gas consumption while in fig. 5 the electric consumption is shown to be very smoothed and more regular than the gas one.

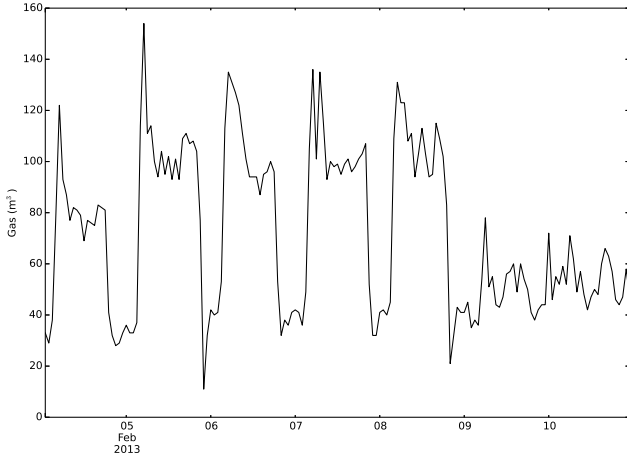
### 4.1.2 Data cleaning

The data was presenting some irregularities like repeated and missing data-points. Although the first one may not influence the performance of the ANN, it could lead to problems when other algorithms are used (like ARMA in this paper). Repeated data-points were deleted keeping only the last one, while missing data-points were reconstructed by linear interpolation. Cubic and spline interpolation were also considered but the performance was heavily affected by these methods.

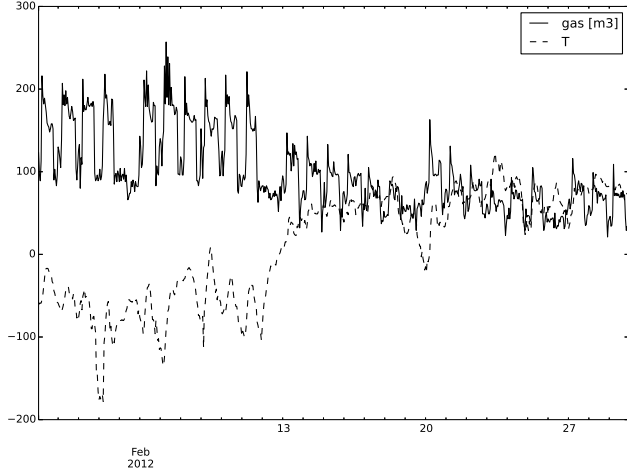
## 4.2 Artificial neural network forecaster

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

**Known** dependencies like date-time dependencies.



**Figure 3: Typical weekly and daily gas consumption behaviour.** 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.



**Figure 4: Typical monthly gas consumption behaviour and its relation with the temperature**

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

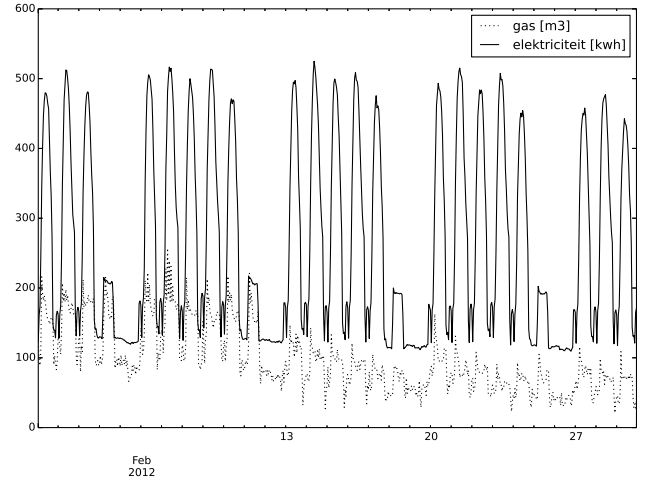
**Short/long-term** dependencies between variables.

Data scientist and experts are focused in known dependencies while the proposed ANNs will be focused on the hidden ones. The short/long-term dependency is realized by a moving window containing a “memory” of the previous states for the interesting variables, using a *Tapped delay line memory* [38]. These memories form a new set of states

$$\{\bar{x}_1(t), \bar{x}_2(t), \dots, \bar{x}_n(t)\}$$

from the original states

$$\{x(1), x(2) \dots x(n)\}$$



**Figure 5: Relation between electric and gas consumption**

where  $\bar{x}_i(t) = x(t - i + 1)$ . The window types will be explained later in this section.

