FORECASTING WITH TAUT STRINGS

Bas Meeuwissen
October 09

Sandjai Bhulai
VU university Amsterdam
Faculty of Science
De Boelelaan 1081a
1081 HV Amsterdam

1 PREFACE

This paper is part of the mandatory curriculum of Business Mathematics and Informatics.

During a project at the university I came across the subject of forecasting with taut strings. In this project we had to forecast the number of calls arrived in a call centre via the smoothing extensions of the taut strings. Since we did not do a very good comparison with other methods, I wondered if the taut string algorithm would be a good algorithm to use in a call centre. This was exactly what my supervisor, Sandjai Bhulai, also wanted to know and the idea of this paper was born.

I want to thank Sandjai for his support and comments during the writing of this paper.

Bas Meeuwissen
Amsterdam 2009

2 SUMMARY

In this paper the possibility to forecast incoming calls in a call centre with extensions of the taut string algorithm has been researched. A forecast is for example the average number of calls received during one interval. To see if the algorithms can be used, a comparison has been made with a number of common forecasting techniques. The forecasting techniques are;

- Mean per interval
- Median per interval
- Moving average per day
- Exponential smoothing per day
- ARIMA model per day

The extensions of the taut string algorithms do not work that good on forecasting the number of incoming calls. 12 days were forecasted, but only 3 times the taut string algorithm was the best to choose. A remark here is that none of the other forecasting techniques outperformed the rest.

3 TABLE OF CONTENTS

4 INTRODUCTION

This paper is about forecasting with Taut Strings. The technique is newly developed to describe a dataset with better accuracy than the current techniques. The field of the taut strings is being researched by A. Kovac, L. Dümbgen, P.L. Davies, E. Mammen, S. van de Geer and others. In this paper, the taut string algorithm will be used to forecast and compare the results with the results of more common techniques.

Forecasting is a mathematical technique to say something about the future while certain parameters are unknown. For example, what can be said about the number of calls a call centre will receive next Monday, next month or maybe next year. Throughout the paper, the call centre will be used as an illustrative example. Thereby, the math on call centres is highly developed and a lot is known about the patterns, waiting and service times.

4.1 BACKGROUND

In the current economy, the forecast has become a highly important tool to run a business. A bad forecast means a head start for the competition and less income, a smaller market share for your business or a worse service to your customers. These effects can occur because, for example, the cost price of a product is actually less then what is forecasted. This can happen if the forecast of the sales are too low and all the overhead costs and material costs are based on the sales forecast. Using this for the product cost will make you charge the customer to much(1). Businesses with a bad forecast model even went bankrupt!(2)

Since all private businesses need income to survive, they aim at the biggest net profit they can get. Therefore a lot of forecasts are made to evaluate the cost or profits increase or decrease of certain future decisions made by the management. Other topics which are forecasted regularly are the market share, the customer satisfaction and goods consumption.

Business can be divided into segments. Every segment has its own specific problems, which is the result of specific characteristics. I divide the business into two segments; those with an inventory (the classical businesses, like supermarkets), those without an inventory (the service businesses, like call centres).

Businesses with an inventory are likely interested in how much they sell this summer, so they can fill their stock to the right level. An incorrect inventory costs too much money, since an inventory level which is too low wastes income, but an inventory which is too high produces extra costs, but no extra income.

Service oriented business, like ENECO and NUON, want to know how much power the consumers will consume on a certain day. Electricity cannot efficiently be stored in an inventory at the moment(3), so all the electricity which will not be used is wasted. If too few electricity is produced, the customer satisfaction will drop. These companies have the difficulty to make a trade-off between service level and costs.

For all these problems forecasting methods exist. Every method has its strengths and weaknesses for a certain problem. For most problems, it is important to find the best method, since methods perform weakly on problems they are not designed for. The preconditions per problem are important and not to be violated to get a good result. However, it is not the intention of this paper to find the right forecasting methods to the right problems, but to compare the different forecasting techniques with the Taut Strings forecasting method on one specific problem, forecasting for a call centre.

4.2 ISSUE AT HAND

For any given call centre, there are a two characteristics(4) which I will explain first. The characteristics are arrivals and service level and the relation between those two characteristics.

The arrivals in a call centre are assumed to be Poisson distributed with an arrival rate λ. This arrival rate depends on time, size of the call centre, the weather outside, business rules and lots of other (unknown) factors. Unexpected changes, like the weather outside or dramatic events, cannot be forecasted and always result in a unreliable forecast.
The arrival rate is also dependents on itself. The morning is often a good indication for the rest of the day. A busy morning, thus a high arrival rate in the beginning of the day, often results in busier than usual day.

In the figure below it can be seen that the number of arrivals vary during the day. For this call centre the busiest moments are between 8 A.M. - 11 A.M. and 1:30 P.M. – 5 P.M. There are also differences between a Tuesday and a Sunday. The Tuesday is busier than the Sunday. It is also clear that the beginning of the week is quiet and it gets busier at the end of the week.

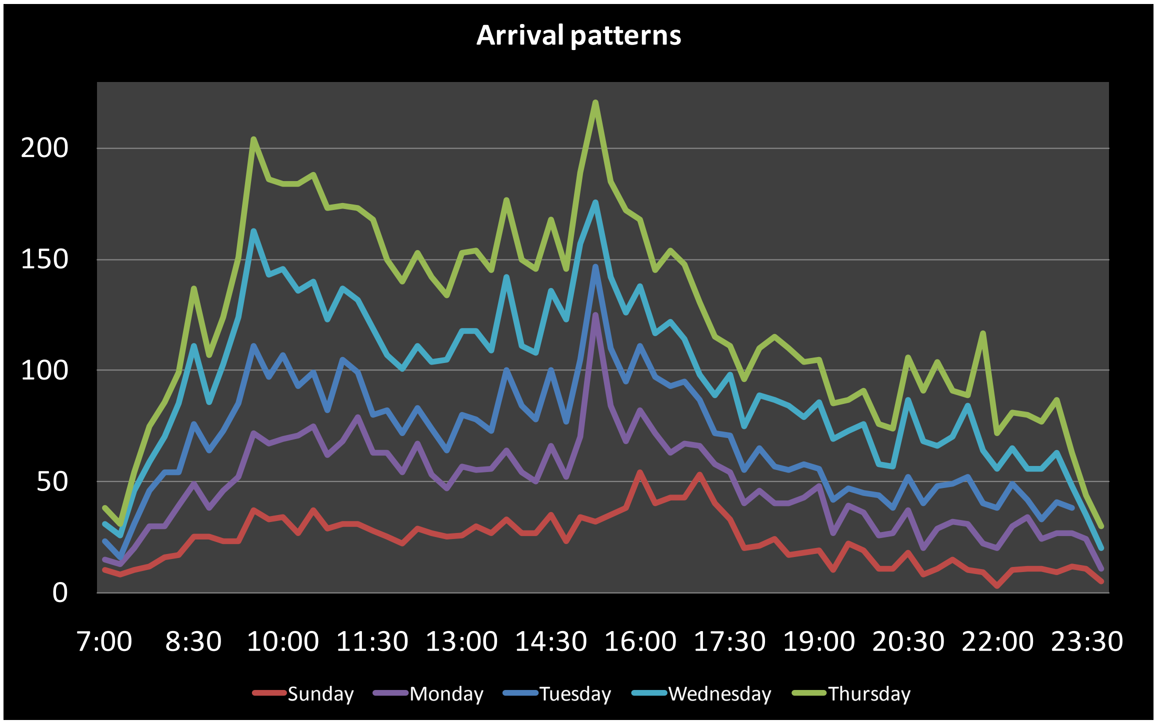

FIGURE 1: ARRIVALS ON A TUESDAY, A THURSDAY AND A SUNDAY

In the figure above it is not clearly visible that for example Sunday has the same characteristics as Thursday. This is because the axis has to fit all data, but if the axis is zoomed in on Sunday, the same pattern as Thursday is visible (Figure 2).

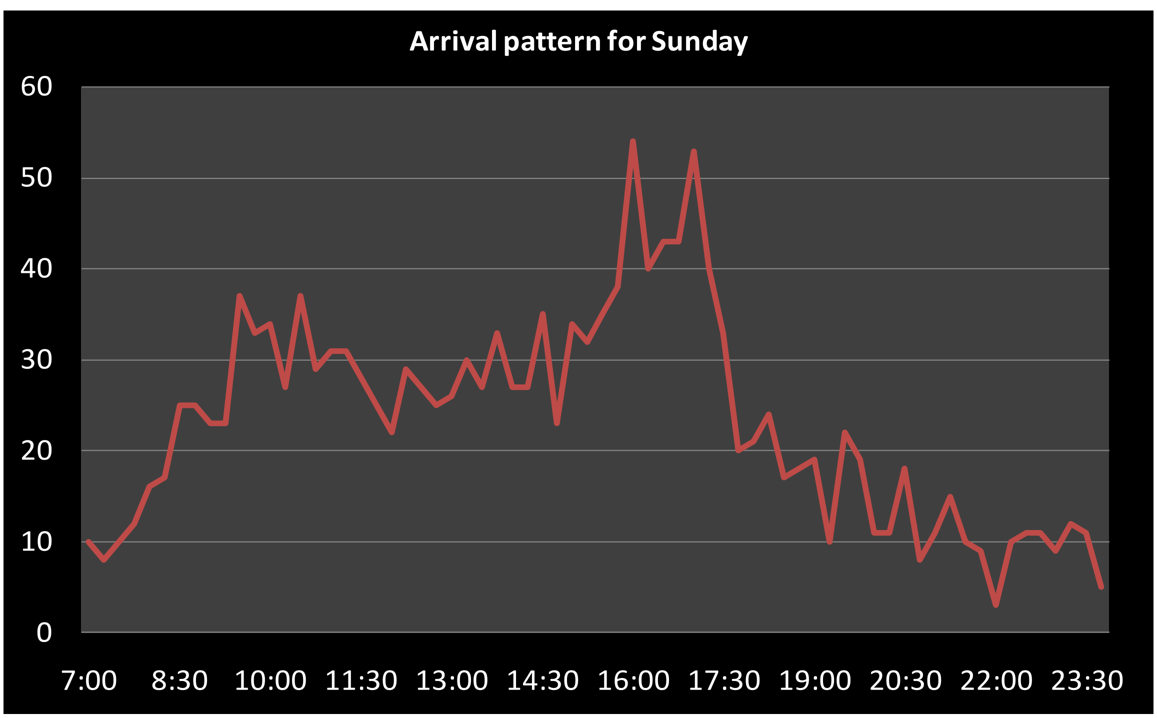

FIGURE 2: ARRIVALS FOR A SUNDAY

Looking at the arrival rate, a uniform forecast for all days is not desirable. The forecasts have to be specific for the days they are made for. The techniques have to take these patterns into account for a proper forecast.

