THE 23rd ECMI MODELLING WEEK
EUROPEAN STUDENT WORKSHOP
ON MATHEMATICAL MODELLING
IN INDUSTRY AND COMMERCE

Wrocław University of Technology, Wrocław, Poland, August 23-30, 2009

Report of project group 5 on

Influence of physical and economical factors on electricity spot market price

---

The 23rd ECMI Modelling Week
23-30.08.2009, Wrocław, Poland

Influence of physical factors on electricity spot market price

Hassine Baya
Chhitiz Buchasia
Martin Rykfors
Philip Anton de Saint-Aubain
Irene Vecchio
Ibrahim Wangwe
Esubalewe Lakie Yedeg

Instructor: Matylda Jabłońska

Contents

1 Introduction

This project investigates the influence of physical and economical factors on electricity spot market price. Electricity prices depend on many different factors. They reveal seasonality respective to electricity demand periodicity. However, they sometimes also show outstanding behaviour called *price spikes*. If we can recognise data points which are caused by rare unexpected events, we can remove them, because the time series model should only depend on the normal circumstances. The aim of this study is to identify those of the spikes in the available data that can be clearly explained by particular physical events, then remove them from the data in an appropriate way and, finally, fit regression models explaining price trends with use of background (explanatory) variables.

In Figure 1 we show an example time series of the electricity prices in New Zealand over a range of ten years.

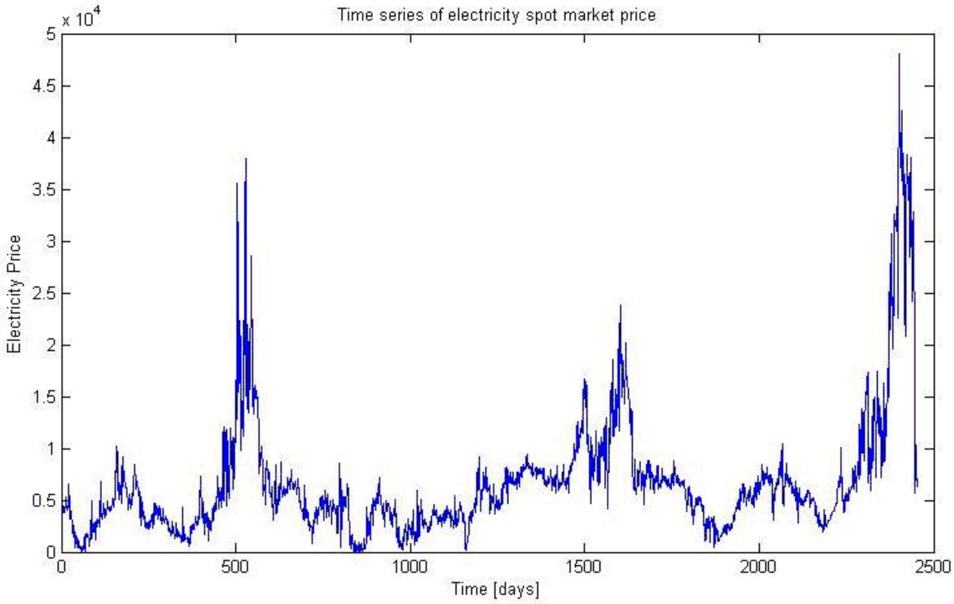

Figure 1: Example of daily price series. Data relative to New Zealand collected in Benmore over a ten years range.

We can see that sometimes the price goes rapidly up/down. There is no simple pattern in this behaviour. In the following, these events are called spikes. If one wants to construct a base model for the electricity prices (describing the general trends but not the spike occurrence) the spikes may bias the model estimates and, therefore, make it less accurate. Thus our task is to find the spikes and see if there is a simple explanation of their occurrence. E.g. if there has been a power supply breakdown, it could explain why the prices would suddenly raise up ten times as high as they were before. If there is a simple explanation then we can argue for removing the spiky data points and replace them with an approximation. This trick will make it easier to make an appropriate estimation of a regression model for the electricity prices because the variance of the data would be smaller. This makes sense because regression models are not the best ones to cater for extreme observations.

The report is structured as follows: Section 2 introduces the data sets available for the study, Section 3 defines the spikes and shows an algorithm to remove them, Section 4 presents two models of the prices: a simple linear regression and a probability distribution model. Finally, Section 5 concludes.

2 The Data

We handle data relative to two different locations. One is form New Zealand electricity market and the other is from Nord Pool. By Nord Pool we understand Nordic Power Exchange, i.e. the power market for Sweden, Norway, Finland, Denmark and a northern region of Germany. Both areas contain data from a number of different locations (see Figure 2).

New Zealand electricity market functions on the two main country islands, the North Island and the South Island. There is a high dependence of the two islands on each other for the electricity supply. The production of electricity is shared by the two islands equally, but the consumption is about 70% in the North Island as compared to 30% on the South Island. Also, during the winter the water reserves of the South Island are protected and therefore North Island has to supply electricity to South Island.

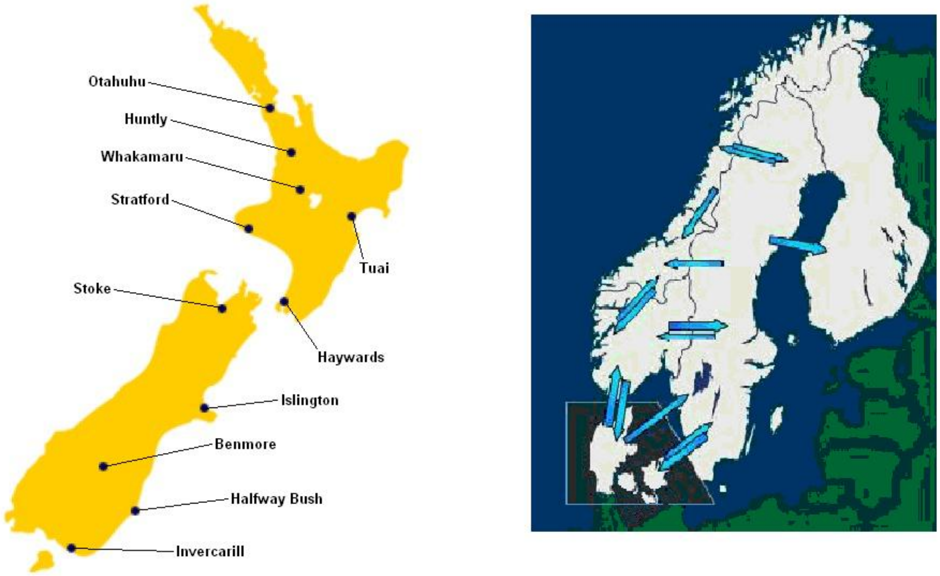

Figure 2: On the left, the two New Zealand islands with the electricity nodes (11 out of the total of over 200). On the right, the Nord Pool area. The arrows indicate the main electricity exchange lines on a particular day.