Since the value to predict is time dependent, the first element to consider is adding the time feature. 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 4.1.1, fig. 4 and fig. 3). The day of the week is 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<sup>2</sup>. The day of the year is a number between 0 and 366, where the first one is the first of January. All these date-time features by means of their sine and cosine values as usual reported in literature [43, 13, 21]. This transform 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. This was believed to be an important measure to find the building usage and the holidays.

There are many factors that affect the energy needs of buildings. These factors can be divided in three main groups namely the *physical environmental*, the *artificial designing parameters* and the *human thermal discomfort*. The first 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 [15], but these variables were unavailable 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 all these data was available, the only significant weather variables found, were the temperature and the wind speed.

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

<sup>2</sup>Thanks to an Open source dutch week-end list <https://github.com/PanderMusubi/dutch-holidays>

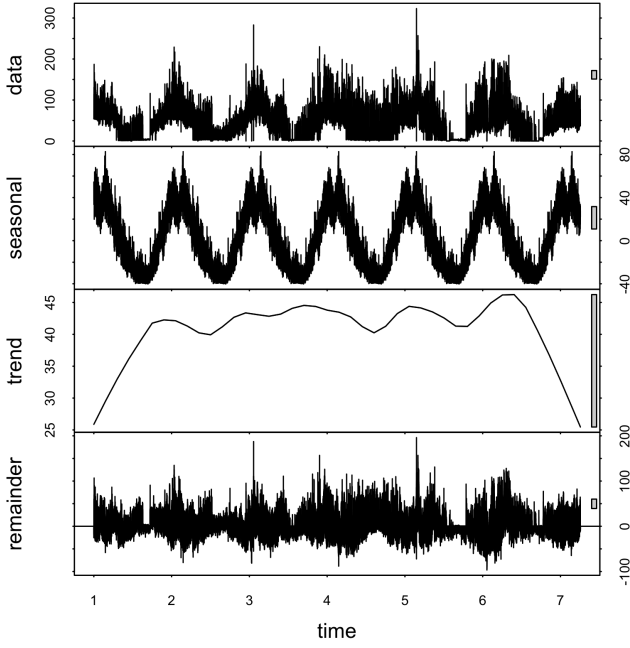


Figure 6: Yearly STL decomposition by LOESS

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, so it was not possible to know if this change would have an immediate effect on the HVAC system and/or its reaction time.

Gas usage has a daily cycle but there are also secondary weekly and annual cycles 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*, irregular component or difference. The time-series was analysed by the STL decomposition by LOESS [9](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* and the trend can be found from the historical part, the daily, weekly and yearly remainders were added as feature. For the same reason also the Temperature, the wind speed and the electric consumption was decomposed by the STL decomposition, resulting in stationary time-series added in the input.

In Zhang et al [55] it is stated that ANNs models really have advantages while dealing with a large amount of historical load data with non-linear characteristic, but the researchers neglected the linear relations including the data. For this reason a hybrid approach is proposed, where the ANN will be helped in linear forecasting by the popular method ARIMA (auto-regressive integrated moving averages), commonly known as the BoxJenkins approach. In order to apply ARIMA, the time-series was processed iteratively with a moving window of 21 days where the ARIMA

Table 2: ANN inputs

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	$\sin(2\pi(d)/366); \cos(2\pi(d)/366)$
Gas peak load'	$\max_{1 \leq k \leq 5} G(t-k)$
Gas sum load'	$\sum_{i=1}^5 G(t-i)$
Gas mean load'	$\frac{1}{288} \sum_{i=1}^{288} G(t-i)$
Temperature	$T(t)$
Wind speed	$FH(t)$
$\Delta T_{k+1}$	$T_{k+1} - T_k$
ARIMA forecast	$forecast(ARIMA(3, 0, 5), 1)$
STL year res.	$YearRes(t)$
STL weekly res.	$WeeklyRes(t)$
STL day res.	$DayRes(t)$
Year res. peak	$\max_{1 \leq k \leq 5} YearRes(t-k)$
Year res. sum	$\sum_{i=1}^5 YearRes(t-k)$
STL T res.	$Res(T(t))$
STL FH res.	$Res(FH(t))$
STL E res.	$Res(E(t))$

model was fitted. After the fitting it was forecasting the values of the next 24 hours, before moving the window and doing the same for the day after. The seasonal ARIMA fitting was done by the help of the Forecast R package [27] and its *auto.arima* method, which finds the 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)<sub>24</sub>. An ARIMA model with temperature dummy variables was tested, but it was not helping further the ANN: the simplest models was preferred.