Another property about the arrivals is that they will shift during the day or even during the week. People who call a call centre and have to wait too long before getting served hang up and probably will call back some other time. These calls will shift for example 30 minutes. This can happen if the call centre does not employ enough personnel or if an exceptional high number of people call at once. Not employing enough personnel disturbs the data and will make it difficult to give a good forecast.

A change in business rules can have a changing effect on the arrivals as well. If management should decide to close the call centre on Sunday mornings, then the arrivals will change. If this decision is made the last month and the data is one year old, then the forecasts will be useless. Most of the data is not based on the current business rules, thus not applicable for forecasting.

Another characteristic of call centres is service level. The service level of a call centre consists of different measurements. A few of these measurements are average speed of answer, the number of calls answered within a certain time span and the number of abandonments. If the number of arrivals are known then the number of agents can be calculated to achieve a certain service level. A specific number of agents can handle a limited range of arrivals. The more arrivals agents have to handle, the less good the service level will be. Up to the point were just a fraction of all arrivals will be served.

Taking the two characteristics into account the problem with forecasts is that they have to be detailed and precise. It is not very useful to have a forecast of the number of calls for one whole day, since the arrivals vary per interval. And the forecasts have to be precise, because a difference in the number of calls will result in a difference of employees needed, and thus on costs and service level.

4.3 PROBLEM

For an imaginary call centre we are asked what the best method is to forecast their incoming calls. They have heard about quite a few forecasting methods and want to know which works best for their situation. The manager has read something about taut strings and is very interested in this method. At the same time he wants a comparison with the more common techniques.

4.4 APPROACH

I will discuss the most common forecasting techniques and their performance on a real world dataset. The dataset is a large matrix of numbers with a new day at a new row. For every day (261 from one year), the incoming calls are registered per 15 minutes interval and put in columns (68 in total), thus the dataset looks like the following table.

		7:00	7:15	7:30	7:45	8:00
3-1-1999	Sunday	11	25	21	14	22
10-1-1999	Sunday	7	9	8	12	12
17-1-1999	Sunday	10	8	10	12	16
24-1-1999	Sunday	5	5	5	14	7

TABLE 1: A SMALL PART OF THE DATASET

Since the taut string is the main interest of this paper, this technique will have its own chapter. Within the chapter some more detailed information and explanations will be given.

After these theoretical parts, the techniques will be used on the dataset to create forecasts. These experiments are conducted in the following way.

At first a few known days are taken and removed from the dataset. From these days a few intervals are taken. These intervals always start at 7 o'clock in the morning and end somewhere between 8.30 A.M. and 11.30 A.M. The intervals serve as input to the Pearson Correlation Distance (PCD). The PCD is a method to calculate the likeliness of two matrices of numbers with the same number of elements. I will use it to calculate the likeliness of two strings with the same number of numbers. The result of this calculation is a dimensionless value.

The PCD needs two strings of numbers to calculate a value, and that is what the dataset is for. If we know for example the first 10 intervals of a day we need another 10 intervals for the PCD to calculate a value. These 10 unknown intervals are extracted from the dataset. For every day, the first 10 intervals are taken and put in the PCD together with the 10 intervals we know. This gives the Pearson correlation distance between the days in the dataset and the day we want to forecast.

After calculating the PCD for all days, the 40 days with the highest PCD will be used for the forecasts. The reason these 40 days will be used is that of the arrivals rate. We know that the characteristics for arrival rate is more or less the same for every day. We also know that the number of calls in the morning is a good indication about the number of calls in the afternoon. Thus it is only necessary to find the days where, for the known intervals, the likeliness is the best.

The forecasts will be created according to the descriptions of the methods in chapters 0 and 6. An extra operation is needed to adjust the forecasts to usable numbers. The forecasts will give real valued numbers, but the number of arrivals are integers. Therefore the forecasts will be rounded off. This will have an effect on the Pearson correlation distance, but I have never heard of 9,3 arriving calls in 15 minutes.

The forecasts have to be measured again. Therefore the PCD is used again, but now on the forecast and rest of the known day. If we knew 10 intervals, there were 58 intervals not used the first time the PCD was calculated. These 58 intervals had to be forecasted, which is done during the experiments. The forecasted intervals, together with the unused intervals, will be put in de PCD. This gives the measurement used to compare the forecasts with each other.

5 FORECASTING TECHNIQUES

There exist a lot of different forecasting techniques. They are based on different mathematical calculations, called statistics, like mean and median. Statistics can also be more complex. The exponential smoothing uses weights for the past statistics, which are calculations by themselves, and weights for the current data.

A forecasting technique can be simple, like the average number of calls received on the past five weeks, which will be the estimate of the next intervals. More complex techniques like the ARIMA exist. This technique is constructed of multiple models, where statistics are dependent and error estimates are taken into account.

At first, the simple techniques will be discussed to get to the more complex techniques and models at the end of this chapter.

To explain the forecasting techniques, the usage of these techniques on the dataset will be explained. The dataset used is the dataset created by the 40 rows which have the highest PCD. This means that the procedures written in the next sections is also the approach for the forecasts. One main concept about the forecasts is that the data has to be seen as a 2 dimensional dataset, an matrix. Some techniques use one dimensional data in their forecast (either along the intervals or the days) and other techniques use both dimensions. The techniques that use one dimensional data for their forecast are adjusted to work with the 2 dimensional dataset.

5.1 AVERAGE(5)

The average is the expected value of a dataset. The most commonly used average is the arithmetic mean, or just called mean. Other examples of an average are the trimmed mean and weighted mean. Many more averages exist, but they are not used frequently.

A definition of average is : *the value obtained by dividing the sum of a set of quantities by the number of quantities in the set*[1]. If the number of quantities is big and there is just one big number in the set of quantities, then the contribution of the large number is not noticed very well.

5.1.1 WEIGHTED MEAN

The weighted mean is a method to calculate an average. It is possible to give weights to data points. This is usually done for time dependent data. Although the data is time dependent, this property is not used for the forecasts.

The formula for the weighted mean is

[1] www.thefreedictonairy.com

$$s = \frac{\sum_{i=1}^{n} \alpha_i x_i}{\sum_{i=1}^{n} \alpha_i}$$

Where n is the number of observations, α_i is the weight which is applied to an observation x_i. Since all observations have the same weight, the formula becomes

$$s = \frac{\sum_{i=1}^{n} x_i}{n}$$

which is called the arithmetic mean.

In the forecasts, the arithmetic mean is used in one dimension. The dimension is the interval. Thus for a forecast for an interval, all observations for the same interval for all 40 days are used.

The graph below shows an example of the above text.

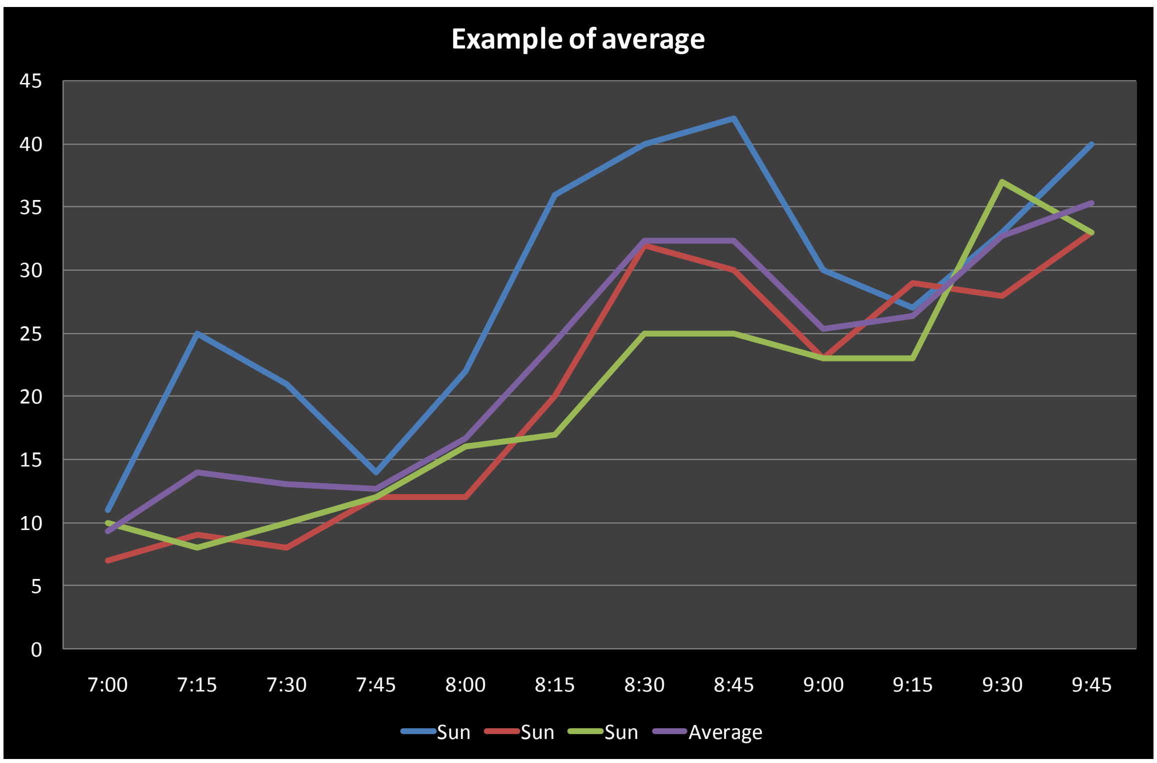

FIGURE 3: EXAMPLE OF THE AVERAGE USED ON THE DATASET

The purple line shows the average of the 3 other lines, where the average is calculated per interval.

5.1.2 TRIMMED MEAN

The trimmed mean calculates the mean by excluding a certain number of numbers. First, the number of calls received will be sorted and then the top and bottom $x\%$ will be excluded for the calculation.