Each area uses and produces some electricity. Sometimes a particular area consumes more and sometimes less electricity than it produces. What then happens is that they buy electricity from each other. The electricity that has been bought and sold is called the flow. Due to the wire and other physical limitations, it is not possible to send an infinite amount of electricity from one area to another. The limitations can change due to physical factors of different weather conditions.

In particular, for each region we have the following data:

New Zealand

- Spot market price
- Temperature
- Rainfall
- Weather phenomena (Gale, Snow, Hail, Lightning, Thunder, Fog and Dew)
- Constraint (1 if the inflow is equal to inflow capacity, 0 otherwise)
- Information from operational reports.

Nord Pool

- Spot market price
- Temperature
- Rainfall
- Snow depth
- Production (How much electricity the area produced)
- Consumption (How much electricity the consumers used)
- Inflow (How much electricity the area buys from neighbours)
- Capacity (How much electricity there maximal can be transmitted from the neighbours)

Most of the mentioned variables are claimed to drive the general electricity spot price trends.

3 Spikes

Section 2 presented a list of explanatory variables available for the study of the mentioned markets. These are the data we use in order to explain the occurrence of the spikes. In theory, the market spot prices should depend on two phenomena: supply/demand and the electricity generators' costs. If a particular area owns wind farms and there wind is strong particular period, one can use the wind energy as the cheaper alternative. Thus it should lower the prices. Also, if much more energy is produced than the consumers can use, the energy price should go down.

Some of the data can tell something about the spikes. For instance, if consumption suddenly one day is much higher than production, that could maybe explain a spike. Norway gets a lot of electricity from hydroelectric plants, so if the rainfall has been very low for some time, it could make a spike. The operational report can tell if there has been a power supply breakdown, which could make the prices goes up.

Our objective is first to determine which spikes can be removed due to simple explanations on physical or economical factors (Section 3. In Section 4 instead, we will deal with the problem of creating a good model for approximating the prices, in the hope that it would possible to foretell their future trends.

There has been a lot of work done in modelling of the Electricity price time series. The most difficult part of modelling the electricity price time series is sudden extreme changes in price in a short duration of time. These changes are of great importance to the manufacturers and providers. If they were able to predict the reasons for these sudden changes then they could hedge against them to eliminate risk of loss. This would then help them to save a lot of money and acquire the required amount of electricity on time. Some of these changes are explainable and they can be removed for modelling purposes. Here we try to look at the prices from the New Zealand & Nord Pool and try to identify if the sudden dramatic changes were explainable from the data provided.

We call these sudden changes as spikes. Furthermore, the spikes can be positive and negative. By positive spikes we mean there was an increase in the price, whereas, by negative spikes we identify those spikes where there was a decrease in price. The negative spikes are easier to explain. They do not add to losses of the buyers, therefore we are not going to consider them in our report. It is the positive spikes which are more important as the reason for their occurrence is not easily explainable and they may lead to huge losses to the buyers.

3.1 Spike Definitions

The next question is how do we define them? There can be many ways in which we can look at a sudden change and define it as a spike. Some of them which we considered in our study are described below.

3.1.1 Global Spikes

This way of defining is based on the idea that we identify those prices in the time series which do not lie in an α-confidence interval around the global mean. This method looks at the entire time series at once and then identifies the spikes based on how data behaves over the entire duration.

In Figure 3 we can see that this method does not appear to be a good one as there are many local trends in the prices including especially the last year of the available data. Therefore, we look at a different method.

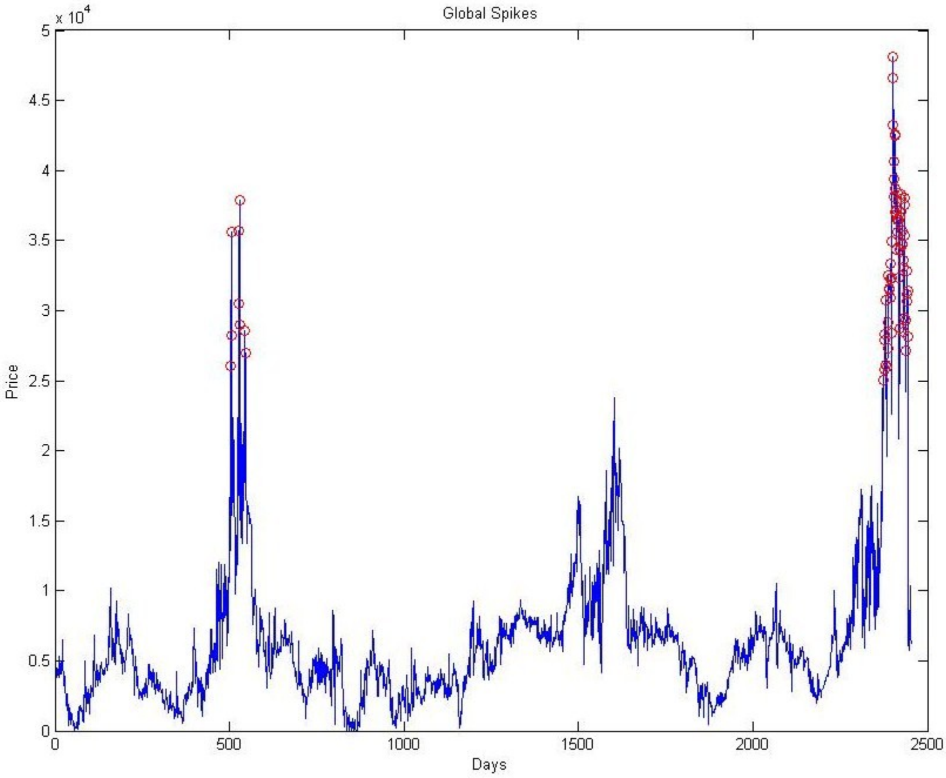

Figure 3: Price time series with Global Spikes

3.1.2 Window Spikes

The above method works fine if the average price is more or less constant over the time (the series is stationary). But if there is a trend in the price time series then the above definition does not work. Therefore, we decide to consider only a window around the price under consideration, not the entire price time series. Thus, in this method we identify the spikes as those prices which do not lie in an α-confidence interval around the mean of the window. This method is better than the previous one as it works fine with both constant price time series as well as with ones having trend.

We fix the size of the observation window equal to an interval of 50 days and $\alpha = 0.95$. As we can see from Figure 4, it is better suited for identifying spikes. But as can be observed there are many points which could have been identified as spikes by eye but statistically speaking they are not. We only consider this method to identify the spikes in our further study.

