Development of a prediction model for the number of deliveries at the obstetric department

Bachelor Thesis

Frank Buisman

S1862502


isala

Development of a prediction model for the number of deliveries at the obstetrics department

Bachelor thesis report

17-06-2019

Auteur

Frank Buisman
S1862502
f.buisman@student.utwente.nl
Bachelor Technische Bedrijfskunde
Industrial Engineering and Management

Universiteit Twente

Dr. Ir. M.E. Zonderland
Dr. Ir. A.G. Leeftink

Isala

L. Hofman, Msc

Foreword

My bachelor thesis is called: “Development of a prediction model for the number of deliveries at the obstetrics department”. This report is written as a final test for the bachelor Industrial Engineering and Management at the University of Twente. The research is executed at the obstetrics department of the Isala hospital in Zwolle. In this research the current situation of the obstetrics department in analysed and an analysis with a theoretical background about birth forecasting and prediction models is included as well. At last, the research is focused on the development of the forecasting model.

In this foreword I want to thank my thesis supervisor at Isala, Laura Hofman. I enjoyed working on the project, due to the helpful guidance and the open and kind attitude of the hospital. You invited me for several meetings about the logistics and capacity management in the hospital, which were interesting. It gave me the insights in what it would be like to work in a hospital. I have learned a lot from you about working in the hospital and about the preparation and validation of data. Besides, I want to thank Ada Pot, the head of the obstetrics department, for her openness. I was always able to reach out to her and ask her everything.

Further I want to thank Gréanne Leeftink and Maartje Zonderland, both my bachelor thesis supervisors of the University of Twente. You provided me useful feedback and I have learned a lot from you.

Finally, I want to thank all my family and friends for their support and interest in my thesis.

I hope you will enjoy reading my bachelor thesis!

Frank Buisman
Enschede, June 2019

Management summary

Problem analysis

The obstetrics department of Isala is a place where women give birth and the hospital is entitled to high quality care by qualified staff, at all times. The department is not always able to provide enough and sufficient care. This is caused by a large fluctuation in the number of deliveries while the department uses a static planning. On a tactical basis the personnel should be planned smarter in order to adapt to the peaks. However the obstetrics department has no input for the personnel planning and therefore a prediction model for the number of deliveries at the department is investigated and developed.

Literature

Currently, there is no suitable prediction model for the number of deliveries or for a comparable situation available. Based on the literature two possible models are found. The first model is based on historical birth fluctuations. The peaks in the number of deliveries take place in January and September, according to the CBS. If this seasonality exists at the obstetrics department as well, a periodically distributed model that predicts the number of deliveries at the obstetrics department is suitable.

The second developed model states that there are a lot of other factors and events that influence the number of deliveries and therefore the prediction of the number of deliveries is independent of the past, unlike the periodically distributed model. This second model computes the expected number of deliveries at the department based on the population pregnant women in combination with:

1. The average percentage of women giving birth at Isala compared to the whole population pregnant women
2. A probability distribution that states the weekly chance of giving birth in e.g. week 36 till 42
3. A statistical method, called convolution, which computes the expected number of deliveries and the most likely interval of the number of deliveries per week.

The analysis of both models is described in the next two chapters.

Periodically distributed model

The basis of the periodically distributed model exists of a level, trend and seasonal factor. In the basic periodically distributed model, these variables are all fixed and the seasonal factor is computed by the average seasonal factors of the years the model is based upon. Two models are developed, one based on the data of 2016 – 2017 and one based on 2016 – November 2018.

Because these two models all have fixed variables, the model is not flexible for changes in seasonality. Therefore two extensions of the models are investigated.

The first extension states that seasonality can change over the years. Therefore the seasonal factor of the past year should be of more importance than the seasonal factor of the former year(s). Again, there are two models developed based on the input of the different years. In order to find the most important and suitable weighted factors the solver tool is used.

The second extension is using exponential smoothing on top of a periodically distributed model. Exponential smoothing is normally based on the former forecast period and it is used to determine to what extend the forecast should adapt to the forecast error in the former period. The adaption in the forecast error is based on the exponential smoothing factor, $0 \leq \alpha \leq 1$. If this value is zero, the error of the forecast is attributed towards coincidence and a value of zero indicates a structural change in the number of deliveries. This model is not suitable for the obstetrics department since it is not possible to update the number of deliveries every week. However, this method can be adapted and then it can be used in two ways:

First method: $F_{t \mid Y} = F_{t \mid Y-1} + \alpha * E_{t \mid Y-1}$

In this way the forecast for 2016 is determined with exponential smoothing and for the following years the formula is used. In this case, the exponential smoothing is used to investigate whether there is a structural change in the number of deliveries for what should be compensated. So the exponential smoothing is built on top of the periodically distributed model that forecasts the number of deliveries in 2016.

Second method: $F_{t \mid Y} = P_{t \mid Y} + \alpha * E_{t \mid Y-1}$

In the second method the forecast for every year is based on a periodically disturbed model. The periodically distributed forecast per week can be adapted with the forecast error of the period a year ago by using exponential smoothing.