The formula of the trimmed mean is

$$s = \frac{\sum_{i=j}^{n-j} \alpha_i x_i}{\sum_{i=j}^{n-j} \alpha_i} \text{ where } x_1 < x_2 < \cdots < x_n \text{ and } j < n$$

j is the number of observations which will be left out on both sides, and α_i is a weight which can be applied.

The 50% trimmed mean is a regularly used statistic. It even has its own name, the median. The median is insensitive to outliers at all. It just looks at the single value that is halfway the dataset. For an uneven number of rows, the median is the number in the middle. For an even number of rows there is a problem. Let us take a look at a small example. If we have four numbers, say 3, 5, 6 and 8, the number in the middle would lay between 5 and 6. But that number doesn't exist, thus we take the average of those two numbers.

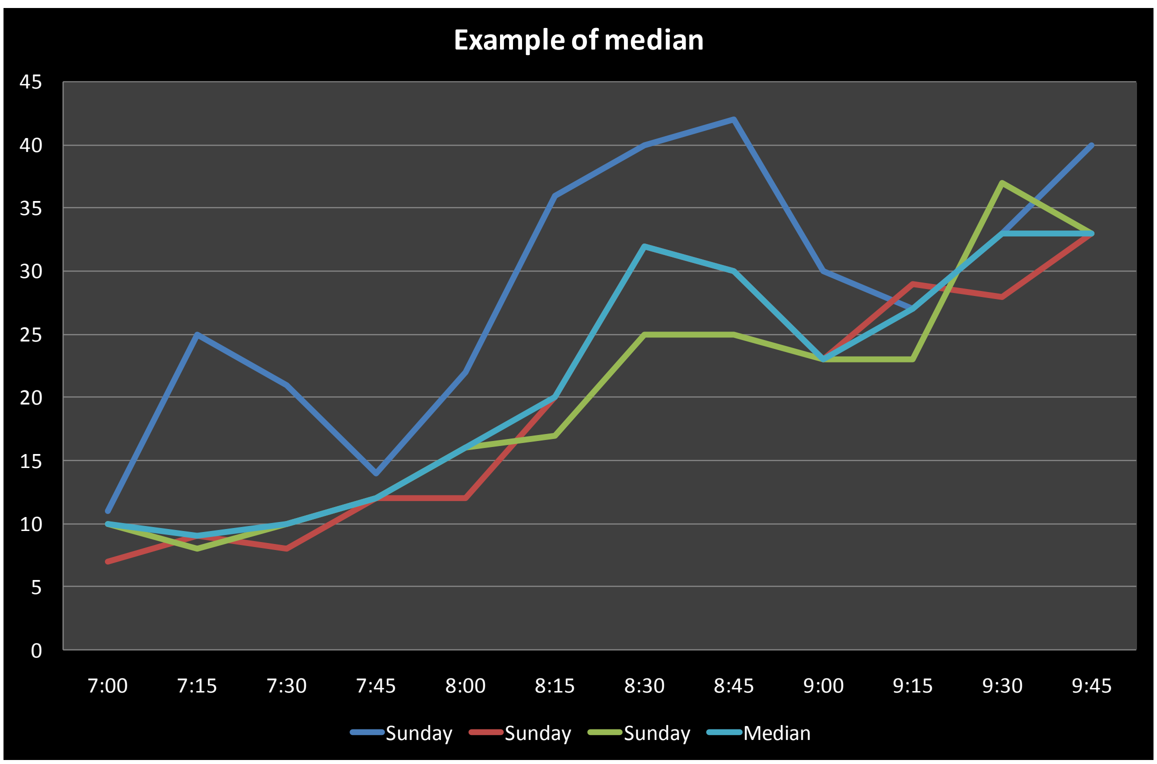

FIGURE 4: EXAMPLE OF THE MEDIAN USED ON THE DATASET

The median works on the same dimension as the average. Thus the average is calculated every interval. It is clear from the graph that the median just takes the number in the middle, since it overlaps the lines which are in the middle. It also can be seen that the median is calculated per interval, otherwise the median would have followed the green line.

The advantage of a trimmed mean is that is less sensitive to outliers the more data is excluded.

There is a disadvantage using the trimmed mean. If the number of calls in a certain interval have a steep rise of sharp fall, then the median probably would not notice it. It would only notice it if more than half the selected days have the same pattern on that interval.

5.2 SIMPLE MOVING AVERAGE(6)

The moving average is a method to use one of the above statistics to create a continuously changing prediction in a time series. The moving average is not a single value to describe a complete dataset, but it takes a few points to calculate for example the weighted mean or trimmed mean to calculate the current data point. The total number of points taken is called the step. A two step moving average uses two data points to calculate the average.

The formula for the moving average at time t is

$$s_t = \frac{1}{j} \sum_{i=t-j+1}^{t} \alpha_i x_i$$

j is the step, the number of data points taken into account. The t is the time of the current statistic, for example interval 10. α_i is the weight given to the number of calls x_i.

The formula itself says that the calls for this interval are calculated as the calls of $j-1$ intervals before and this interval itself.

There are two possibilities to calculate the moving average on the selected days. The first possibility is to calculate the average per interval and then use this calculated number of calls with the moving average. This would be no more than calculating the moving average of the prediction made in section 5.1.1. The second possibility is to calculate the moving average of all 40 selected days and take the average over these calculated days.

It is the second possibility which will be used in this paper to create a forecast. I will explain this method with a little example. In the table below I have taken two days which are not very much alike. I have done this, because the graph which comes with this example would be unreadable if the days would be very much alike.

The moving average is a 2 step moving average and both data points are weighted equally. So, to calculate the new [7:15 7:30] interval using the Sunday, the calculated interval becomes $\frac{1}{2}(11+25) = 18$. This is also the value on the third row. The same trick is used to calculate the rest of the row and to calculate the moving average (MA) of Thursday. The average per interval of those two rows (MA Sunday and MA Thursday) is the value of the final moving average. In the table the round off can be seen as well. The average of 18 and 5.5 is 11.75, which is rounded off to 12.

	7:00	7:15	7:30	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45
Sunday	11	25	21	14	22	36	40	42	30	27	33	40
Thursday	6	5	5	13	17	25	49	31	31	15	34	24
MA Sunday		18	23	17,5	18	29	38	41	36	28,5	30	36,5
MA Thursday		5,5	5	9	15	21	37	40	31	23	24,5	29
Final MA		12	14	13	17	25	38	41	34	26	27	33

TABLE 2: THE MOVING AVERAGE

The figure below is a graphical representation of the table above. The moving averages for the days are the dotted lines and the final moving average is the blocked line. In the figure it is clear that the moving average is always at least one step behind the changes in the data.

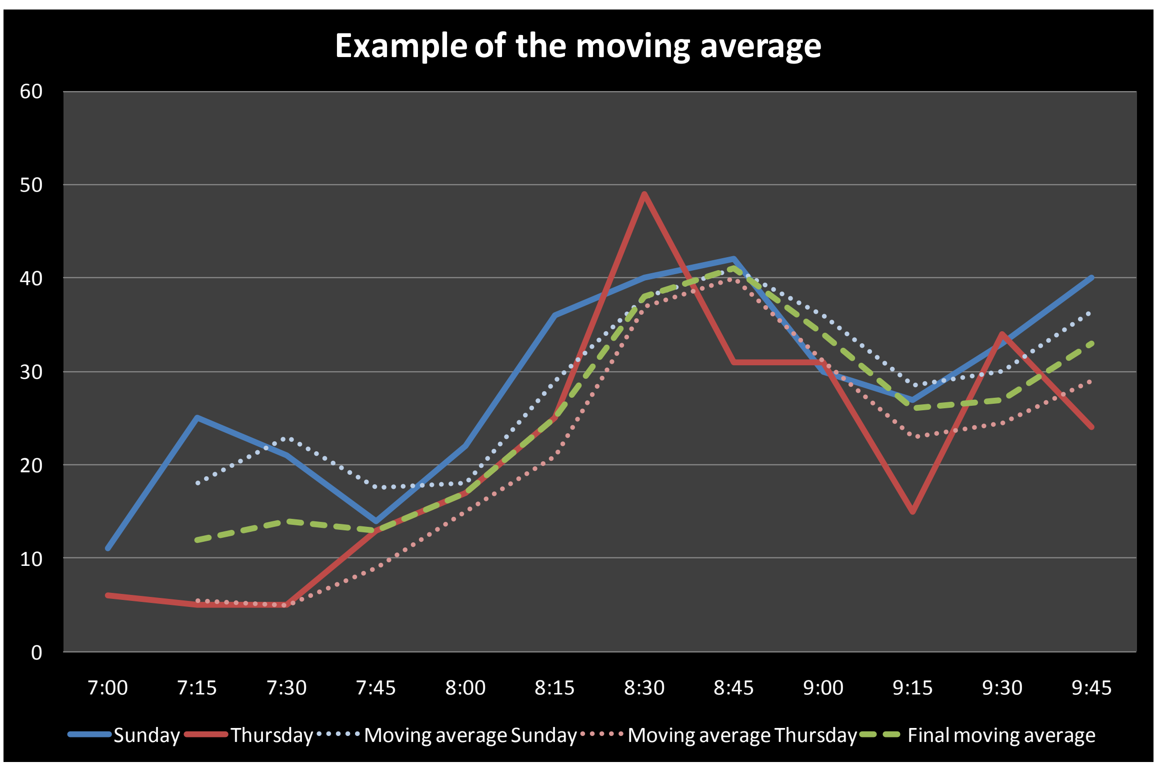

FIGURE 5: GRAPHICAL REPRESENTATION OF TABLE 2

Because the moving average needs two data points, the first interval does not have a value. This is solved for the forecasting problem by beginning the moving average one step before the actual forecast is needed. The number of intervals known is always more than one, thus this trick can be applied on every forecast.

5.3 EXPONENTIAL SMOOTHING(7)

Exponential Smoothing is a special case of the moving average. The exponential smoothing does not need a few data points to calculate the next value. Instead, it calculates the next value by applying a smoothing factor to the old data and the new data.

The formula for the exponential smoothing at time t is

$$s_0 = x_0$$
$$s_t = (1-\alpha)x_t + \alpha s_{t-1}$$

Where α is the smoothing factor and $0 < \alpha < 1$. The value s_t is a simple weighted average of the last observation x_t and the previous smoothed statistic s_{t-1}.