Taking into account the points made in this section, the ANN is predicting the gas consumption “seeing” without knowing its shape, and its behaviour in the previous hours/days. This limit is surpassed by some rolling windows, which will somehow simulate the Recurrent neural networks behaviour. Two rolling windows were created for the gas consumption, memorizing the sum and the peak load of the previous five hours, and two moving rolling windows were created for the STL yearly residuals, memorizing sum and peak of them.

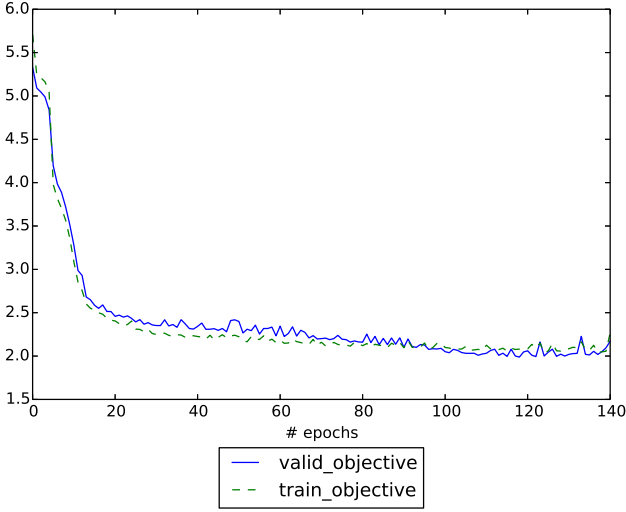
All the features that were not between the limits  $[-1, 1]$  were scaled to have a faster convergence [34] of the *Stochastic Gradient Descent* (eq. (4)).

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

where  $x_i$  is the original value and  $x'_i$  is the scaled one. Many practical tricks like the shuffling of the elements, the normalization and initialization were taken from [34, 6].

All the process described so far is also called *feature engineering* and was done almost iteratively, cumulatively introducing and removing features from the model and comparing the performance.

Choosing number of hidden units for the ANN is always a tricky task. As stated by [33, 47], 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 avoid bad local



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

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*.

#### 4.2.1 Architecture

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

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

where the error function  $p$  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.

Using the notations defined above, 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 [35]. In order to control the damage caused by outliers, in this paper the Least Mean Log Squares (LMLS) method (eq. (5)), presented by [35] is used. The ANN will try to minimize the Mean Log Squared Error (MLSE).

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

The ANN is a 1-hidden-layer *Multilayer Feedforward* ANN with a feedback structure, called *Backpropagation*. This ANN is composed by *Rectifier* neurons and one output linear node. Training is done by the *Stochastic Gradient Descent* algorithm with 10 batch size and is 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* [22] to construct and test the ANN and the R system with the *zoo* [54] and the *Forecast* [27] packages for the ARIMA process.

### 4.3 Outlier detection

According to Chebschev's theorem [3], almost all the observations in a data set of system states falls into the interval  $[\mu - 3\sigma, \mu + 3\sigma]$ , where  $\mu$  and  $\sigma$  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 interval<sup>3</sup> 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  $\sigma$ .

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

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

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

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

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

Therefore

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

## 5. EXPERIMENTAL EVALUATION

The ANN has been trained with early stopping, with a fixed number of training epochs (phases) or stopping the training when the validation error rate was increasing. All the results showed are obtained from a k-fold cross-validation techniques, 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 is always trained with 70% of the data, 15% is used for validation and the other 15% for testing.

Although most the Mean Absolute Percentage Error (MAPE) is considered a standard for examining the quality of the models prediction of energy load, it is an adequate error measure only if the loss function were linear and recent studies demonstrated it is not [30][29]. Moreover the percentage error is infinite if there are zero values on the series, frequent in intermittent data and in consumption data, and it puts a heavier penalty on positive errors than on negative errors [26]. Because of these disadvantages, this paper will only

<sup>3</sup>For the following description user *fabee* of *CrossValidated* needs to be mentioned: <http://tinyurl.com/19gvz65>.

consider the minimization of the Root Mean Square Error (RMSE), which penalize large errors, as suggested in [53]. As suggested in [25], for every experiment also the Mean Absolute Error (MAE) will be calculated.

$$\text{MAPE}^4 = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100 \quad (12)$$

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

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

where  $\hat{Y}$  is the vector of the  $n$  predictions and  $Y$  is the vector of the true values.