For both methods the solver tool is used to find the optimal value of alpha. For both methods, regardless of the used periodically distributed model, the value of alpha results in a value of zero.

Validation

The models are validated with the bias, MAD, MSE, percentage error, MAPE, Average Tracking Signal and the graphical representation of the forecast and the actual number of deliveries. The KPI's with a weighted average seasonal factor show a less accurate forecast than the model with the average seasonal factor. The forecast, and thus the values of the KPI's, of the basic periodically forecast for the model based on 2016 – 2017 and the model based on 2016 – November 2018 are comparable, with an exception of the week 38 and 39, also called the Christmas and New Year's Evening peak. Taking this peak and the amount of data the model is based upon, the model based on 2016 – November 2018 is considered as most reliably.

Since the optimal values of alpha for exponential smoothing are all zero, the exponential smoothing does not change the forecast. It could either be stated that exponential smoothing on top of a periodically distributed model is not acceptable or it shows that the developed periodically distributed model is suitable and that there is no need for compensating in forecast error.

Convolution model

By analysing the data of all pregnant women with their parturition data in the region of Zwolle with the actual number of deliveries per month, the average percentage of women giving birth in Isala is 83,0% with a standard deviation of 6,8%. This standard deviation results in a possible difference of 58,2 deliveries per month, which is 13,1 deliveries per week. On top of this uncertainty the probability distribution of a woman giving birth in e.g. week 36 up and until 42 of the pregnancy is added.

Comparing the variation in the model with the actual deviation in the number of deliveries per week, which is 8,8, the variation of the model exceeds the actual variation. Therefore the uncertainty in this model is considered too large. In this research the choice is made to stop developing this model.

Conclusion

The action problem is that the obstetrics department is not always able to provide enough and sufficient care. The core of the problem is that the obstetrics department has no input for their employee planning in order to adapt the planning to the large fluctuations. Therefore the goal is to develop a model that predicts the number of deliveries, which is the input for the personnel planning.

Since the convolution model is considered unreliable, the focus is on the periodically distributed model. Based on the KPI's, the graphs and the amount of data the model is based on, the basic model based on 2016 – November 2018 is considered as most reliable.

There are some external changes, which should be compensated. Some patients in the neighbourhood of Lelystad, Harderwijk and Hoogeveen are going to Isala due to bankrupted hospitals or hospitals which close their doors temporarily. In order to compensate for the external changes, an analysis for 2019 in excel is executed and it results in a compensation of 1,6%. Therefore the forecast is multiplied with a factor of 1,016.

With the developed model and compensation, the research accomplished its goal to develop a model that predicts the number of patients at the obstetrics department.

Recommendations

First of all, the head and planners of the obstetrics department should investigate how the forecast will affect the personnel planning and when the department should shift up or down on its personnel planning.

Besides, the obstetrics department should use the developed template, including instructions, of the periodically distributed model that predicts the number of deliveries. The periodically distributed model predicts the number of deliveries and the exponential smoothing part will automatically compensate for changes in seasonality. If there are too many changes in seasonality, the data of the last two years should be used as an input for the next periodically distributed model, so the template is used again, in order to make a new forecast.

A lot of information in the data lack value. Therefore the obstetrics department should think about what data they want to keep track of, considering both the logistical and the health care aspects.

Besides that the periodically distributed model is not able to forecast the Christmas and New Year's evening peak because Christmas falls in a different week every year. It is hard to predict in which week, 38 and or 39, this peak will be. Therefore the planners of the department should be aware of this unpredictable peak.

Inpatient and outpatient deliveries need different caregivers. For this research this distinction does not matter since the hospital needs to handle an inpatient delivery and facilitate an outpatient delivery. If the business intelligence employees are able to make this distinction, it would be interesting to investigate the difference between the two in order to make a better forecast.

Table of Content

Glossary of Terms

ATS	Average Tracking Signal
MAD	Mean Absolute Deviation
MAPE	Mean Average Percentage Error
MSE	Mean Squared Error
OD	Obstetrics Department
TS	Tracking Signal

1 Introduction

This chapter will provide some general information about the Isala hospital, the obstetrics department (OD), the motivation and a brief description about the research.

1.1 Isala

Isala is a large regional hospital located in Zwolle, Meppel, Steenwijk, Kampen and Heerde. Isala offers basic care and is recognized as a top clinical hospital, meaning that Isala has a number of high quality specialized departments. This means that Isala has some high specialized departments. Examples of this within Isala are dialyse, electrophysiology of the heart and 15 other departments.

The 'Vrouw-kindcentrum' in Zwolle is the department of Isala for healthcare for women and their partners, pregnant women, newborn babies and children. This department is the largest birth centre of the Netherlands and exists of four specialisations: Fertility, gynaecology and obstetrics, neonatology and paediatrics.