The value of α has an impact on the weights for the current statistic. An α close to zero gives more weight to the current observation which means that the statistic will be less smoothed, but is more responsive to changes in the observations. An α close to one gives more weight to the smoothed statistic s_{t-1} which is less responsive to changes in the observations, but has a more smoothing effect.

The exponential smoothing can also be used in two ways. Here, I have chosen to use it the same way as the moving average. First, the exponential smoothing is calculated for every selected day and then the average is taken for the calculated intervals. Since there is no optimal value for the smoothing factor α, its value will be 0,5. I have chosen for this value, because it is in the middle of all possible value.

I will explain the exponential smoothing with a little example.

	7:00	7:15	7:30	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45
Sunday	11	25	21	14	22	36	40	42	30	27	33	40
Thursday	6	5	5	13	17	25	49	31	31	15	34	24
ES Sunday	11,0	18,0	19,5	16,8	19,4	27,7	33,8	37,9	34,0	30,5	31,7	35,9
ES Thursday	6,0	5,5	5,3	9,1	13,1	19,0	34,0	32,5	31,8	23,4	28,7	26,3
Final ES	9	12	12	13	16	23	34	35	33	27	30	31

TABLE 3: EXPONENTIAL SMOOTHING EXAMPLE

The data for the Sunday and Thursday is the same as the data for the moving average example. It is clear that the exponential smoothing does not have the problem of needing an extra data point to create a forecast. Instead, the original value is used as the first forecasted value for the first interval. After that, the average of the forecasted value of the first interval and a new data point from the second interval is a new forecasted value for the second interval. Thus to forecast the [7:15 7:30] interval, the ES Sunday [7:00 7:15] value and the Sunday [7:15 7:30] value are averaged: $\frac{1}{2}(11+25)=18$. This trick continues until the end.

The 2 forecasts from the exponential smoothing are averaged per interval to create the final forecast. The figure below (next page) contains the example, but then visually.

For the exponential smoothing I thought of two options how to use this method. The first method is to begin 1 interval earlier than the forecast is actually needed. With this, the forecast would not contain any original numbers from the 40 selected days. The other option was to use the exponential smoothing as it is described in the formula at the beginning of this section.
I have chosen to use the exponential smoothing the way it is described in this section. The method has coped with the problem to forecast the first value and I wanted to leave the method intact.

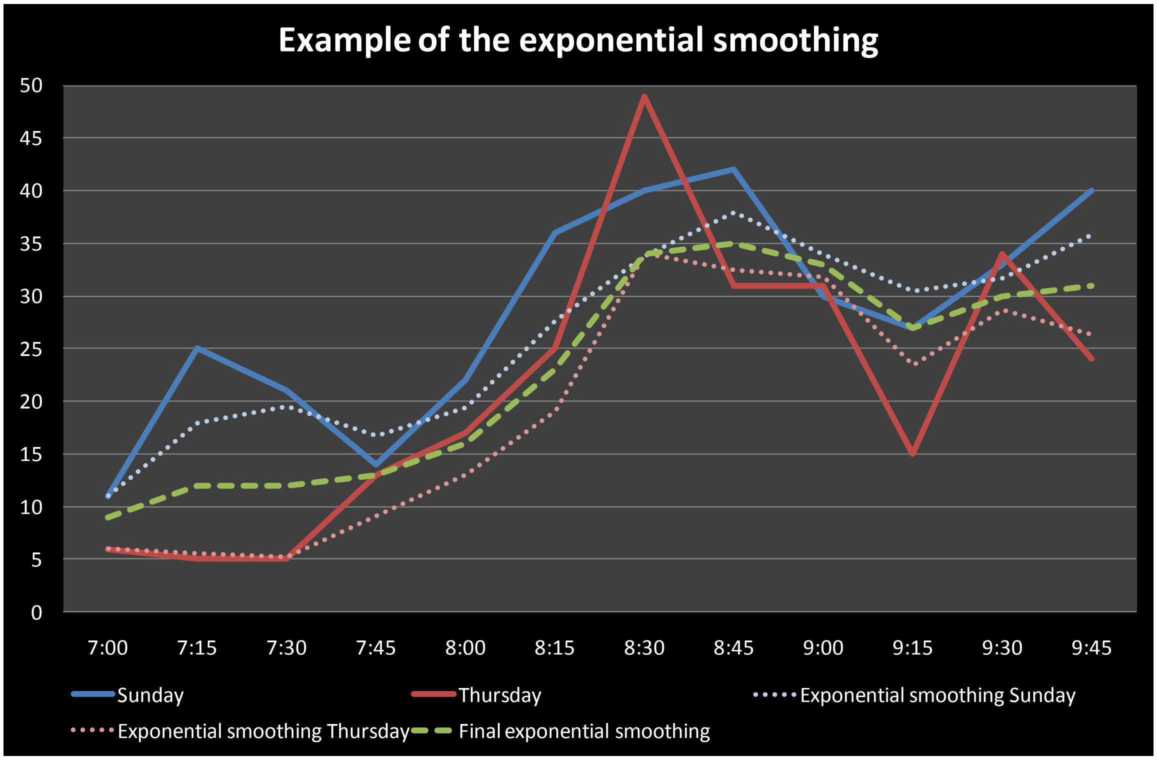

FIGURE 6: THE EXPONENTIAL SMOOTHING

5.4 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE(8)

The ARIMA model is the integrated variant of the autoregressive moving average (ARMA). Therefore I will explain the ARMA first, and later on the ARIMA. The ARMA model itself is a compounded model, which exists of an autoregressive process (AR) and a moving average process (MA). So the explanation is divided into four parts. First, the concept of an AR process is explained. After that the MA will be explained. This moving average is different than the moving average explained in 5.2. The whole ARMA comes as third. Thereafter the explanation of the ARIMA will follow.

The ARIMA process is a time series model. This means that the model is used to predict or describe an event which is dependent on time. For example the call centre, the number of calls at 11 o'clock depends on the number of calls received at 10 o'clock. So this model can be used with the call centre data without manipulation.

5.4.1 AUTOREGRESSIVE

The autoregressive (AR) part of the ARIMA is a process which looks at **p** previous events and gives them a certain weight. To calculate the current statistic, the **p** weighted events get an extra component of white noise. White noise is a number which is drawn from the normal distribution with mean 0 en standard deviation 1. The formula for the AR(p) process is:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

The formula means that the current observation is dependent on the previous observations with a parameter φ per observation, plus an undetectable error, which is assumed to be normally distributed with mean zero. This is the white noise.

5.4.2 MOVING AVERAGE

The moving average (MA) part of the ARIMA is a process which takes q white noise data points and gives them a certain weight. To calculate the current statistic, the q weighted events are added. The formula for the AM(q) process is:

$$X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

5.4.3 AUTOREGRESSIVE MOVING AVERAGE

The ARMA process is the combination of an AR(p) and an AM(q) process. The two processes are added, but the parameters per process are independent of each other. The formula for the ARMA process is:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

The formula above can be expressed in the Lag operator. This is a function which operates on an element of a time series. For example;

$$LX_t = X_{t-1}$$

With the Lag operator, the previous formula for the ARMA becomes;

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

This is an equation which can be directly used for the ARIMA.

5.4.4 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

The ARIMA process is the ARMA process with an integrated MA process. The integration has an advantage. The ARMA process needs a stationary dataset to work on. This means that the dataset itself is not allowed to have seasonality or trend, and is supposed to follow a random walk. The ARIMA process does not have that precondition on the dataset. Even if the dataset has a trend or seasonality, the ARIMA process can handle it. The formula for the ARIMA process is:

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1-L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t$$

For the ARIMA process, there are also two possible ways to use the method. The ARIMA can work on the whole dataset at once (a multivariate ARIMA) or it can work on a whole day. It would have been nice to use the multivariate ARIMA, but the program I used did not support multivariate ARIMA (xlstat). Therefore, the ARIMA is calculated per day and the average is taken for the predictions given by the multiple ARIMAS.

5.5 COMPARISON

In this section the strengths and the weaknesses of the models above will be listed in a table.

Model	Strengths	Weaknesses
Weighted mean	Easy to compute	Sensitive to outliers Needs a tweak to work with time series
Trimmed mean	Easy to compute	Does not respond very well to changes in the data Needs a tweak to work with time series
Simple moving average	Easy to compute Does work with time series without tweaking	Is always a few steps behind the changing data
Exponential smoothing	Does work with time series without tweaking	Not a uniform answer for alpha.
ARIMA	Does work with time series without tweaking	Complex. Sometimes needs extra preparation of the data No uniform answer due to a lot of input parameters

6 TAUT STRINGS(9)

The concept of taut strings is relatively new. Half way the previous century, the concept was developed. The idea of a taut string is to create some kind of function, which is able to describe a dataset. The exact formula is usually unknown, but some properties of the dataset are needed. This is the field of the regressions. Regression methods are iterative processes where the best fit of a curve or line is computed. The taut string algorithm is an iterative process where small continuous functions are created to come up with one long string of numbers.

The taut string method is an integrated process with linear interpolation between design points. These design points are created as follows;

$$Y_0 = 0$$

$$Y_j = \sum_{i=1}^{j} y_j \quad (j = 1, \dots, n)$$

These design points create the function Y. The function Y is the basis for a tube, which is created by adding and subtracting a number $\lambda \Rightarrow [Y + \lambda, Y - \lambda]$. This tube is the basis for the taut string. The left panel of Figure 7 shows the function Y (the line in the middle) and the tube, created by adding and subtracting λ. This middle string is tightened, which can be visualised as two hands pulling the string to the sides. Because the tube does not change, the string is pulled along the different sides of the tube, creating a *taut string*. The right panel of Figure 7 is the result. This gives the function F^{λ}. Differentiating the function F^{λ} yields the approximation f^{λ}. This is the function we were looking for.

A few properties of the taut strings are;

- Piecewise constant
- Modality increases monotonically with decreasing tube width, dus het aantal pieken of relatieve maxima neemt toe als de tube kleiner wordt.

In the next part, I will explain the properties step by step.