## 5.1 Synthetic experiments

The method is tested with synthetic generated data. Two days outliers were generated by different algorithms. In the first one the real consumption was 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.

## 5.2 Measured data experiments

The method is also tested with measured data, coming from a different type of the day. For example the gas consumption of a week-end was placed in a week-day, simulating a holiday. The purpose of this test was to show that the unusual pattern was detected. 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 table 1.

## 6. FUTURE WORK

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 multi-seasonality is present, an algorithm like TBATS can overpass the ARIMA one, detecting it. This non-parametric model described in

<sup>4</sup>MAPE errors will be calculated only on the non-zero values, to avoid the problems described before.

[12] could be substituted to the ARIMA one as feature of the ANN. At the moment it is very slow, but it is very recent, so it will be probably improved.

The daily pattern could be seen in the transformed Fourier space applying the Modified Discrete Cosine Transform (MDCT) [36]. In theory this could help as well to understand the pattern, but it was only applied once by [8], with scarce results.

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 perfect data, and this probably affected the performance. It is necessary to contact these building managers to further help the training of this algorithm.

The input variables were scaled standardizing them to a midrange 0 and 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. Some examples are [28] and the recent [37] can be the base of a future refine of the ANN inputs.

Before 2006, ANN 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 [24], 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 [8, 51, 50]. The *Pylearn2* [22] RNN framework is under development.

## 7. CONCLUSION

Some interesting behaviours were found though this work:

**Holidays** Regardless of whenever the school was operational, the first tests showed that the Ebatech system was normally heating the buildings (Christmas, in Tuesday 25th of December 2012, was heated like a normal 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. It seems that the system is wasting energy and this shape is totally different from the usual one (fig. 3). This lasts for months and it is clear that also the ANN training is affected by this outlier-like behaviour.

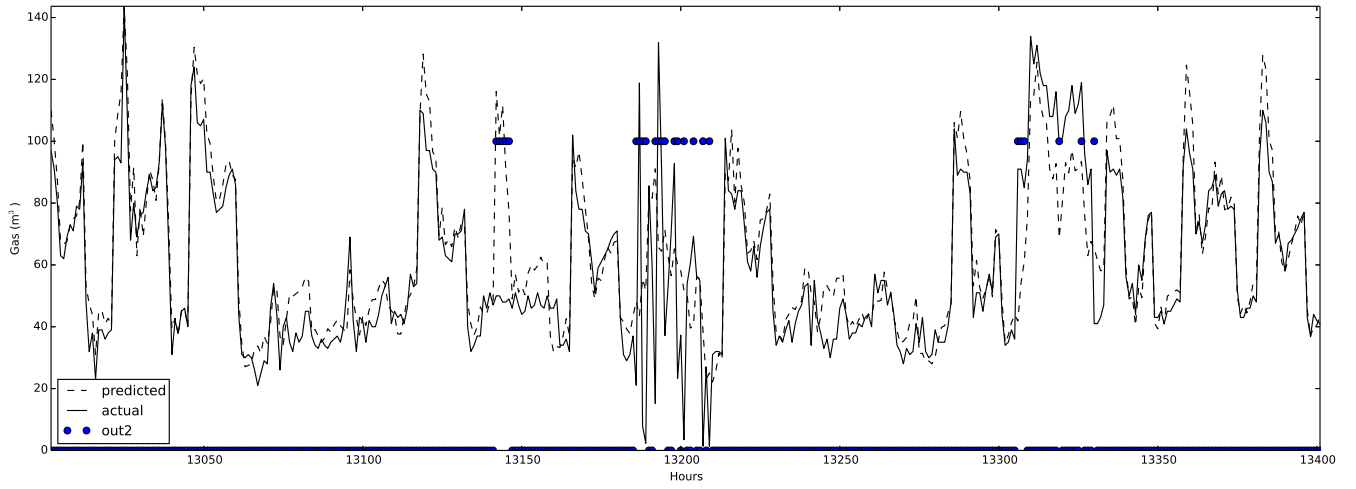
**Peaks** Around the initial days of September there is a huge bounce of the consumption (up to three times more than the maximal consumption of the year). Is it a test?