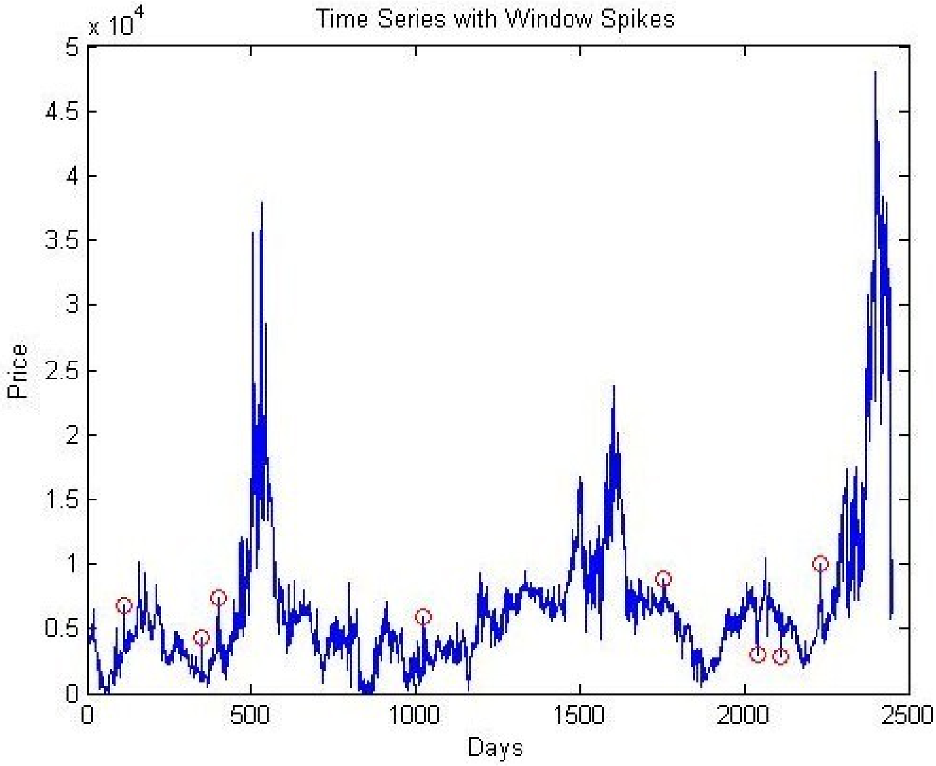

Figure 4: Price time series with Window Spikes

3.1.3 Local Spikes

Sometimes there are sudden changes of price for few days. The situation is that one day it increases suddenly and then the other day it decreases and it continues like that for next few days. In that case none of the above two methods will be efficient. In this method we consider only those prices as spikes which show a certain % change in the price over the previous day. This method is best suited for fluctuating markets where prices change drastically very often.

3.2 Spike Removal

Spikes which can be explained by the environmental, economical and other factors have to be removed when one wants to apply regression model. Spikes can make the model estimates biased and, therefore, mislead possible prediction. We hereby present some methods which we considered for removing the spikes by an appropriate value.

3.2.1 Mean value

This is the most simple and the most naive method. In this method we just replace the price at the spike by the mean. This mean can be

- Global mean: replacing the spike by the global mean. Not so effective as all the spikes will be replaced by the same value. The prices of the recent past do not play any role in determination of the corrected value of price.
- Window mean: Better than the above method. We take a window around the spike and replace the value at spike by the mean of the window.

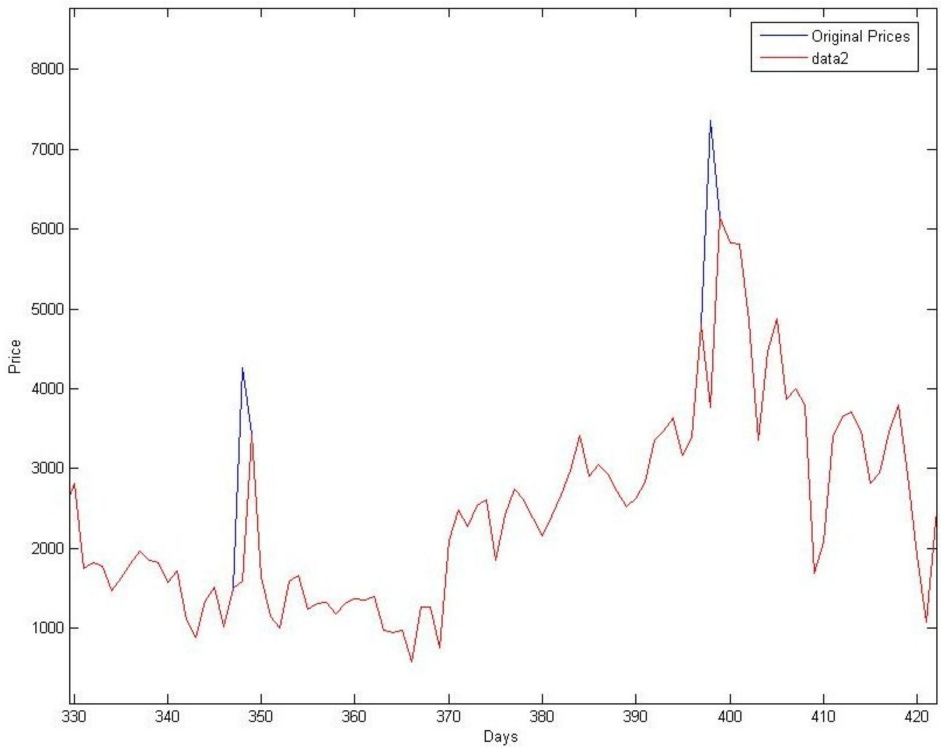

Figure 5: Spike removal by window mean

Figure 5 shows how this method works. We can see only the price at the spike has been modified to bring it down to the mean value of the selected window around it. The spike removal methods can be improved further as described below.

3.2.2 Window Shift

We observed that the price around the spike depends on it. When removing the spike therefore these values should also be changed in a logical way. In this method we modify the window around the spike in such a way that the values nearest to spike are modified by a large amount. As we move away from the spike the modification becomes less and less and at the two ends of the window it is almost zero. This method better suits to the real world. Figure 6 illustrates this method.