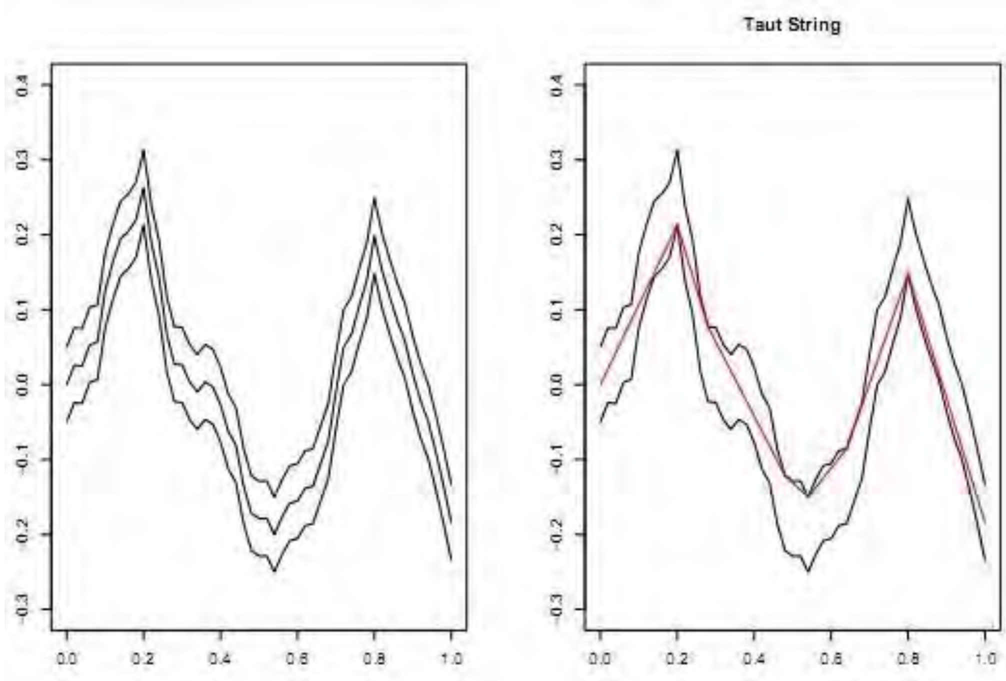

FIGURE 7: DESIGN POINTS WITH λ ADDED AND SUBTRACTED AND THE TAUT STRING

The taut string is piecewise constant. This comes from the fact that the string inside the tube makes a jump at a local extreme. Between local extremes the value of the function is the average of two local extremes (see Figure 8).

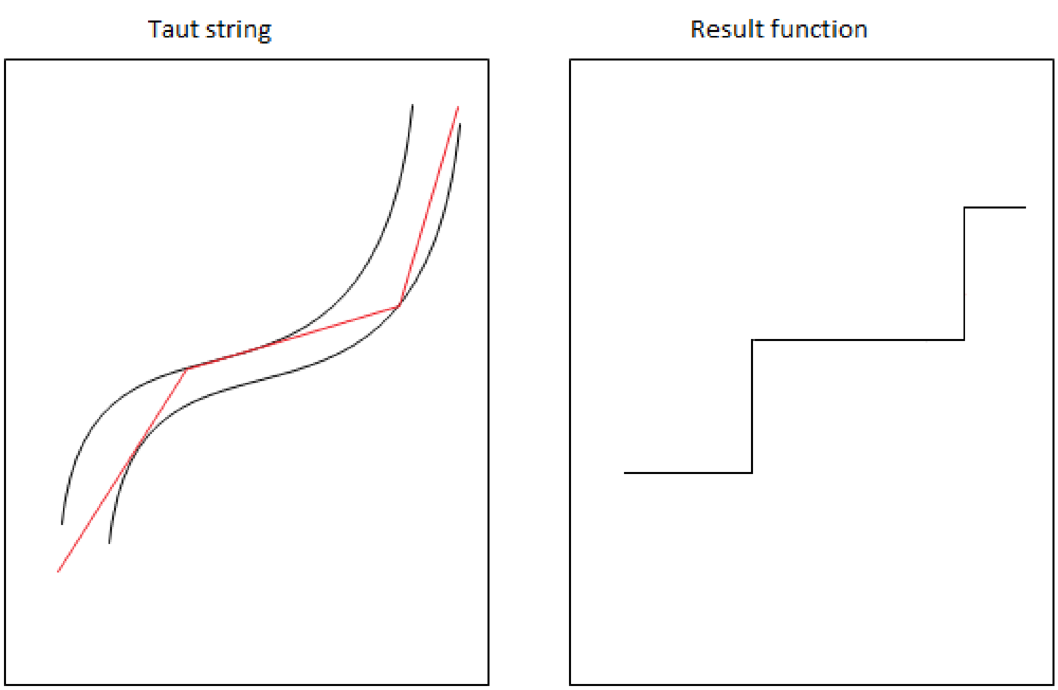

FIGURE 8: THE TAUT STRING WITH THE RESULTING FUNCTION

The modality, i.e. the number of local maxima or minima, increases monotonically with decreasing tube width. This is true because the string within the tube bounces between the upper and lower bound. If a local maximum is passed, the string will lay on the lower bound. For a local minimum it is the counter case, the string will touch the upper bound. If the tube width is large, the string will have a lot of space to go from one point to another without touching the bounds. By decreasing the tube width, the string will have less space to go from one point to another and will encounter a bound more frequently. In the figure below the right panel shows a large tube and there are no local extremes. The line can pass through the tube without touches the boundaries. In the left panel, the tube width is smaller and the line touches the boundaries.

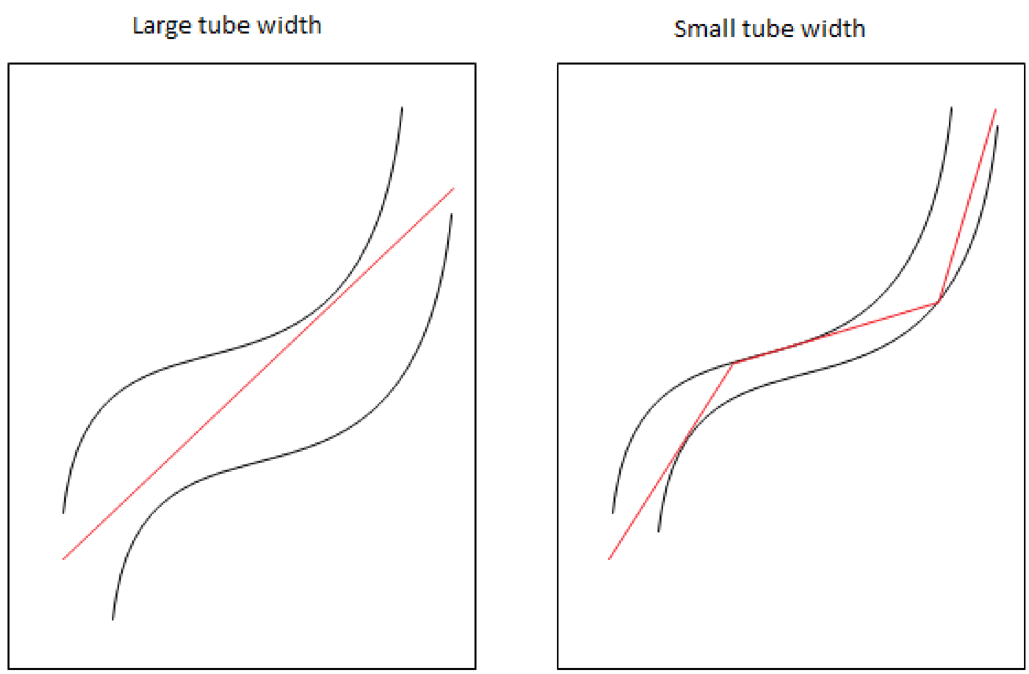

FIGURE 9: MODALITY WITH LARGE AND SMALL TUBE WIDTH

A smaller tube width, creating more local extremes, will create a more precise function. This is because at every extreme value, the function F^{λ} makes a jump and becomes the average of the values before reaching another extreme. This effect is visible in the figure below. The taut string is used to fit the Doppler effect[2]. On the left panels is the tube with the string within. The upper panel has a large tube width, the lower panel a smaller tube width. The right panels show the result of the fit.

In the upper panels it is visible that for every extreme value, i.e. the string touches the upper or lower bound, the value of the function makes a jump. The value of the

[2] http://en.wikipedia.org/wiki/Doppler_effect

function becomes the average of the string between two extremes. The lower panels show a better approximation of the Doppler effect.