**August with heaters** In August 2009 and August 2011 the heaters were active even with not apparent cold summer.

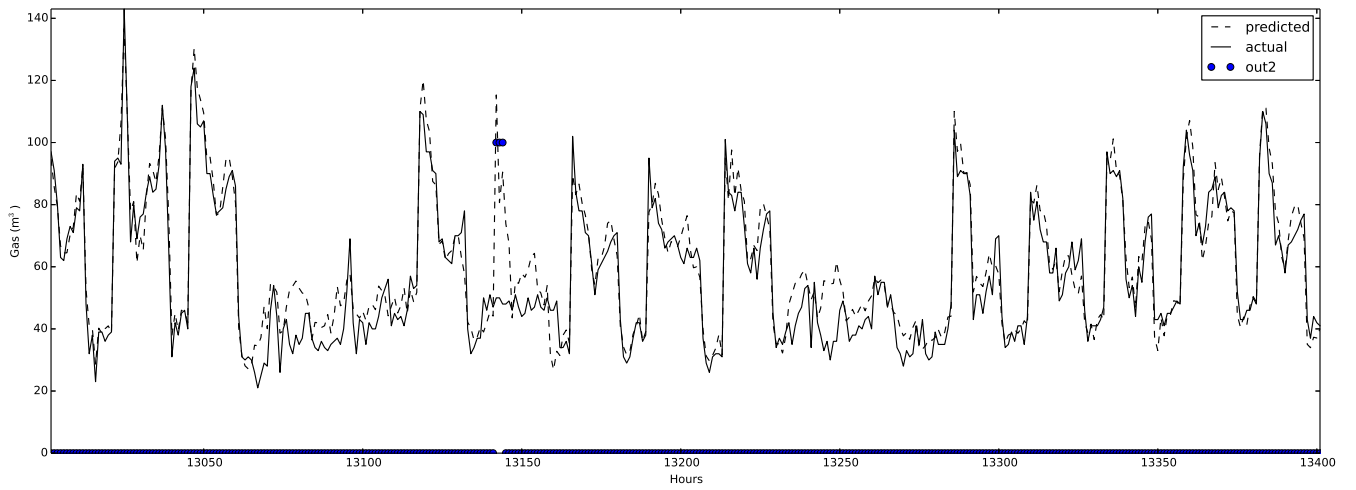
**Outliers** Some other outliers are found but they need to be confirmed by the managers, hopefully after the verification of the previously mentioned behaviours.

All these problems will be immediately reported to the company.

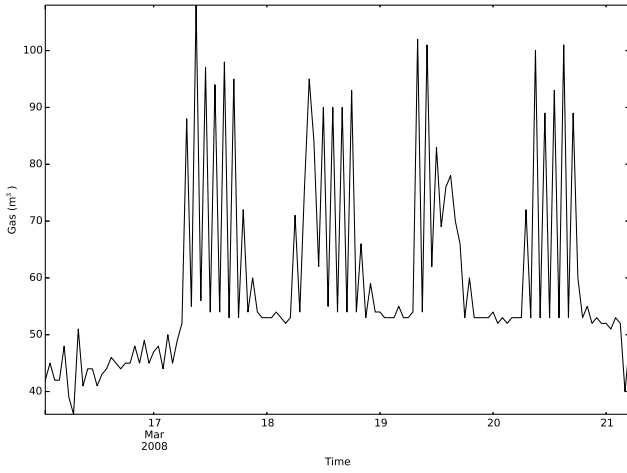




**Figure 8: Outlier detection with synthetic generated data. The circle represents the hours where an outlier is detected.**



**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.**



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

**Table 3: Best selected results 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.**

Model	neurons	epochs <sup>5</sup>	RMSE	MAPE	MAE
ARIMA <sup>6</sup>	-	-	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

An ANN with the standard cost function MSE was also trained, apparently resulting in a smaller RMSE error in a faster way (table 3). Although this can be true, the Hybrid MLSE model was more precise, detecting better the possible outliers. They contributed the most to the error.

Any model can't be treating accurately all the situations for a large amount of historical load data. The irregular fluctuation of the gas consumption was hardly predictable, for this reason 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 to forecast short-term gas consumption, which are very irregular and not easily predictable with classic methods. Since the predictor is very accurate (with RMSE of  $10 m^3$ ), the outlier mechanism is able to easily detect strange behaviours defining a threshold value in the confidence interval 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 electric consumption time-series. It is hoped that this could lead to new analysis of the energy consumption in public buildings.