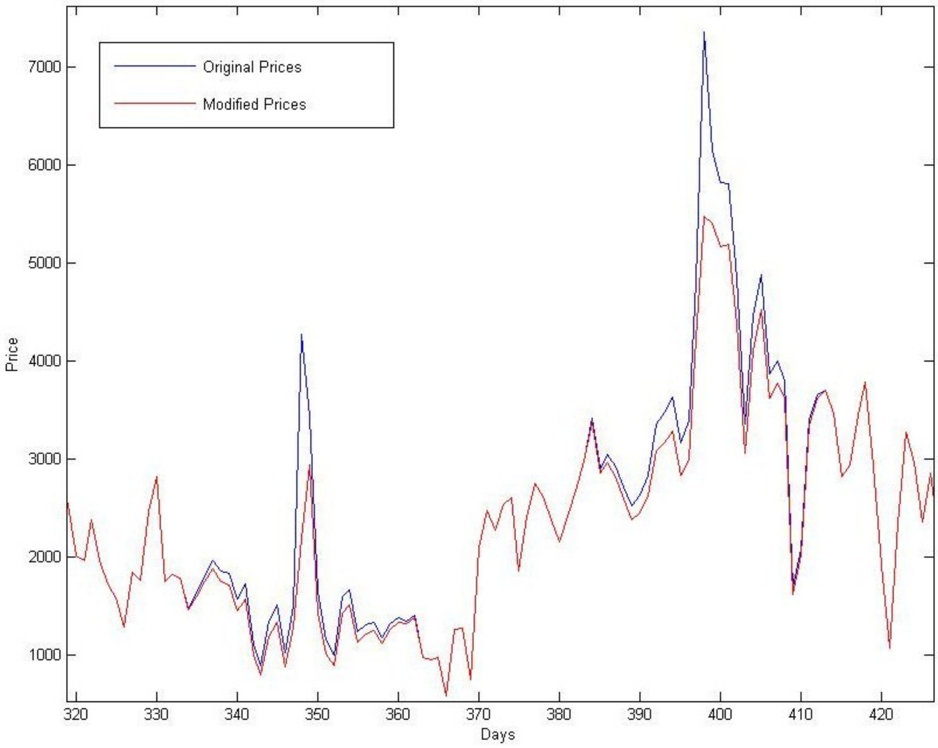

Figure 6: Spike Removal by Window shift

3.3 Explanation of Spikes in New Zealand

We used the Window Spikes method to find the spikes in the New Zealand electricity price time series. For all the 11 nodes combine the method identified 170 spikes, 144 positive and the remaining negative. Below is the list of days when a spike occurred with the Island in which they were identified.

- North Island:

  - Positive Spikes: Days 46(6), 87(6), 113(1), 209(3), 210(1), 348(6), 401(6), 462(5), 630(2), 800(6), 803(6), 861(6), 896(1), 897(1), 972(6), 1025(1), 1030(5), 1130(6), 1131(6), 1197(5), 1756(6), 1843(5), 1871(6), 2277(6).
  - Negative Spikes: Days 1810(6), 2040(6), 2111(4)
  - Altogether 124 spikes.

- South Island:

  - Positive Spikes: Days 113(5), 348(5), 398(5), 1021(5), 1129(2), 1144(2), 1438(2), 1756(5), 2234(5).
  - Negative Spikes: Days 2040(5), 2111(5)
  - Altogether 46 spikes.

The digits in bracket denotes the number of nodes which had spike on the same day.

We can clearly see that most of the times if there is a spike at one of the nodes in an Island there is a very high probability that there will be a spike at all the other nodes as well. This however is very true if we look at this phenomena from the no-arbitrage point of view. Now we present the reasons for days which we could considered as explainable from the data available.

- One of the most important reasons for the occurrence of spikes was the Forced Outage of HVDC or HVAC at more than one instance during the day. Days: 46, 87, 113, 209, 210, 348, 800, 803, 861, 1021, 1025, 1129, 1130, 1131.

- Inflow Constraints met. If the Inflow constraints are met than there is a possibility of a spike. If the inflow constraint is met at a very critical node like Haywards(3), then it is likely to affect the entire North Island prices. Days: 87(1, 3), 113(1, 11), 348(3), 398(3), 401(3), 462(3), 861(3, 8, 11), 897(2), 972(1, 3), 1021(1, 3), 1025(1, 3), 1030(1, 3), 1144(6), 1197(1, 2, 3, 10), 1843(1), 1871(3, 8). The values in the bracket respectively denote: Benmore, Halfway Bush, Haywards, Huntly, Invercargill, Islington, Otahuhu , Stoke, Stratford, Tuai, Whakamaru.

- Sometimes the spike should have continued for a few days because of very long duration of power outages lasting for days in some regions, but we observed no spikes the next day. This was explained by heavy rains following the day that had a spike. The

New Zealand market is highly dependent on Hydro Power generation. Day 209 at Otahuhu, 1129, 1130, 1131.

- Heavy rainfall on the following day may also be a reason for the spike. If there is a sudden drop in the prices because of heavy rainfalls this can lead to a forced spike. Days: 113.

- Another interesting observation was that there were spikes around the same time of the year every year. Perhaps there was yearly maintenance of the plants which may have lead to spikes. Days: 46, 401, 1131, 1871.

- Hydrological Storage is very important. Whenever there was a spike the hydrological storage was much less than the average hydrological storage for the entire time series. We later on study the hydrological storage to design the pricing model.

- Prediction of hailstorms, lightning for the next days.

  – Day 2040: Lightning 2 days later at node 8.
  – Day 2111: Snow and Hale one day later at node 5.
  – Day 1144: Lightning and Thunder 1 day after, Lightning and Hale 2 days after at node 8.
  – Day 2234: Lightning after 3 days at node 8.

- Sudden changes of temperature.

- One other reason could have been some festivals where there is a sudden increase in the consumption of electricity.

If we look at all these reasons combined we can remove a large number of spikes.

3.4 Explanation of Spikes in Nord Pool

In case of Nord Pool we got 140 spikes by method of Window spikes. Out of those, 88 were positive and the rest negative. As we can already observe there are considerably more negative spikes in this region as compared to the New Zealand region. The reasons for spike in this region was sometimes different from the ones observed in the above case. The list below gives an idea of the distribution of spikes in the region.

- System:
  Positive Spikes: Days 29, 347, 350, 389, 767, 1098, 1848, 2576, 3660.
  Negative Spikes: Days 905, 1270, 1956, 2026, 2368, 2684, 3041, 3489.

- Finland:
  Positive Spikes: Days 29, 347, 350, 389, 516, 767, 1098, 1848, 2265, 2379, 2493, 2534, 2576, 2577, 2960, 3086.
  Negative Spikes: Days 905, 2026, 2096, 2313, 2684, 3041, 3489.

- Sweden:
  Positive Spikes: Days 29, 347, 350, 389, 516, 581, 631, 767, 1098, 1848, 2265, 2540, 2576, 2577, 2594, 2960, 3086.
  Negative Spikes: Days 905, 1820, 2026, 2096, 2684, 3041, 3489.

- Norway:
  Positive Spikes: Days 347, 350, 390, 767, 1098, 2254, 2580, 2947.
  Negative Spikes: Days 668, 905, 1270, 1956, 2018, 2368.

- Denmark West:
  Positive Spikes: Days 347, 350, 389, 767, 1098, 1848, 2576, 3660.
  Negative Spikes: Days 905, 1270, 1956, 2026, 2368, 2684, 3041, 3489.