The obstetrics department is the department where women give birth. This department is open 24 hours per day, 7 days a week. Every patient in this department is entitled to high quality care offered by qualified staff, available at all times. This obstetric staff exists of gynaecologists, doctors, clinical obstetricians, paediatrician, anaesthesiologists, interventional radiologists, sonographers and obstetric nurses with an O&G education.

1.2 Motivation for research

It occurs that the staff occupation at the obstetrics department is insufficient to fulfil a safe and adequate care for patients. The number of patients and the availability of staff is out of balance and Isala has no view on the expected number of deliveries for the upcoming period. Therefore Isala has started an international research concerning this problem. A team existing of an obstetrics professional, an improvement coach and a capacity advisor has started to investigate the problem and they agreed on the need for a prediction model. This research is a thorough investigation on the matter and the possible solution(s). The action problem is stated as follows:

Action problem: *"The obstetrics department is not always able to provide enough and sufficient care for its patients."*

1.3 Problem statement

The stated action problem is the result of a misbalance between the number of obstetric nurses and the number of deliveries at the clinic. This gap is undesirable for the employees, because they are called in last-minute during high peaks and sent home when there are no patients in the department. This leads to less employee satisfaction and less flexibility.

The misbalance is caused by the fact that the availability of obstetricians is not adjusted to the number of deliveries, which has two causes: On the one hand the number of deliveries at the obstetric department fluctuates a lot and on the other hand the obstetrics department has a static personal planning.

Since this static planning model is not reactive on large fluctuations, it is not sufficient. Isala has a static planning because there is no sufficient input for the planning. The input for this model should be the expected number of deliveries at the department in the upcoming weeks but there is no model available which can predict this. Therefore the core problem is described as following:

Core problem: "There is no model that predicts the number of deliveries at the obstetrics department for the upcoming week."

This core problem has the following purpose:

Purpose of research: "Develop a model that can predict the deliveries of patients for the upcoming weeks at the obstetrics department."

Figure 1 summarizes this problem description.

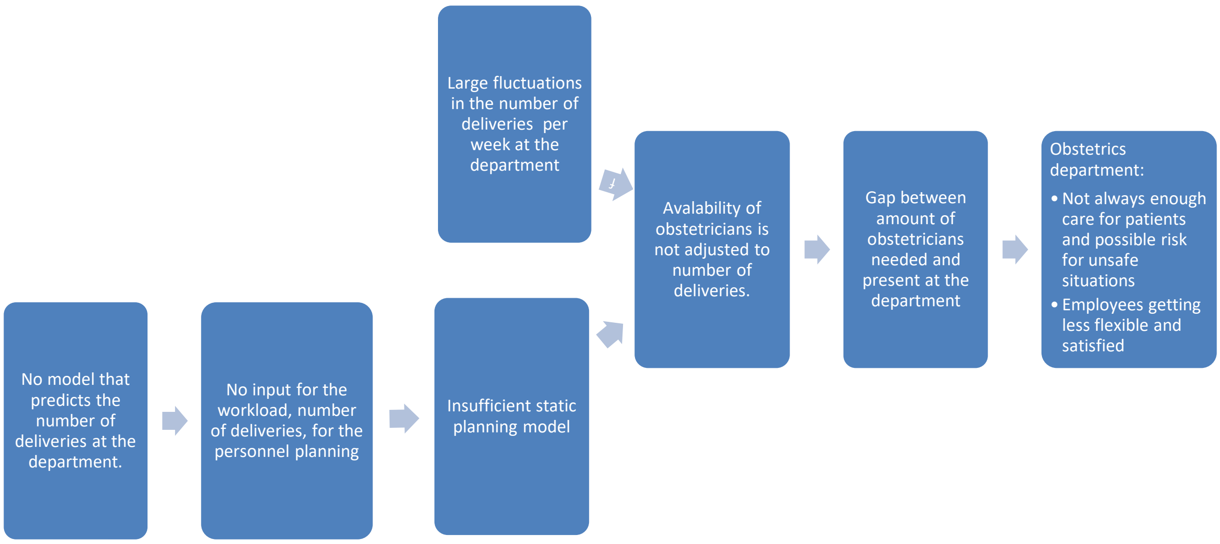


Figure 1: Problem cluster

1.4 Research questions

To execute the research several sub questions should be executed. This section provides an overview of the sub question of every section with its subject or sub question, including motivation and problem solving approach.

1.4.1 How does the current situation looks like? – Chapter 2

To get a good overview of the current situation it is important to speak to several people with different backgrounds related to the problem. Suitable people are the head of the obstetrics department, who requested this research, the medical coordinator at the department and the improvement coach, who made a start with the research. Besides, the situation should be supported by a short data analysis.

1.4.2 What kind of models exist for this situation? – Chapter 3

When the situation at the department is clear the question rises what kind of model is suitable for the situation. First a short background about forecasting and forecasting in a healthcare environment is given. Thereafter it is important in what way the number of deliveries is distributed and what variables could predict or influence the birth-rate. At this moment it looks like there are two type of models that should be investigated: A periodical model based on historical periodical trends and a model that computes the number of deliveries for the upcoming weeks based on the data of all the pregnant woman in the region. This part of the research is the theoretical part and will be based on literature.