<sup>5</sup>epochs to converge

<sup>6</sup>Calculated iteratively as described in section 4.2

## 8. ACKNOWLEDGMENTS

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

## 9. REFERENCES

- [1] M. Adya and F. Collopy. How effective are neural networks at forecasting and prediction? a review and evaluation. *J. Forecasting*, 17:481–495, 1998.
- [2] C. C. Aggarwal. *Outlier analysis*. Springer, 2013.
- [3] 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.
- [4] C. M. Bishop et al. Neural networks for pattern recognition. 1995.
- [5] C. M. Bishop et al. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.
- [6] L. Bottou. Stochastic gradient descent tricks. In *Neural Networks: Tricks of the Trade*, pages 421–436. Springer, 2012.
- [7] 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.
- [8] E. Busseti, I. Osband, and S. Wong. Deep learning for time series modeling. Technical report, Technical report, Stanford University, 2012.
- [9] 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.
- [10] T. Czernichow, A. Piras, K. Imhof, P. Caire, Y. Jaccard, B. Dorizzi, and A. Germond. Short term electrical load forecasting with artificial neural networks. *Engineering Intelligent Systems for Electrical Engineering and Communications*, 4(LRE-ARTICLE-1996-003):85–99, 1996.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] European Parliament and Council of the European Union. Directive 2009/28/EC, 2009.
- [17] E. C. Eurostat. *Energy balance sheets - 2010-2011 - 2013 edition*. Publications Office of the European Union, Luxembourg, 2013.
- [18] 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.
- [19] S. Few. *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press, 2009.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] D. Hawkins. Identification of outliers. *London: Chap*, 1980.
- [24] G. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [25] 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.
- [26] R. J. Hyndman. Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, 4(4):43–46, 2006.
- [27] 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.
- [28] B. Iglewicz. *Robust scale estimators and confidence intervals for location*. New York: Wiley, 1983.
- [29] 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.
- [30] S. A. Kalogirou. Artificial neural networks in energy applications in buildings. *International Journal of Low-Carbon Technologies*, 1(3):201–216, 2006.
- [31] M. Khashei and M. Bijari. An artificial neural network ( $i_1, i_2, \dots, i_n$  model for timeseries forecasting. *Expert Systems with Applications*, 37(1):479–489, 2010.
- [32] 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.
- [33] S. Lawrence, C. L. Giles, and A. C. Tsoi. What size neural network gives optimal generalization? convergence properties of backpropagation. 1998.
- [34] 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.
- [35] K. Liano. Robust error measure for supervised neural network learning with outliers. *Neural Networks, IEEE Transactions on*, 7(1):246–250, 1996.
- [36] H. S. Malvar. *Signal processing with lapped transforms*. Artech House, 1992.
- [37] I. Mizera and C. H. Müller. Location-scale depth. *Journal of the American Statistical Association*, 99(468):949–966, 2004.
- [38] M. C. Mozer. Neural net architectures for temporal sequence processing. 2007.
- [39] Nest. Energy savings from nest white paper preview. <https://nest.com/downloads/press/documents/efficiency-simulation-white-paper.pdf>, 2014.
- [40] 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.
- [41] S. J. Nizami and A. Z. Al-Garni. Forecasting electric energy consumption using neural networks. *Energy Policy*, 23(12):1097 – 1104, 1995.
- [42] O. U. of Catalonia. The importance of being (r)eproducible. <http://labs.biblioteca.uoc.edu/blog/?p=4332>, 2014.
- [43] 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.
- [44] 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.
- [45] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*, volume 589. John Wiley & Sons, 2005.
- [46] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. Technical report, DTIC Document, 1985.
- [47] 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.
- [48] T. Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Rückstieß, and J. Schmidhuber. PyBrain. *Journal of Machine Learning Research*, 2010.
- [49] X. Song, M. Wu, C. Jermaine, and S. Ranka. Conditional anomaly detection. *Knowledge and Data Engineering, IEEE Transactions on*, 19(5):631–645, 2007.
- [50] I. Sutskever. *Training recurrent neural networks*. PhD

thesis, University of Toronto, 2013.

- [51] 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.
- [52] J. W. Taylor and R. Buizza. Neural network load forecasting with weather ensemble predictions. *Power Systems, IEEE Transactions on*, 17(3):626–632, 2002.
- [53] R. Yao and K. Steemers. A method of formulating energy load profile for domestic buildings in the uk. *Energy and Buildings*, 37(6):663–671, 2005.
- [54] A. Zeileis and G. Grothendieck. zoo: S3 infrastructure for regular and irregular time series. *arXiv preprint math/0505527*, 2005.
- [55] 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.
- [56] G. P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.
- [57] 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”, as stated in [42]. To avoid this, the author provides all the sources through this website: <https://github.com/denadai2/energyUva>. Some sources are also available for the popular Pybrain [48] library.

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.