- Denmark East:
  Positive Spikes: Days 767, 1098, 1848, 2576, 3660
  Negative Spikes: Days 905, 1270, 1956, 2026, 2368, 2684, 3041, 3489

- Kontek:
  Positive Spikes: Days 2576, 3645
  Negative Spikes: Days 2684, 3041, 3489

- Bergen:
  Positive Spikes: Days 29, 347, 350, 389, 767, 1098, 1848, 2576, 3660
  Negative Spikes: Days 905, 1270, 1956, 2026, 2368, 2684, 3041, 3489

- Tromso:
  Positive Spikes: Days 29, 347, 350, 389, 516, 767, 1098, 1848, 2265, 2379, 2493, 2534, 2576, 2577, 2960, 3086

Negative Spikes: Days 905, 1820, 2026, 2096, 2313, 2684, 3041, 3489

Data for some of the stations was not available from the starting dates, that is why we see inconsistency in the occurrences of spikes for these places. The main reasons possible are:

- Local Production. In Denmark the local production is mainly based on oil and coal. Production of electricity by these materials is costly. So whenever there is a sudden requirement which cannot be met by buying electricity from providers, Denmark has to produce electricity at local plants which may have led to spikes in both parts of Denmark. Moreover, the share of wind generation is considerably high, but this type of production is the most uncertain in capacity forecasting terms, therefore, prices in Denmark are a lot more volatile.

- Very cold winters when the consumption increases. If this continues for a long time then the water reserves are used up. This then leads to usage of other more costly sources.

- Norway on the other hand depends entirely on other nations to get the extra requirements to be fulfilled. If there is a sudden increase in demand then it has to buy electricity from Sweden or Denmark which could have led to spikes.

- Low rainfall at Oslo may have also led to some of the spikes as it is an important node in the entire Nord Pool circuit.

- We observed that whenever there was a spike there was a high ratio of inflow to consumption.

Most of the spikes in Denmark, Norway could be explained by combining more then one of the above. From the other regions only some of the spikes can be considered as explainable. But as already stated above, the price spikes occur at almost all the places simultaneously.

If we can explain the occurrence of spikes at key nodes then it may tell us a lot about the spikes at other locations. The main point now is to identify those nodes which are very critical to the entire network under consideration. We think it will be sufficient to study only these nodes and then the condition of the other nodes can be determined exactly form these few nodes.

4 Price Models

In this section, we will present some models to simulate the prices. Mainly we will use regression models and one based on probability measures. For the theoretical deliberation here reported, we refered to [1].

4.1 Linear Regression Model

The linear regression is an important part of the linear model. Generally, the linear model describes a situation in which the average of a random vector Y, that represent the observations, are linearly dependent on an explanatory variable X with coefficient θ

$$Y = X\theta + \epsilon, \quad E(Y) = X\theta, \quad Var(Y) = Var(\epsilon) = \sigma^2 I \tag{1}$$

The random part is represented by an additive error ϵ whose components are centered, have the same variance σ^2 and uncorrelated. The observations are random variables of same variance and remain uncorrelated with each other. Their averages are different and represented by the component of vector $X\theta$. This vector is an unknown linear combination, whose coefficients are the components of the parameter θ, columns of the matrix X, which itself is known. The diversity of models come from the nature of this matrix. In the case of simple regression or multiple columns of X are quantitative variables as Y, possibly observed, but assumed perfectly known that is not random. In analysis of variance, the elements of X is valued in $\{0, 1\}$, they indicate the presence or absence in each observation, the different modalities of external quantitative variables. The analysis of covariance is the juxtaposition in X of these two types of variables.

The value of the linear model is to provide a simple way, by the method of least squares estimators of parameters θ and σ^2, their properties and confidence intervals and tests under Gaussian assumption. Moreover, a geometric representation facilitates the understanding of results. However the context of linear regression is very different from that of analysis of variance or covariance and computational techniques are very specific. The simple linear regression approach allows the linear model so elementary. In this case the variable of interest is facing one other variable.

The multiple linear regression is an immediate extension of the simple linear regression, particularly in its geometrical interpretation in space observations. The essential difference lies in the formalism through different scriptures matrix estimators and their variance. The generalized linear model is to consider a matrix X containing p columns and hence a parameter θ of dimension p. Specifically, the model of multiple linear regression implies that $y_1, y_2, ..., y_n$ are the observations of random variables $Y_1, Y_2, ..., Y_n$ satisfying

$$Y_i = \sum_{j=1}^{p} \theta_j x_{ij} + \epsilon_i, \quad i = 1, ..., n. \tag{2}$$

4.1.1 Application

As we have seen in the previous section, in the linear regression model, the dependant variable is assumed to a linear function of the independent variables plus the error introduced to account for all other factors

$$y_i = \beta_1 x_{1i} + \cdots + \beta_n x_{ni} + \epsilon_i$$

In this regression model, y_i is the dependent variable which depend on the independent variables $x_{1i}, \ldots, x_{ni}$ and ϵ_i is the disturbance or error term. Our goal is to obtain the unknown parameters $\beta_1, \ldots . \beta_n$, which indicate how a change in one of the independent variable affects the value taken by the dependent variable.

Nonlinear regression model extends the linear regression model for use with a much larger and a more general class of functions. As the name suggests, a nonlinear regression model is any model of the form

$$y_i = \beta_0 + \beta_1 f_1(x_{1i}) + \cdots + \beta_n f_n(x_{ni})$$

where at least one functions $f_1, f_2, \ldots, f_n$ are nonlinear.

To apply this regression model, we will consider the data in the New Zealand and Nord Pool separately.

In the New Zealand, we need to find a relation how the price depend on temperature, rainfall, storage capacity and demand. Since we have the dependant variable y = price and independent variables x_1 =temperature, x_2 =rainfall, x_3 =storage capacity and x_4 =demand, the nonlinear regression model will be

$$y_i = \beta_0 + \beta_1 f_1(x_{1i}) + \beta_2 f_2(x_{2i}) + \beta_3 f_3(x_{3i}) + \beta_4 f_4(x_{4i})$$

where the parameters $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 have to be determined for the appropriate choice of functions f_1, f_2, f_3 and f_4.

The choice of the functions is in such a way that the regression model best fits the data. To determine whether the functions give as the best regression model or not we use the least square error method. That is we tried to find a regression model with minimum residual sum of squares.

To determine the parameters $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 we solve the following system of equations

$$y = X\beta$$

where

- y is an m dimensional vector,
- β is a vector of parameters (i.e. $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$),
- X is an $m \times 4$ matrix, note that $m =$ the number of samples on the data such that the ith row of X is given by $X_i = (1, f_1(x_{1i}), f_2(x_{2i}), f_3(x_{3i}), f_4(x_{4i}))$.

Therefore, the parameter vector β is given by

$$\beta = X \backslash y.$$

For instance, to the first node in the New Zealand we choose the functions $f_1(x_1) = x_1, f_2(x_2) = x_2, f_3(x_3) = (\frac{1}{10})^{x_3}$ and $f_4(x_{4i}) = x_4$, with minimum mean square error. For these functions the parameters are determined and given by

$$\beta = 10^3\dot{(} - 7.6607, 0.1012, 5.5153, 0.0001, -0.0495)).$$

A plot of the original price data and the price obtained from the regression model can be seen in Figure 7.