1.4.3 What data can be used? – Appendix A

The next step is a check if we have sufficient and reliable data to develop the models that were found in the literature. There are different sources available for the data, so the right data should be collected, validated and processed in a way that meets the privacy and security rules of the hospital and the Dutch law. When the date is found suitable, the two models can be build.

1.4.4 Is the number of patients at the obstetrics department periodically distributed? – chapter 4

Based on the literature found in chapter 3 and the assumption that the fluctuation in the number of patients is comparable to the national birth fluctuation, the prediction for the upcoming periods could be based on a level, trend and seasonal factor. The data from the deliveries of the last few years is prepared and used to develop a periodically distributed model. After the development, the model is immediately validated. The validation is checked with the bias, Mean Absolute Deviation, percentage error, mean average percentage error and tracking signal. The developed periodically distributed model is validated and compared with the current model that is based on the average number of patients at the department, because it should be measured if the developed model is more accurate than the static model.

1.4.5 Convolution model – Chapter 5

As described in chapter 3, there are other factors that influence the number of deliveries. So, a periodically distributed model would not be not be suitable. A model that predicts the number of deliveries at the OD could be based on the data of the pregnant women in the region. This model is created via convolution model and is programmed in VBA. This model should also be validated with KPI's. The same KPI's should be used as the validation of the periodically distributed model to measure the improvement.

1.4.6 Conclusion and recommendations – Chapter 6

What model provides the most accurate forecast and is the forecast sufficient enough? It is also important to understand what the meaning of the outcome is and what the consequences are for the obstetrics department.

2 Analysis of current situation

This chapter describes a thorough analysis of the problem statement.

2.1 Patients

The patients that give birth at the OD are patients who are pregnant and in 32 to 40 weeks. The patients are classified into two different groups and depending on the classification there is a different guidance during the delivery:

- Inpatients: These are patients who have a medical indication and are guided by a gynaecologist and assisted by an obstetric nurse, which is second line.
- Outpatient: These patients do not have a medical indication and are guided by a first line, so someone extern. The assistance is from an obstetric nurse from Isala or a maternity assistant (NL: kraamverzorgende) or an extern obstetric nurse.

2.2 Planning & fluctuations

Planning schedules can be categorized in three hierarchical levels (Hans, Van Houdenshoven, & Hulfshof, 2011):

- Strategical level: Planning for upcoming years.
- Tactical level: Planning for upcoming weeks / months.
- Operational level: Planning for upcoming hours / days.

The situation at the department would not have been a problem if there was flexibility on operation level. So if the OD was able to adapt to the busy and calm moments really quick. But this is not the reality.

Every department in Isala has to deal with fluctuations in the number of patients. Isala has created a flex pool of nurses that can be called in to cover a shift during busy moments at the department. However the obstetric department cannot make use of this flex pool because the nurses at the obstetric department need a special O&G certificate. Due to the fluctuation in the number of childbirths and because of the need for an obstetric nurse during every childbirth, it is important to have a forecast for the number of deliveries in order to make a planning.

Besides the fact that an arriving patient almost immediately needs to be treated, a woman who needs a caesarean is dependent on the fixed OK block. There is only one patient type that can be treated in a flexible way and that is a patient that needs to be 'primed'. This means that the pregnant woman is being artificially prepared to give birth. Since this moment of 'priming' can be determined, it is easy to shift this process a few days to a more quiet moment. At maximum there are 8 patients planned that needs to be primed, but it is not possible to delay a patient again and again. So with every delay of a patient less flexibility is created.

On operational level there is not much flexibility. An excel file with all the data of the 'Vrouw-Kindcentrum' from 2016 until week 13 in 2019 is available. All the childbirths are filtered and analysed. Comparing the busiest and the most quiet weeks shows there is a statistical difference.

On average, there are 10,1 deliveries at the clinic with a variation of 3,1 deliveries per day, see table 1. However there is a difference in the fluctuations if you compare the busy and the more quiet week. In the six most busy weeks this average was 13,1 with a variation of 3,1 deliveries and in the six most quiet weeks the average was 7,8 with a variation of 2,6 deliveries per day. This information shows that there is a big difference between the busy and quiet weeks at the OD.

Table 1: Average number of deliveries and the deviation per week and per day

	6 most quiet weeks	2016-2019	6 most busiest weeks
Average per day	7,8	10,1	13,1
Variation per day	2,6	3,1	3,1
Average per week	54,4	70,8	91,6
Variation per week	3,7	8,8	8,1

If we zoom out and take a look on a tactical level we see an average of 70,8 deliveries per week with a standard deviation of 8,8. Furthermore, there is a big difference between the busiest and most quiet weeks which have an average of 54,4 and 91,6 respectively. Over the weeks there is a big difference in the number of childbirths. Since there is a significant difference in fluctuation on a tactical level and since there is not much flexibility on operational level, it would be important to shift up or shift down on the personal planning over the weeks so that enough care for women at the OD is always provided.