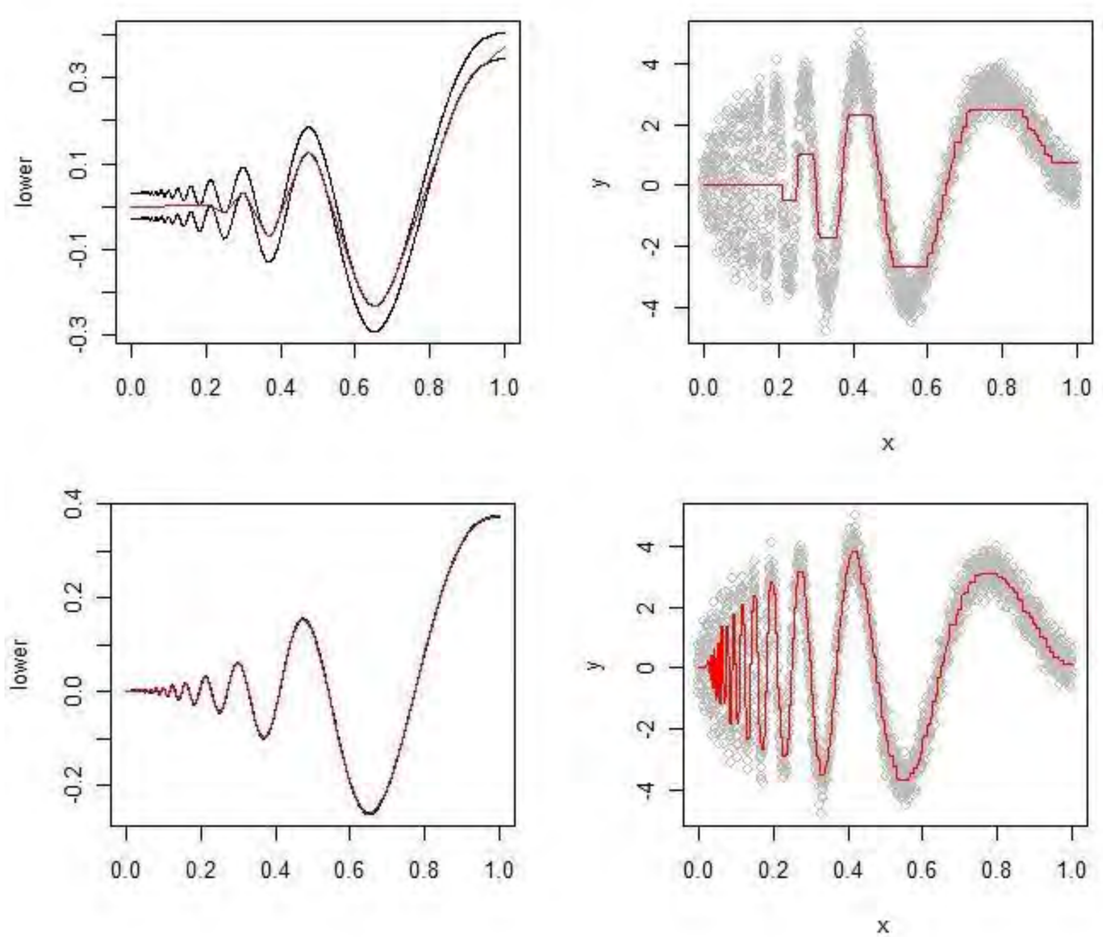

FIGURE 10: THE DOPPLER EFFECT VIA TAUT STRING

The formula for the taut string method is

$$T(F) = \sum_{i=1}^{n} \left(y_i - \hat{f}_i\right)^2 + \sum_{j=1}^{n-1} \lambda_j \left|\hat{f}_{j+1} - \hat{f}_j\right|$$

The formula consists of two parts. The first part is the difference between the found function and the original data points squared. The second part is the total variation of the function $\hat{f}$.

The total variation of a function is the distance travelled along the y-axis (one dimensional function) if we follow the function along the x-axis. The easiest function to calculate the total variation is a straight line. The total variation of this function is $|f(x_{max}) - f(x_{min})|$.

The simple formula above is actually the same formula as the second part of the taut string formula. Because every function other than a straight line cannot be computed by subtracting the minimum value of the maximum value, the summation is needed. A complex line is cut into n equally distanced parts to create small lines which are approximated by a straight line. The formula to calculate the total variation of a straight line is known, thus summing the approximations of the n pieces gives the total variation of the complex line.

These two parts combined make the taut string method a fast algorithm for minimising.

6.1 QUANTILE REGRESSION(10)

Quantile regression is a form of regression where the regression results in estimates of approximating either the median or other quantiles of the response variable. It is possible to solve these problems with the taut string method. Therefore, the formula is a little bit adjusted;

$$T(F) = \sum_{i=1}^{n} \left[\frac{|\hat{f}_i|}{2} - \left(\beta - \tfrac{1}{2}\right) \hat{f}_i \right] + \sum_{j=1}^{n-1} \lambda_j |\hat{f}_{j+1} - \hat{f}_j|$$

Where β is a quantile.

6.2 POISON REGRESSION(10)

If the data available is Poisson distributed, Poisson regression is the key to the problem. Another extension of the standard taut string is the ability to let it solve Poisson regression problems. The formula of the original taut string is adjusted in the following way to handle Poisson distributed data.

$$T(F) = \sum_{i=1}^{n} \exp(\hat{f}_i) - \hat{f}_i y_i + \sum_{j=1}^{n-1} \lambda_j |\hat{f}_{j+1} - \hat{f}_j|$$

Although the data is based on a call centre, the number of calls is probably not Poisson distributed. The data is real life data, which is most of the times different from the theoretical data. In theory this extension of the taut string should work best on the data.

6.3 USING THE TAUT STRING EXTENSIONS

The taut string algorithm is capable of using two dimensional data at once. This means that the 40 selected days can be given to the algorithm and that it will come up with a one dimensional function; our forecast.

The implemented algorithm does not really see the data as two dimensional, but it will see it as values on a two dimensional grid. This grid is like the x-axis and y-axis of a graph and the values of the days are fit into that grid.

In (10) a section is devoted to the choice of the tuning parameters λ. This parameter can be fixed for all data points or it can be adaptable during the calculation of the algorithm.
Since we want the deviations between the real data and the estimation to be as small as possible, we will set the algorithm to use an adaptable λ.

7 EXPERIMENTS

The experiments consist of numerous forecasts of the number of calls received. Every method described in the previous two chapters will be tested on 12 different data strings. A data string is for example the number of calls received on a Tuesday during the first 20 intervals. The 12 data strings will differ in length (number of intervals known) and day.

For every method and for every data string a string with numbers will be created by the forecast. The original data string will be compared with the forecast by a Pearson correlation distance. This distance calculates the likeliness of the two strings.

The methods used for the forecasting are;

- M1: Arithmetic mean
- M2: Median
- M3: Moving average
- M4: Exponential smoothing
- M5: ARIMA(1,1,2)
- M6: Taut string with quantile regression ($\beta = 0.5$)
- M7: Taut string with Poisson regression

A data string is chosen from the dataset available. The chosen data string will be deleted from the dataset as long as the data string is used for forecasting. This prevents the methods from using the original data within a forecast, which would influence the forecast.

The dataset contains 261 days with data. Every row has 68 intervals (columns). To create a forecast, the algorithm will search for 40 rows which have the smallest Pearson correlation distance with the given data string. These 40 rows are used to create the forecast.

The more intervals are known, the more data is available for the Pearson correlation distance (PCD). For every known interval, the PCD is calculated. With more intervals known, the difference between rows increases. Two rows can have the same number of calls on the first interval, but the rest can be very different. With a single interval known, this difference won't be noticed by the PCD.

The data used as input for the forecast is displayed in the table below. The first column contains the day, the rest of the columns contain the number of calls for the intervals [07:00 07:15] (1) till [11:15 11:30] (19). If an interval has no data in it, the interval is not known.

Code	Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
D1	Mon	9	12	10	11	14	19	32	19	27										
D2	Mon	9	17	11	19	18	17	20	16	26	37	48	54							
D3	Mon	14	7	10	10	19	15	24	27	22	25	26	32	28	36	28	31	20	28	20
D4	Tue	6	9	9	13	11	21													
D5	Tue	15	15	24	21	24	35	35	41	35	50	54	53	60	56	46	44	43		
D6	Wed	6	7	6	19	13	11	26	20	27	46	33	28	21	26	26	27	20	24	32
D7	Thu	7	9	9	18	6	17	22	36	34	29									
D8	Thu	15	11	15	12	29	31	37	29	39	27	25								
D9	Thu	6	7	11	13	21	19	23												
DA	Sun	11	25	21	14	22	36	40	42	30	27	33	40	51	41					
DB	Sun	8	5	10	7	18	19	37	32	33	32									
DC	Sun	2	7	8	15	11	19	29	25	37	40	45	47							

TABLE 4: THE INTERVALS USED FOR INPUT

8 RESULTS

> I have to make a remark about the data first. After all the tests were performed, I saw that there was an error in the data of D1. This day had a 15 minute interval with 137 incoming calls. The rest of the day looked normal, and I changed the value to 37. The number of calls in the previous interval is 39 and in the next interval is 20. The value of 37 looked like a good value, because an extra 1 could have been added by error.
> This had no implications on the 40 selected days, but it had an impact on the PCD of the forecasts since the error did not occur in the intervals to select the days. The only difference is that the PCD for all methods has gone up, but the order of best to lower performance is exactly the same.

The results show that forecasting the number of calls in a call centre is difficult by using the number of calls per interval. The forecast changes relatively much as more and more data becomes available during the day. And it would probably be easier to forecast the total calls of a new day.

The table below shows the PCD of the forecasted number of calls. It can be seen that none of the methods used really outperforms the other methods. The simple methods, like mean and median, outscore the more complex models on more than one occasion.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC
Mean	0,782	**0,187**	0,718	0,849	0,823	0,403	0,779	0,777	0,793	0,730	0,845	0,609
Median	0,779	0,178	**0,725**	0,840	0,827	0,387	0,759	**0,795**	0,798	0,716	0,839	0,596
Moving average	0,770	0,163	0,691	**0,864**	0,824	0,372	**0,784**	0,789	0,807	0,748	0,847	**0,612**
Exponential smoothing	0,785	0,151	0,700	0,856	0,819	0,373	0,783	0,790	**0,808**	0,748	**0,850**	0,612
ARIMA	0,772	0,112	0,672	0,833	0,802	0,335	0,772	0,782	0,781	**0,754**	0,843	0,598
Taut string quantile	0,782	0,186	0,724	0,843	**0,828**	0,403	0,765	0,792	0,800	0,713	0,830	0,593
Taut string Poisson regression	**0,789**	0,106	0,681	0,793	0,782	**0,450**	0,726	0,695	0,412	0,749	0,790	0,563

TABLE 5: THE PEARSON DISTANCE CORRELATION FOR THE FORECASTED DATA

The moving average has the best overall performance. Three times it creates the best forecast, and several other times it is second best. Still it cannot be said that one method is better than another method. Therefore the differences in PCD are still very small and the number of occurrences to little.

For the experiments where the taut strings performed best, the forecasts and the actual data are displayed in the figures below. The figures show that, although the forecasts have the best likeliness to the real data, the forecasts are not that good.