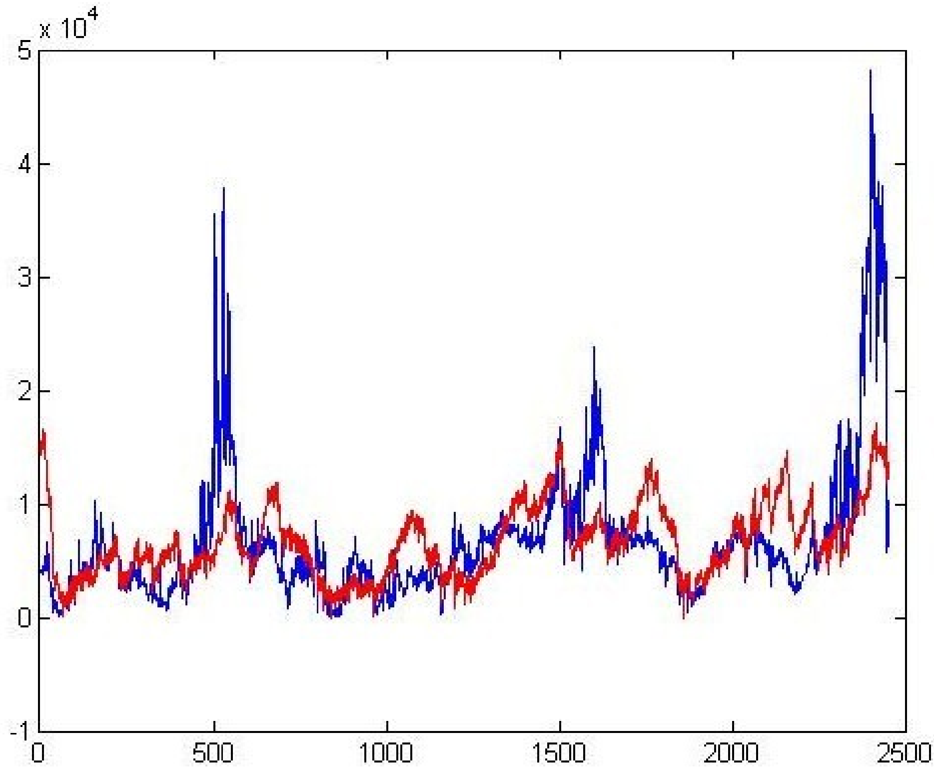

Figure 7: Plot of the original data and the price obtained from the regression model

Similarly we can extend the same procedure to all the remaining ten nodes.

In the Nord Pool we have 9 areas. Since we use the same procedure for these nine areas, here we will consider only the data on Stockholm.

In Stockholm there are five independent variables (factors) which influence the dependent variable (price). Therefore, the regression model is

$$y = \beta_0 + \beta_1 f_1(x_1) + \beta_2 f_2(x_2) + \beta_3 f_3(x_3) + \beta_4 f_4(x_4 + \beta_5 f_5(x_5)$$

where,

$y =$ price
$x_1 =$ temperature
$x_2 =$ rainfall
$x_3 =$ consumption
$x_4 =$ production and
$x_5 =$ snowfall

Here we choose the functions $f_2(x) = x_4$ and the other functions are linear. For thus functions the parameters obtained are

$$\beta_0 = 3.2172, \beta_1 = 0, \beta_2 = -0.0023, \beta_3 = -0.0001, \beta_4 = 0.0005, \beta_5 = 0.0001).$$

In Figure 8 it is shown the plot of the original data and the price obtained from the regression model.

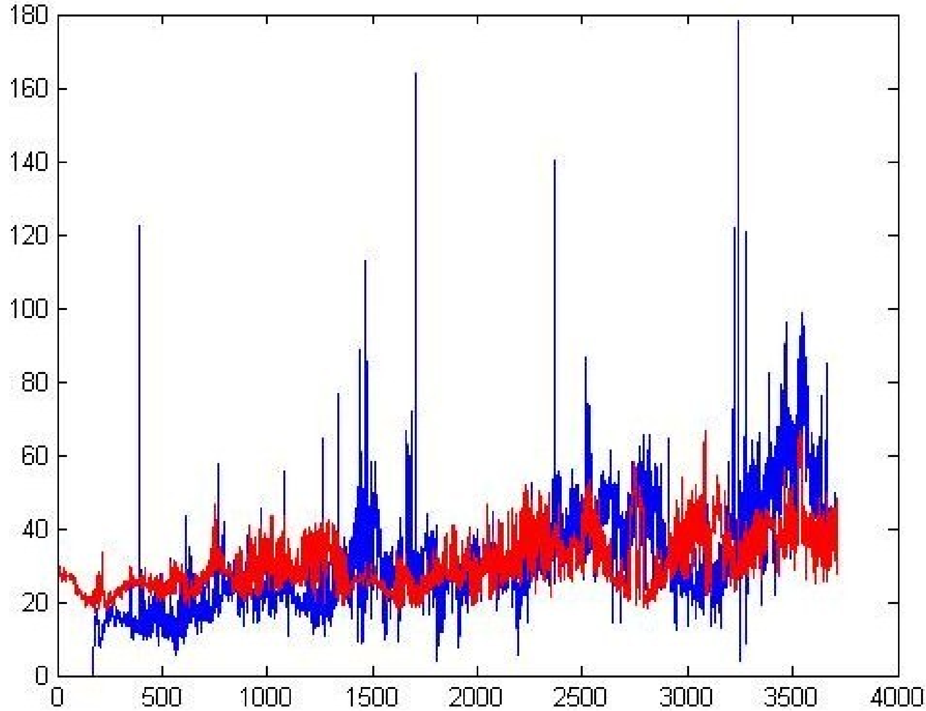

Figure 8: Plot of the original data and the price obtained from the regression model

4.2 How to improve the regression model

In the previous section, we showed how to apply a regression model to the electricity prices. As it can be seen in Figure 7 and 8, although the approximation generally tends to follow the same trend of the original prices, we were not completely satisfied with this model. Therefore, we searched for new ideas to improve the model.

For what concerns the Nord Pool, for instance, the data we are provided cover a range of ten years and we have one datum per day. Besides, it is evident that there are many sudden changes in the behaviour of the prices, which makes it hard to create a model valid for all the years based only on a few parameters. We then decided to split the data year-wise, i.e. resolving different regression equations for each year. The result for Stockholm prices is presented in Figure 9. One can easily see that for some years the approximation is very good, while for others unfortunately there still exist a gap between the real prices and the fitted model. This distance between the data and the approximation can be due to the fact that in that specific period the prices were influenced by other factors not available in the provided data set.