The amount of staff is not the problem at the obstetrics department, according to the head of the OD. Besides, there has been an internal research of Isala concluding that the amount of FTE at the OD is not the bottleneck, which is surprising since there is a personnel shortage in the healthcare sector in the Netherlands (V&VN, 2017). The current amount of FTE at the department is sufficient, but the personnel needs to be planned smarter and more efficient.

The preferred situation is a basic personnel planning with a layer of flexible employees in order to manage the busy week. So for the peak hours the OD is scaling-up with personnel that are able to work extra hours.

In order to tackle the problem, the model is developed on a tactical level. However, there are some changes worth mentioning on a strategic level that influence the number of deliveries:

- There is a trend going on that more women are using medication during the delivery (Medisch contact, 2016). The increase is caused by the desire of being in control and having a painless delivery by the new generation of pregnant women. As a consequence of the medication the woman is automatically labelled as inpatient instead of outpatient and thus the hospital is obligated to provide an obstetric nurse. On the long term there is an expansion of the amount of women who come to the hospital and need an obstetric nurse.

3 Literature research

In order to develop a model that suits the problem at the OD, a theoretical review is executed about the definition of forecasting and whether there is an existing forecasting model available for this situation. Thereafter a literature research is done about what factors can predict or influence the number of deliveries and what kind of models could be derived from this research about birth fluctuations.

3.1 Forecasting

Forecasting is the activity of judging what is likely to happen in the future, based on the available information, as stated in the Cambridge dictionary (n. d.). (Dictionary). The forecast has to deal with the uncertainty in the future, relying mainly on data from the past and present and analysis of trends (Business Dictionary, n.d.). Forecast techniques can either be quantitative, based on (historical) data and statistical modelling, or qualitative, based on experiences and instincts from e.g. experts. The more accurate the forecast method, the more reliable the outcome. Forecasting is used in a lot of different fields. The best-known forecast is the daily weather. Within an organisation forecasting is used for sales, production, inventory, personnel, etc.(Sam Ashe-Edmunds, 2018). Within a society it is of important matter as well to do forecast about e.g. the population and demographic variables. Every situation that needs forecasting is different and there is no standard guideline available that suggests what forecast model should be used (Chambers, et al., 1971). When a forecast method is developed, it is important to keep track of the context, the relevance and availability of historical data, the degree of accuracy desirable and the time period to forecast.

3.1.1 Forecasting in a hospital

Forecasting in a hospital is a valuable tool for predicting future health events or situations such as demands for health services and healthcare needs. Forecasting enables the provided health service to minimize risk and manage demand (Soyiri & Reipath, 2012). Forecasting to determine future health situations involves a degree of uncertainty. It is impossible to have a forecast that is 100% accurate. Therefore it is important to validate the model to determine the value of the prediction.

At this moment most healthcare forecasting studies concern certain specific patient groups. The research that has been executed about forecasting the number of patients is focussed on emergency patients. Research or information about prediction models for the number of patients coming to the OD could not be found.

3.2 Birth forecasting

This section describes the different variables that might predict the number of deliveries over time.

3.2.1 Birth fluctuations

The 'Centraal Bureau voor de Statistieken' (CBS) has published an interesting article from Karin Haandrikman, called 'Seizoensfluctuaties in geboorten: veranderende patronen door planning?", about the seasonal fluctuations in the number of deliveries in the Netherlands in the past century. The article concludes that there is a clear and stable fluctuation in the number of deliveries per season over the years. In the period of 1950-1970 the peaks in the number of childbirths were in the beginning of the year and during springtime. These peaks shifted in the 1980's from the springtime towards the summer, specifically the month September. All the other months have a stable amount of childbirths (CBS, 2004).

There is no research concerning the prediction of the number of patients in the OD or the seasonal fluctuation that the number of patients in the OD might has. If this is the case, there is a connection between the total number of deliveries and the number of deliveries at the OD on a tactical basis. However, there is relationship between the number of childbirths at the OD and the total number of deliveries on an operational level. According to the CBS most of the children are born between Tuesday and Friday (CBS, 2004). Compared to the OD in Isala, which also has its peak moments between Wednesday till Friday, some type of relationship can be found. Assuming that such a relationship exists on a tactical level, it is interesting to investigate seasonality at the OD on a tactical level. This periodically distributed model is explained in chapter 3.3.

3.2.2 Events that influence the number of pregnant women

There are other variable that might influence the number of childbirths as well. For example there are some famous, but scientifically unproven stories about events that cause a higher birth rate. Examples of these stories are a higher birth rate at full moon (Laeven ,2010), 9 months after a power failure (Remmers, 2017) or after carnival.

An event that is proven to have caused a higher birth rate is the end of the Second World War. All children born between 1946 and 1955 are part of the 'babyboom' generation. In this time period 2,4 million children have been born. This was an enormous peak and caused several issues. The peak has caused overfull classes at the primary schools in the 1950's and overfull classes at the high schools in the 1960's. In the 1970's these people needed a house to live in, so a lot of infrastructure and building projects have been started. Even nowadays this peak influences the age of retirement. (CBS, 2012).