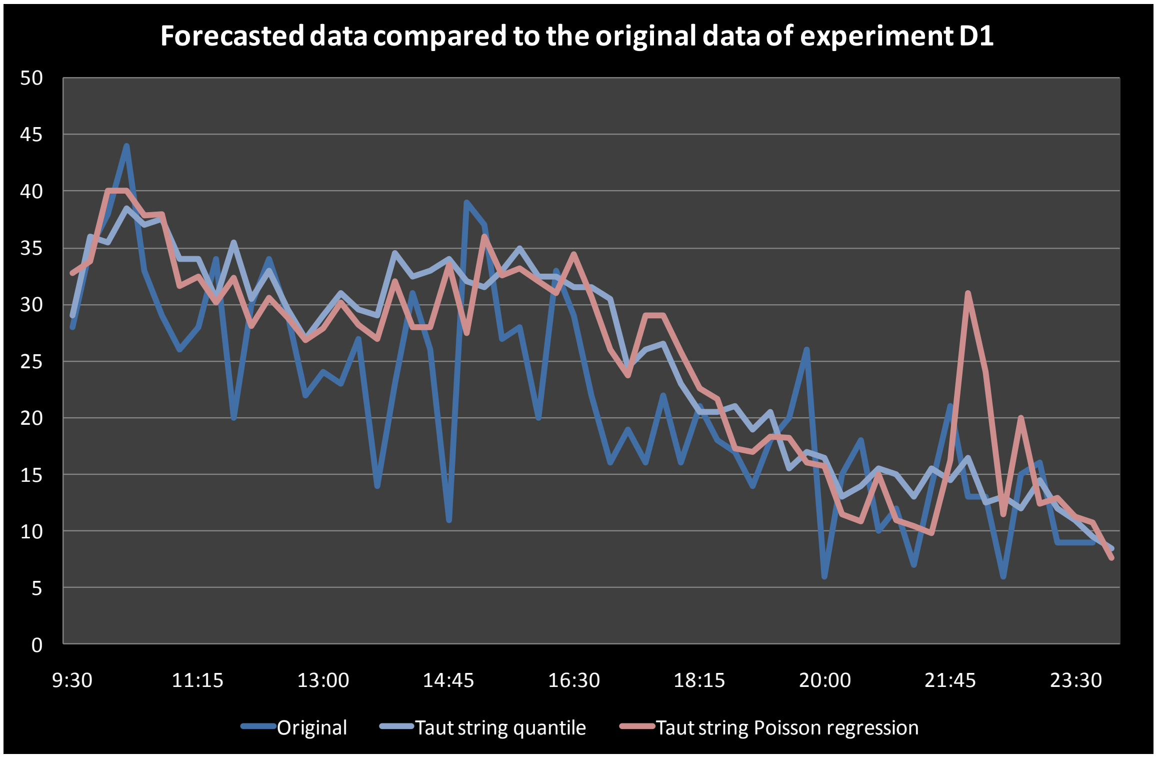

FIGURE 11: THE FORECASTED DATA OF EXPERIMENT D1

In Figure 11 the forecast of the taut string with the Poisson regression has the best likeliness with the original data. It can be seen though that the forecast is mostly too high and does not really follow the pattern of the incoming calls.

The forecasts displayed in Figure 12 are not that good either. The pattern is followed a bit more strictly, but the forecasts are consistently too low.

For Figure 13 the same results as Figure 11 can be seen. The forecast to high and the pattern not really followed. The anomaly here is that the incoming calls start way too high and it takes a long time before the forecasts are reasonably good again.

Another thing we see at the figures is that the original data is very jagged. Between two subsequent intervals, the differences can be up to hundreds of percentages. This is a big change is a short period, which is difficult to keep up with.

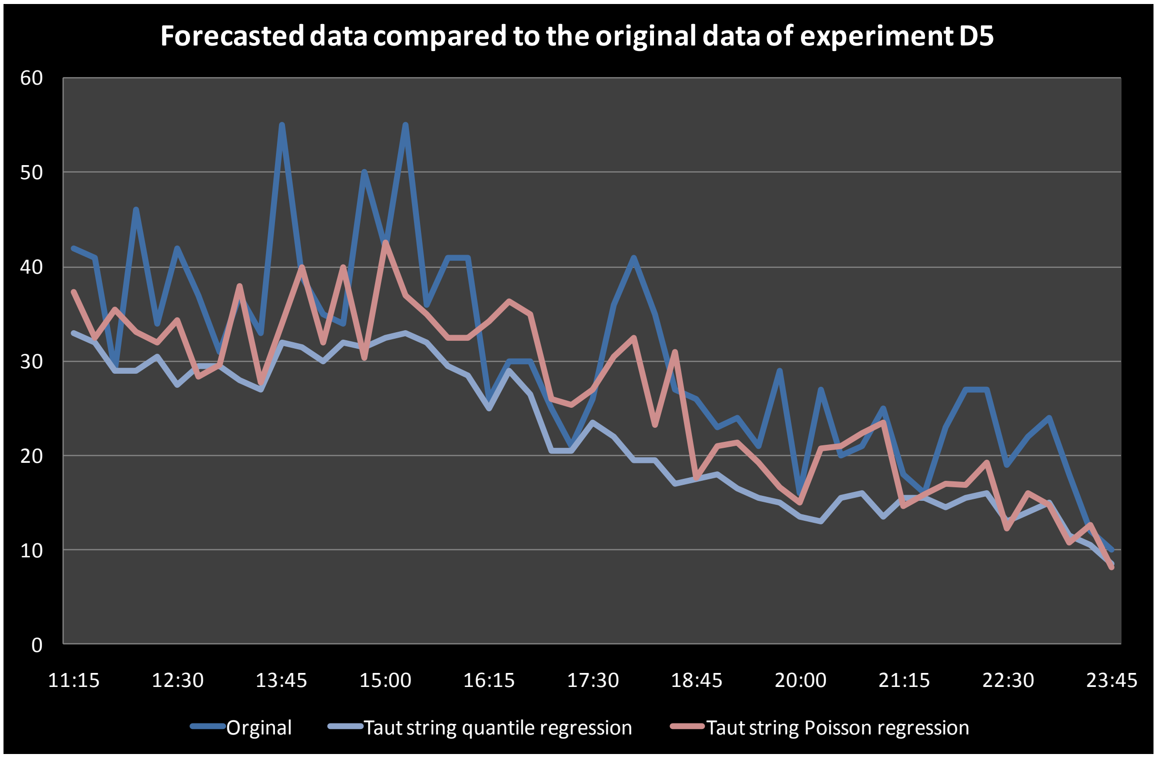

FIGURE 12: THE FORECASTED DATA OF EXPERIMENT D5

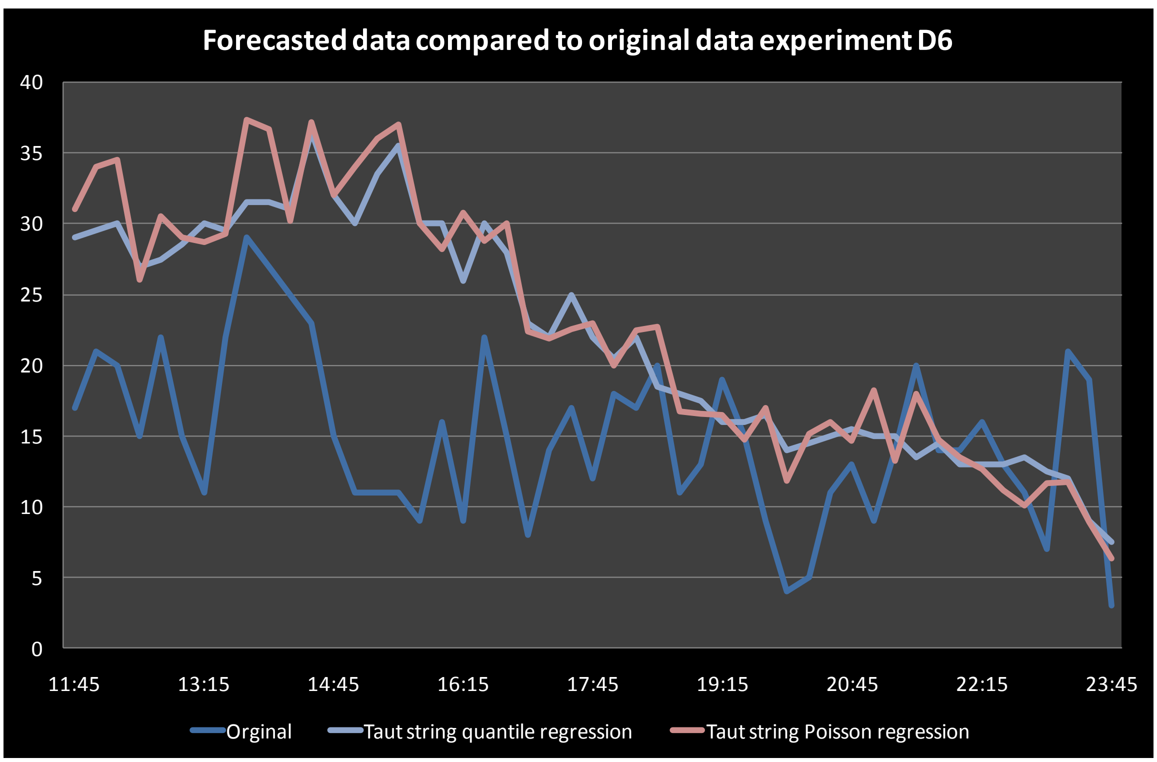

FIGURE 13: THE FORECASTED DATA OF EXPERIMENT D6

9 CONCLUSION

The taut string algorithms can be used to forecast the number of calls in a call centre. In the setting used, it does not perform better than the average or the median though.

A reason why the taut strings did not perform very well might be explained by the form of the data. The days used are alike. For every day put in the data, there are just more data points on one interval. The taut string might want to create a function between those data points, but it can only create a line from interval to interval.

The figure below shows the number of incoming calls for the 40 selected days on the interval [9:15 24:00]. It is clear that there is no function which would touch all the dots in the figure, thus the best function is sought. Since this best function has to minimize the error for 40 days, it is possible that the function does not find the best solution for the problem.