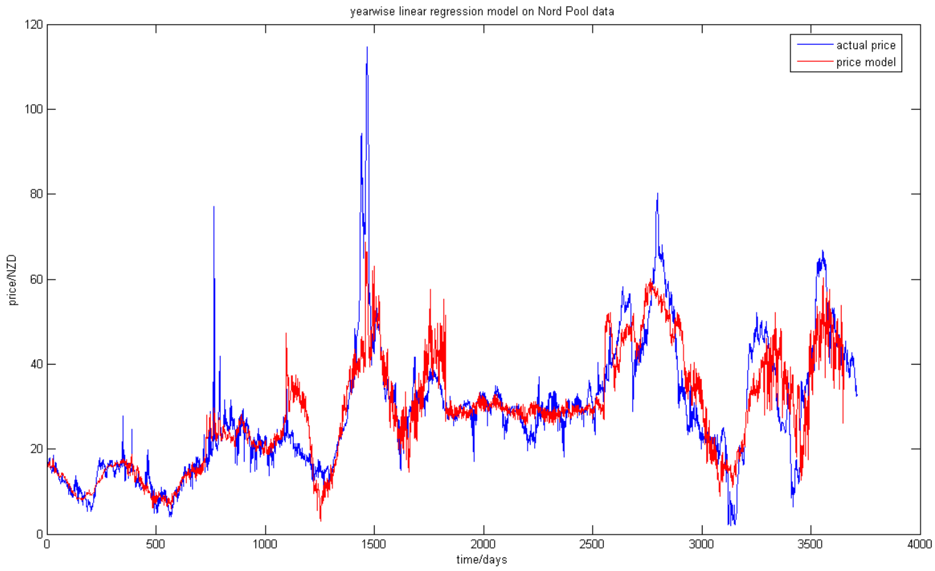

Figure 9: A simulation of the price with the regression model, fitting a different model for each year.

One completely different approach we considered deals with the occurrence of the spikes.

Observing Figure 7, we realised that the biggest distance between the model and the prices is in correspondence of the spikes. Especially, it looks as if the model "is not able" to reach high prices. We thought that such regression model is not suitable for approximating those high prices, but might improve in the other areas when the spikes data are not included in the model.

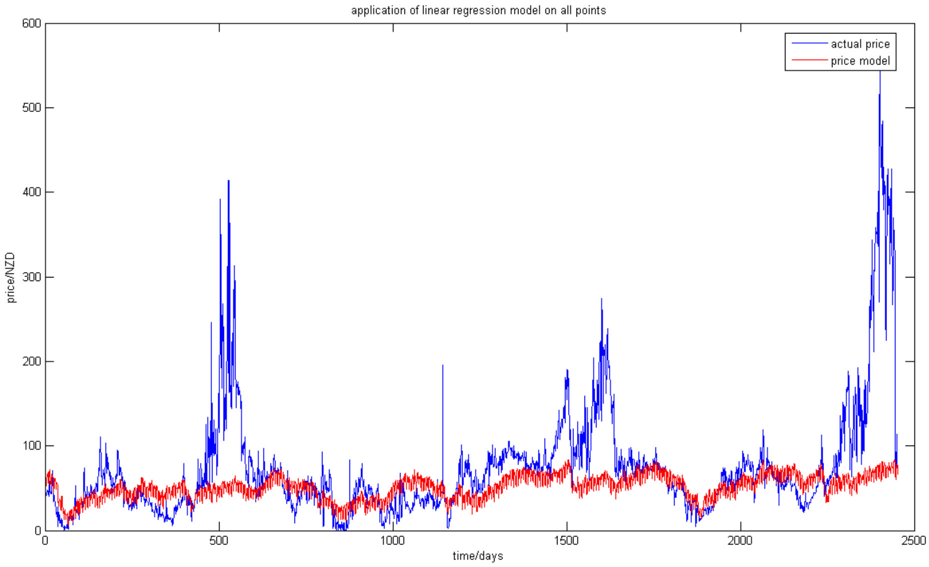

Figure 10: A simulation of the price with the regression model, fitting the model after removing the spikes.

Therefore, we decided to “remove” the spikes from our data before calculating the coefficient of the regression model, in order that these do not depend on the unexpected events of spikes occurrences. Once identified the spikes with the routine earlier introduced, we simply cut a neighborhood of the spike from the price vector and fit the model again. When representing the model prices, we used some noise in correspondence to the spikes areas. In Figure 10, it is shown the model obtained through this method concerning to New Zealand data. It can be seen, that the model is actually more accurate where there are no spikes occurrences.

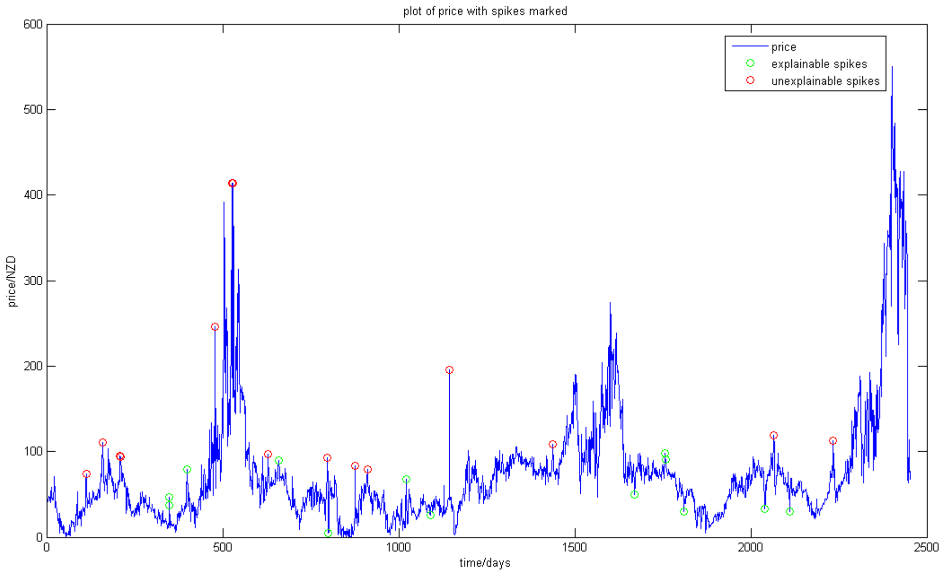

Figure 11: Prices with the spikes marked. The green circles represent the “explainable” spikes, while the red are the “unexplainable”.

We finally decided to take a look at the plot of the differences between the original prices and the model, in order to have a more precise idea of how good the model is. Moreover, we draw some circles in correspondence of the spikes. In this way, we could see how some spikes are well approximated by the model, when the difference is small, while some others occur exactly in the maxima of the difference. We could therefore draw a classification of the spikes, that we represented in Figure 11 through circles of different colors. Our conjecture about the meaning of these remarks, is that the “well approximated” spikes are the ones which can be explained with the data we are using to fit the model (green circles), while the others should depend on different factors (red circles). We discovered that this conjecture is consistent with the results of the analyses performed in Section 3.

4.3 Probability based model

Until now, we were not able to recreate with the model the high prices that often occur in our data. In this section, we present a model developed in order to reproduce that kind of spikes.