For the past decades a lot of researchers have tried to find some variables that can predict the number of deliveries. Some researchers tried to find a variable that could explain the seasonal fluctuation, like the temperature, rain and clouds. Other researchers analysed it with some demographical variables, like the number of weddings, social-economic variables, or with cultural, biological or combinations of some variables. All studies have tried to compute the number of deliveries based on an event or variable. Until now, there is no research that has found a ground-breaking variable or method that could predict the number of deliveries (CBS, 2004).

The variables in the research mentioned above could be variables that affect that number of births. Since it is not proven yet that specific variables influence the birth rate, it shows that the situation is complex and the number of deliveries depends on a lot of factors.

This research concerns the prediction of the total number of deliveries at the OD. It is not necessary to execute research with these kind of variables since a dataset of all the pregnant women in the region of Isala is available. This data will be explained in Section 4.1 data.

3.3 Possible forecasting methods obstetrics department

This section describes the prediction models that might be suitable for the situation at the obstetrics department and are based on the literature research about giving birth.

3.3.1 Periodically distributed model

The periodically distributed model assumes that the demand, so the number of patients, has a seasonal factor. The model is constructed out of a systematic component and a random component. The systematic component is the expected value of the demand and the random component deviates from the systematic component. The forecast is based on the data of the past. So in fact the historical data, thus the observed data, is also based on a systematic and a random component.

The model contains three aspects: A level, a trend and a seasonal factor. There are three different ways to calculate the systematic component with these three factors: A multiplicative, an additive and a mixed model. The most common model, the mixed model, is assumed. The level is the current deseasonalized demand. It is the starting point of the model. From this starting point there is steady growth or decline in demand per time period. This results in 'Level + trend'. The seasonal factor is multiplied with this level and trend:

*Systematic component = (Level + Trend) * seasonal factor*

This model depends on the period, so this should also be included in order to make a forecast model:

$$F_t = (L + (t^*T)) * S_t$$

Where, F_t = *forecasted number of patients at period t*
L = *estimated level of number of patients at t = 0*
t = *number of period*
T = *the trend (so the growth or decline per period)*
S_t = *seasonal factor for period t.*

The random component is computed as follows:

$$\text{Random component} = F_t - D_t$$

Where, D_t = *real number of patients at period t.*

The lower the random component, the more accurate the model is. If the random component is really large, it might refer to unexpected distortions of existing or anticipated trends.

The level, trend and seasonal factor can all be computed with the data in excel. However it is possible to already say something about the seasonal factors and the trend.

As described in section 3.2.1 a clear fluctuation can be seen in the birth rate of the Netherlands. If the number of patients that give birth at the OD is comparable with the fluctuation of the birth rate, the seasonal factors should also be very obvious in the model.

A research from the National Center for Biotechnology Information (NCBI, 1998) has investigated the future workforce for obstetricians and gynaecologists based on trends in patient demographics and care patterns. The output of the research states that there is a slow to no growth workload for obstetric personnel. Therefore I expect the trend in the model to be really small to zero.

3.3.2 Convolution model

The convolution model predicts the number of patients based on the population of pregnant women and by using statistics. This model assumes that historical data cannot influence the events, or number of patients, in the future. So the number of patients does not depend on the past.

There is data available about the women that are pregnant at the moment and what their expected due date is. This expected date should be compared with the actual delivery in order to create a probability distribution that states the chances of giving birth in a specific week. A woman is pregnant for 40 weeks on average, but it is expected that most women that give birth in the hospital give birth earlier than 40 weeks, according to a supervising obstetric nurses, due to difficulties, stress or medications during the pregnancy or delivery.

The result is a probability distribution that states the weekly chances of giving birth in e.g. week 36 till 42 of the pregnancy. If this is analysed and combined with the expected parturition, all the possible chances of all possible deliveries should be listed and used to make a forecast. The forecast per week is executed by combining all these probabilities for that week via a mathematical operation, called convolution.

Convolution is used to compute the sum of two independent variables. These variables can either be continuous function or discrete variables with the same or a with a different probability distribution. There are also applications for situations with a Poisson or a normal distribution (Fall, 2014).

At this moment convolution is used in several forecasting areas, like transport (Cheng, et al., 2018), rainfalls (Wei, et al., 2006), earthquakes (Rhoades, et al., 2011) and in hospitals as well. However, a prediction model that uses convolution in order to predict the number of patients could not be found.

Since the model computes the expected number of deliveries in that week in the hospital, every week is seen as a different sample. Let's state there is a list of all the possible women that might give birth in a specific week, the related probabilities should be combined via convolution.

Let's state that X_1 is the probability of a women to give birth in a specific week. In this case X can be seen as a discrete and independent stochastic variable:

- Discrete: X can either take the value of one or zero, because the delivery will take place in the specific week (1) or in one of the other weeks (0).
- Independent: The chance of the delivery is not dependent on the deliveries of other women.
- Stochastic: The probability distribution of a woman's delivery is stated per week.