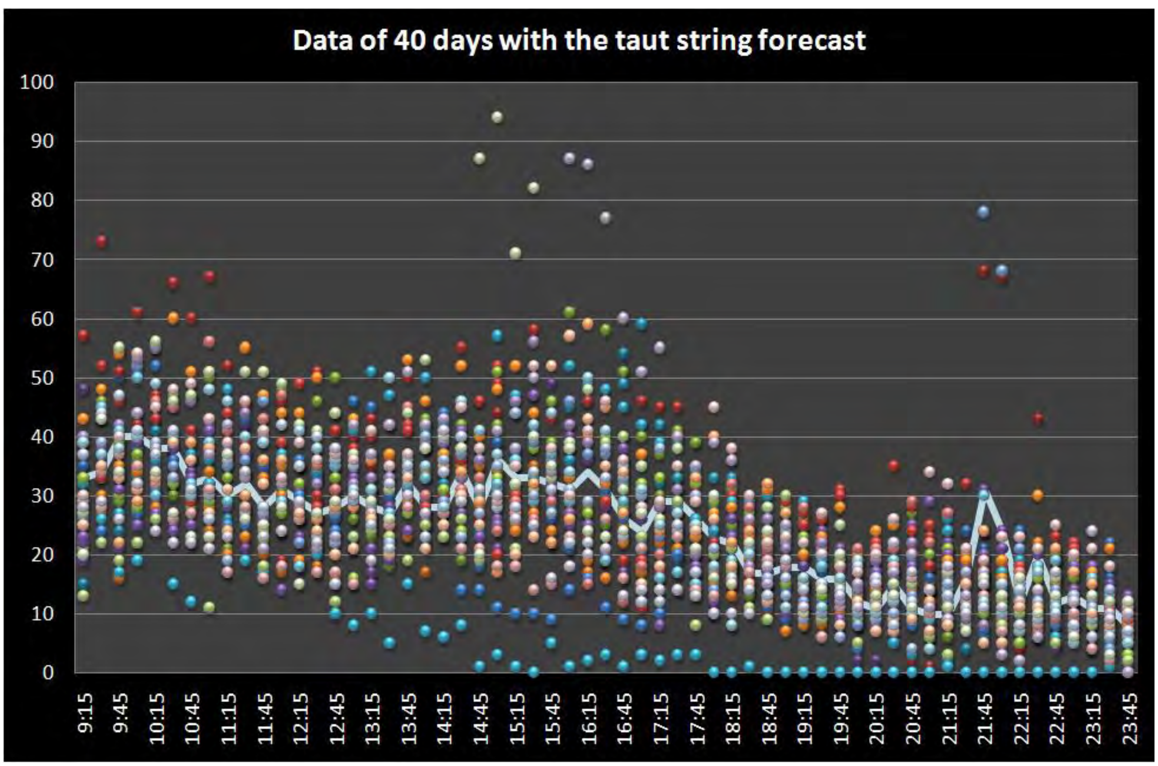

FIGURE 14: THE DATA OF ALL SELECTED DAYS AND THE FORECAST OF THE TAUT STRING WITH POISSON REGRESSION

In Figure 14 all days and the forecast of the taut string Poisson regression is displayed. The figure tells us that the function found is nice. Of all available data points, the line is somewhere halfway. The question is, would the forecast be better if we had created a function per day and then take the average?

It could also be that the number of days is too much. With a smaller amount of days, the characteristics of an individual day should be better preserved. An experiment with more than 40 days can be conducted as well, but in that case, with just 261 days in the dataset, the chance of creating the same forecast for different known intervals increases.

Another interesting fact is about the number of intervals known. I thought that an increasing number of intervals known, would lead to better forecasts. This is not the case. The worst forecasts are made when most information was available. Take experiment D6 for example. 19 intervals were known, but the Pearson correlation distances of the forecasts did not come higher than 0.45.

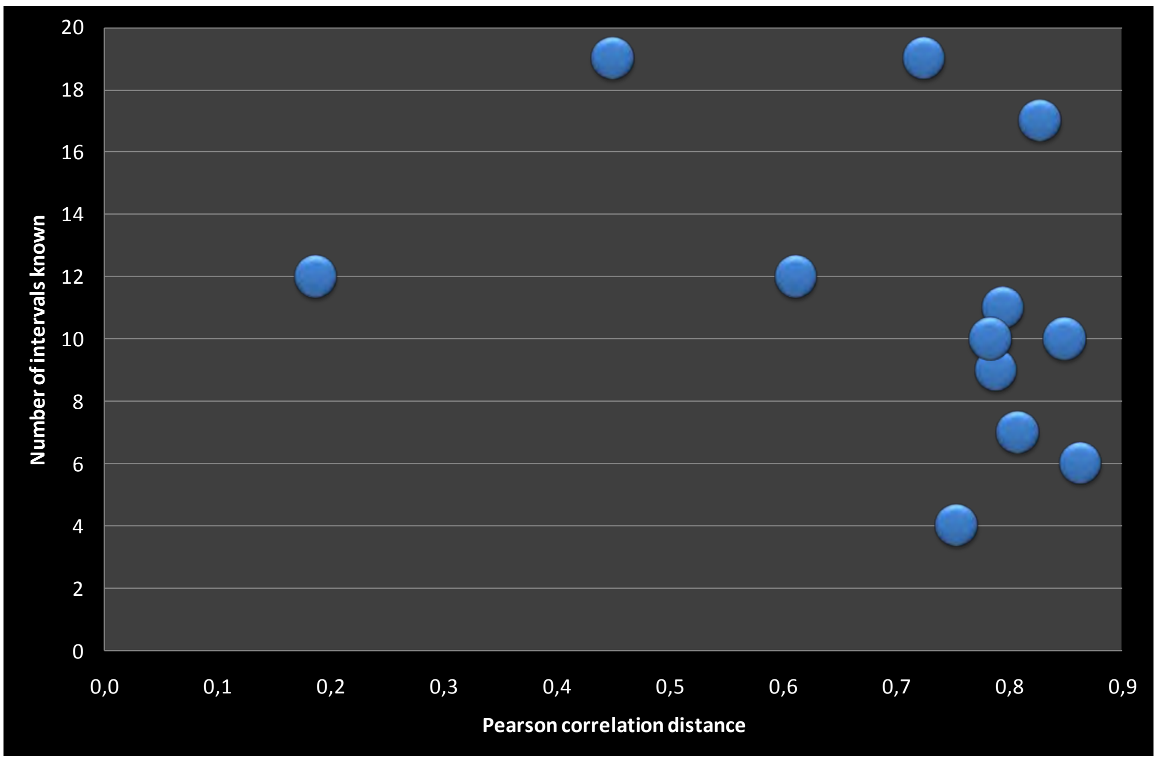

FIGURE 15: RELATION BETWEEN INTERVALS KNOWN AND PCD

According to my thoughts there should have been a line from the lower left corner to the upper right corner. This is clearly not the case. It might be explained by the Pearson correlation distance. The less information is available for this formula, the easier it is for two string to look alike, or just the opposite. While longer strings can have more disruptions to become not alike anymore.

Another explanation for the bad performance when a lot of information is known might be the dataset. The PCD with a lot of information can better select days which are alike, but what if there are just 20 of those days available? Then another 20 days which are not very much alike will be selected as well. These not alike days will disrupt the forecasts.

10 RECOMMENDATIONS

I would recommend to use other settings for testing the forecasting capabilities of the taut strings. During this research I have learned that is more than one solution to this problem.

At first, I want to recommend to use a bigger dataset. 261 days seem to be a lot, especially if you take the effect of changing characteristics of the incoming calls with the change of business rules, capabilities of agents etc, but it seems not to be enough to hold a lot of different patterns. Every day is unique, but with a lot of days available which are about the same, better forecasts could be made.

The second recommendation is to use less days for a forecast. 40 days seems to be too much. With this many days the specific characteristics of the selected days seem to disappear. I think it would be very useful to forecast with 20 days.

As third, it would be more appropriate to test the taut string with more different settings. In this research just two settings were used, but there are a lot more values which can be used. The taut string with the quantile regression should be tested on the most common quantiles at least. And for both variants of the taut string, the value of λ could be fixed. This has an impact on the total variation and will definitely have an impact on the forecasts.

A fourth recommendation is to change the input to the taut strings algorithm. Instead of entering all data in the algorithm at once, it might be interesting what would happen if the taut strings work on the individual days and then take the mean or the median from the results as a forecast.

We have seen that the real world data is not always perfect. A lot of big changes happen during very short periods of time, which are unusual for a theoretical call centre. To prove the potential of the taut string methods, it is possible to simulate call centre data and test the forecasts on the generated data. This might prove that the methods work very good, in theory at least.

At last I can think of varying the number of known intervals per day. The research concentrated itself on the different days, with different number of intervals known. But does the forecasts differ that much as well as the day is kept the same, but the number of intervals is changed?

11 INDEX

12 REFERENCES

1. How to Improve Service with Better Forecasting. *Lionhart Publishing Online.* [Online] August 1997. [Cited: 25 August 2009.] http://www.lionhrtpub.com/IM/IMsubs/IM-8-97/BetterForecasting.html.

2. **Yannick Veys.** Meest voorkomende redenen faillissement. *Higher Level.* [Online] 3 March 2008. [Cited: 25 August 2009.] http://www.higherlevel.nl/forum/index.php?board=47;action=display;threadid=18678.

3. **Vince Beiser.** Store Power in Super Batteries. *Wired Magazine.* [Online] 23 March 2009. [Cited: 25 August 2009.] http://www.wired.com/science/discoveries/magazine/17-04/gp_storage.

4. **Koole, Ger.** Call Center Mathematics. *CCmath.* [Online] 26 January 2007. [Cited: 26 August 2009.] http://www.math.vu.nl/~koole/ccmath/book.pdf.

5. **DragonBot.** Average. *Wikipedia.* [Online] 17 November 2008. [Cited: 2 December 2008.] http://en.wikipedia.org/wiki/Average.

6. **Michael Hardy.** Moving Average. *Wikipedia.* [Online] 30 January 2009. [Cited: 15 February 2009.] http://en.wikipedia.org/wiki/Moving_average.

7. **Zvika.** Exponential Smoothing. *Wikipedia.* [Online] 16 November 2008. [Cited: 03 December 2008.] http://en.wikipedia.org/wiki/Exponential_smoothing.

8. **Melcombe.** ARIMA. *Wikipedia.* [Online] 23 February 2009. [Cited: 15 March 2009.] http://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average.

9. **Kovac, Arne.** Local extreme Values. *University of Bristol.* [Online] 13 December 2006. [Cited: 19 March 2009.] http://www.maths.bris.ac.uk/~maxak/ast.pdf.

10. **L. Dümbgen and A. Kovac.** Extensions of smoothing via taut strings. *University of Bristol.* [Online] 2009. [Cited: 2 April 2009.] http://www.maths.bris.ac.uk/~maxak/L1Strings09.pdf.

11. **Gunst, M.C.M. de.** *Statistical Models.* 2008.

12. **Koole, Ger.** OBP. *VU University Amsterdam.* [Online] 16 March 2009. [Cited: 26 August 2009.] http://www.math.vu.nl/~koole/obp/obp.pdf.