When comparing the New Zealand price series to the time series for the hydrological storage, we find that there is a relationship between the stability of the price and how big the hydrological storage is at a given time. When the storage is high the prices are stable and low, but when the storage decreases the prices grow higher and vary a lot more. Spikes in the price would only occur during these periods as well. This leads us to formulate a model that, instead of mapping the parameters directly to a price, would map the parameters to a probability distribution of the price. By providing stochastic variables with these distributions, a simulation of the price series can be achieved.

The model can be generalized as following: First the parameter space is divided into disjoint subsets. A probability distribution for the price is then associated with each subset. All the probability functions need to be determined appropriately. For every subset we create a histogram of all the corresponding prices of the all the data points in the set. These histograms are then used as the distributions.

First, we looked at how this model worked when the parameter space is represented only by the hydrological storage. The interval between the minimum and maximum storage in the data was divided into six equally large sub-intervals. The price histograms collected for each interval can be seen in Figure 12. A simulation of the price, using this implementation can be seen in Figure 13.

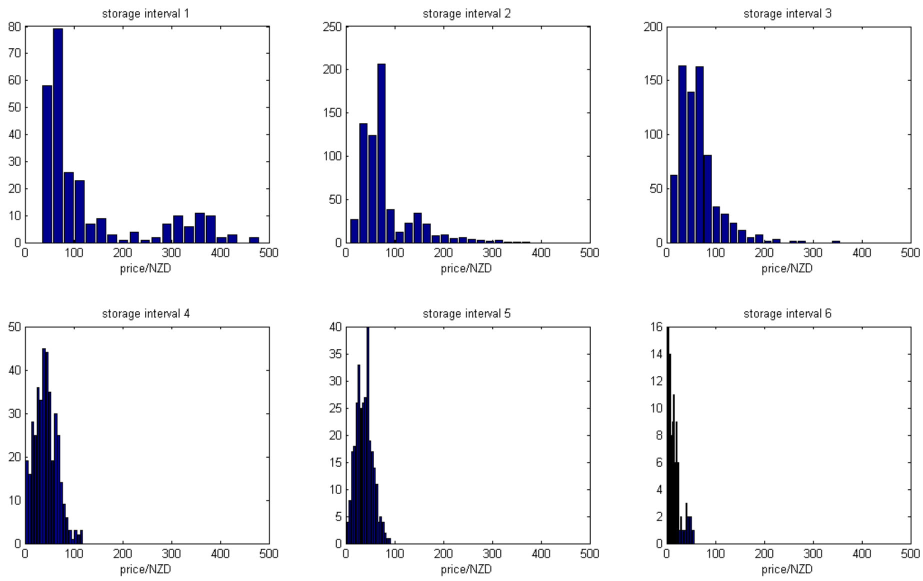

Figure 12: The histograms collected for six different intervals in the hydrological storage. The top left represent the prices when the storage was as at its lowest, and the bottom right when it was at its highest.

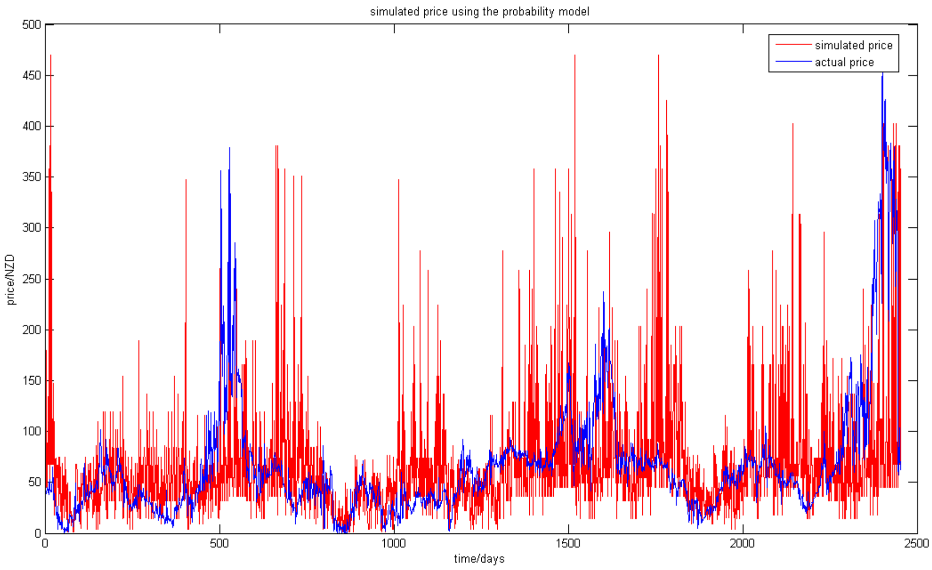

Figure 13: A simulation of the price with the probability model, using only the hydrological storage divided into six sub-intervals.

As we can see this model is rather crude as the residual is quite huge. Still, it manages to recreate some of the behaviour of the price data, like during the times the actual price is low and stable. It also manages to reach the high values that can be seen in the end of the price time series. We attempted to make the model a bit more advanced by using a two-dimensional parameter space consisting of both the hydrological storage and temperature data. Due to time constraints we did not manage to implement this properly. The results we achieved from our attempts must be considered invalid.

Despite the somewhat crude results our use of this model gave us, we still find it promising. We only managed to make it work on a one-dimensional parameter space, but we believe that if it can be implemented in higher dimensions better results can be achieved. The idea of running it on a parameter space of two dimensions, consisting of the hydrological storage and the previous day's price crossed our minds as something that should be studied further.

5 Conclusion

First of all, we worked on the definition of spikes in the spot market prices and worked out ways to classify and remove as many of them as possible. It is known that some spikes occur without any logical explanations, however through the available material, we managed to explain some of the spikes. Namely, the physical factors gave us explanation of some of the spikes, but the information from operational reports was not helpful to explain any of them. Our conclusion in this particular matter, is that more data are needed to allow complete analyses. Especially economical factors or information on electricity outflow from the specific areas might be of value for further investigation.

One final remark on the whole work on the spikes is necessary: although we could dispose only of a limited number of information influencing the electricity prices, it took us lot of time to go through it. Unfortunately, only a little of this kind of analysis could be performed via "systematic" methods, actually most of the work needed a "manual" approach. We believe that in the future it will possible to find faster way to handle this problem.

For what concerns the linear regression model showed, we are confident that there may be a potential in further research in this area. We gained some promising results, even though it required an intensive work before achieving good models. Further developments could be done using more complex models, e.g. nonlinear regression and/or including more parameters into the model. Moreover, it might be of value to analyze how the different parameters are correlated with some simple multivariate statistics.

Finally, the probability price model introduced in Section 4.3 showed some promising features. As before mentioned, we did not dispose of enough time to implement and test some new ideas we had to improve this model. We are sure that further analyses on these fields could improve the probability price model. Simply by increasing the number of parameters taken into account, one can obtain much better results. We sincerely hope that it will be possible to proceed with further investigations of probability based models and that this approach will produce a significant and helpful analysis of the spot market prices.

References

[1] George Rousssas, *A Course in Mathematical Statistics*, Academic Press Inc, 1997