If two discrete and independent stochastic variables should be added and you want to know what value is expected the following theory can be used (Meijer, 2016):

Recall there are two discrete stochastic variables, X and Y, with both another probability function. In the context of the research it could be said that X and Y are both women that give birth in a specific week. In order to compute the expected value of the two births the following formula is used:

$E(X + Y) = \sum_{i=0}^{1} \sum_{j=0}^{1}(i + j) * P(X = i \; en \; Y = j)$

To clarify, there are four situations possible:

- Both women do not give birth in the sample week (0+0) * P(X = 0 and Y = 0)
- Only women Y gives birth in the sample week (0+1) * P(X = 0 and Y = 1)
- Only woman X gives birth in the sample week (1+0) * P(X = 1 and Y = 0)
- Both women give birth in the sample week (1+1) * P(X = 1 and Y = 1)

Where, P(X = i and Y = j) = P(X = x) * P(Y = y) For every possible values of X and Y.

By adding those four situations the expected number of deliveries is computed.

The described theory is focused on two variables, X and Y, that can take different values. However, in this research there should be more variables that can take the value of 1, if the birth takes place in the given week, or 0, if the delivery takes place in another week.

Let's state that 'S' is the sample of all the pregnant women that might give birth in a specific week, according to the expected due date and the founded probability function. All the possible deliveries are termed: $X_1, X_2, \ldots, X_n$, where N is the total number of women that might give birth in that week. In this case the formula for the expected number of patients should be extended:

$E(X_1, X_2, \ldots X_N) = \sum_{1i=0}^{1} \sum_{2i=0}^{1} \ldots \sum_{ni=0}^{1}(1i + 2i + \ldots + ni) * P(X_1 = i \text{ and } X_2 = 2i \text{ and } \ldots \text{ and } X_N = Ni)$

The question is whether it can be assumed that the expected number is equal to the forecast. It could be interesting to show an interval of possible patients. In this case the variation should be added and subtracted to get the possible spreading.

An extra factor that should be taken into account is that only a part of the population of pregnant women gives birth in the hospital. How to deal with this issue requires a data analysis, this will be described in chapter 6, the convolution model.

3.3.3 Summarizing table

In order to summarize both models, an overview of both the periodically distributed model and the convolution model can be found in table 2.

Table 2: Key characteristics of the periodically distributed model and the convolution model.

	Periodically distributed model	**Convolution Model**
Assumes	(Historical) Seasonal fluctuations	Independent on past
Using data	Historical data of the number of deliveries at the department	Data with all the pregnant women in the region of Isala.
Based on	- Level - Trend - Seasonal factor	- Average percentage of women giving birth at Isala - Probability distribution of giving birth before and after the due date - Convolution method that computes expected number of deliveries and the most likely interval of the number of deliveries
Outcome	Forecast that is based on the systematic component.	Expected number of deliveries and the most likely interval of the number of deliveries.
Uncertainty	Random component	- Deviation in percentage error of women giving birth at Isala - Probability distribution of giving birth in each week

4 Periodically distributed model

In this chapter the periodically distributed model is elaborated. Chapter 3.3.1 describes the theory behind the model and the model is executed in this chapter in several ways. Besides, the various models are validated. The periodically distributed model is based on the Cognos data. The validation of the data is described in Appendix A.

4.1 Static planning

As described before, the OD makes use of a static planning. In order to compare the developed periodically distributed models they should always be validated and compared with the current situation. Since the department has no input for its planning the assumption is made that a static planning could be seen as planning on the average number of patients.

4.2 Development basic periodically distributed model

The development of the model starts with the Cognos data, which provides the number of deliveries per week from 2016 up and until November 2018. Besides, the data for the realized number of deliveries, according to the OD, are available on a monthly level, while the model is developed on a weekly level.

Several different models have been developed and they all are based on the same principle. Since this model assumes seasonality, the observed historical data need to be deseasonalized. In order to get the level and trend of the model, linear regression is applied to get the best line through the deseasonalized number of deliveries. With the level and trend a model based on regression can be build: Level + Trend * period number. A slightly increasing line can be seen. The season factor for every week is computed by dividing the actual number of deliveries for that week by the forecast based on regression. The next step in the development of the model is to determine the seasonal factor.

Idealistically, the model is based on a majority of the data and validated with the remaining piece of data. Since the Cognos data contain more detail than the monthly numbers of the OD, it would be good to use the data from 2016 and 2017 and to validate the model on the remaining Cognos data of 2018. Extra validation can be executed with the monthly numbers of the OD. The first model is based on 2016 and 2017 and validated with the Cognos data and the monthly number of the OD. However, the question is whether a model based on two years is reliable and valid enough since one coincidental incident could make a significant difference in the model. Therefore there is also a model developed based on all the Cognos data from 2016 till November 2018 and validated with the Cognos data itself and with the monthly number of the OD.

The seasonal factors for both models is computed by taking the average seasonal factors per week for the years the model is based on. This way the most common model is established. However, the model is not flexible for changes since the level, trend and seasonal factors have a fixed value. Therefore, two extensions for the model are investigated.

4.3 Weighted average seasonal factor

The first enlargement is a different way of computing the seasonal factor. In the former models the seasonal factor for every week is computed by taking the average seasonal factors per week of the time period on which the model is based. In this case it could be that seasonal factors are changing over time and therefore the experiment of these models is to investigate whether e.g. the seasonal factor of last year is of bigger importance and should be taken into account more than a season factor of the year(s) before. In this way it is possible to adapt to new seasonal fluctuations, like the seasonal changes between 1950-1980, as described in chapter 3.2.1 Birth Fluctuations.

There were already two models: One based on 2016 and 2017 and one based on 2016 till November 2018. Both these models can be rebuilt with this new way of computing the seasonal factor.

In the first model, the forecast of 2018 is based on data from 2016 and 2017. The question in the new model is what value should be used to multiply the seasonal factors of both 2016 and 2017. In order to get the seasonal factors used in the forecast for 2018, the variables X and Y are added and used as followed:

$$S_{\text{Week A of 2018}} = X * S_{\text{Week A of 2016}} + Y * S_{\text{Week A of 2017}}$$

Where S_A stand for the seasonal factor of week A in the specific year.

For Example, X = 0,4 and Y = 0,6 would mean that the seasonal factors of 2016 and 2017 would be respectively 40% and 60%. So the seasonal factor of 2017 is more important than the seasonal factor of 2016 in order to predict the forecast of 2018. In order to get the most appropriate values of X and Y, the solver tool in excel is used. With this tool Excel is changing the values of X and Y between 0 and 1. The initial objective of the solver tool is to minimize the MAPE in order to get the least difference between forecast and the realised number of deliveries. The same method is used for the model based on 2016 till November 2018, so with three variables, X, Y and Z.

This results of the model based on 2016-2017 are as followed:

X = 0,74 (2016)
Y = 0,26 (2017)

For the model based on 2016 till November 2018 the results are:

X = 0,154 (2016)
Y = 0,152 (2017)
Z = 0,702 (2018)

4.4 Simple exponential smoothing

Exponential smoothing is a forecasting method with no clear trend or seasonal factor. The new forecast is the old one plus an adjustment for the error that occurred in the last forecast. Thus, the new forecast is compensating in the forecasted error of the former period (Nugus, 2009). In formula:

$F_t = F_{t-1} + \alpha * E_{t-1}$

Where $E_{t-1} = F_{t-1} - D_{t-1}$

E_t = forecast error at period t

F_t = forecasted number of deliveries in period t

D_t = realized number of deliveries in period t

α = smoothing factor between 0 and 1.

If the errors in the forecast are due to random fluctuation, the model should not adapt to the random fluctuation and therefore the value of alpha tends to 0. If the errors in the forecast are due to a shift in the level of the forecast, the model should adapt to the switched level and therefore the value of alpha should be around 1. Exponential Smoothing has been executed but does not have an added value since the forecast is dependent on the forecast error of the last period and the OD is not able to update the realized number of deliveries every week. Appendix C shows the forecasts using exponential smoothing for different values of alpha.

The first figure in appendix C shows a horizontal line. For the forecast of the first week in 2016 the average number of deliveries is taken. The forecast for the second week is computed with the forecast of the first week plus the value of alpha times the forecasted error. Since the value of alpha is zero, there is no compensation for the difference between forecast and error of the former period. Therefore it could be stated that the difference between forecast and realized number of deliveries is fully based on coincidence and there is no need to adapt the forecast for the next period. The second figure of appendix C with an value of alpha of 0,5 is not static. The forecast for the next period is based on the current forecast plus 50% of the forecasted error of the current period. Therefore the error in the forecast is partly compensated. The third graph shows that the forecasted error is fully included in the forecast for the next period. This way it could be stated that the error in the forecast is absolutely not based on coincidence and fully based on a structural change.

As described in the paragraph 3.3.1 Periodically distributed model, $F_t - D_t$ is the random component in the periodically distributed model. The question is whether exponential smoothing can be used in order to compensate for the random component. If the value of alpha is unequally to zero, it would indicate that the random component is decreasing and thus there would be a change in the model '(level + trend) * seasonal factor' and that should be compensated.

Since the data for the number of deliveries are not available every week, this model is not applicable. However, what if exponential smoothing is not based on the last period, but the same period a year ago? In that case it could be stated that the season factor a year ago did not fulfil a 100% accurate forecast and that exponential smoothing is compensating for the forecasted error. By focussing on exponential smoothing on the period a year ago, a new series of forecasts can be created. The value of alpha is again between 0 and 1. A value of 0 indicates that changes in the seasonal factor are due to randomness and a value of 1 means that there are structural changes in the seasonality.

There are different ways for the mathematical elaboration of applying exponential smoothing with a periodically distributed model:
First possibility:

$F_{t \mid Y} = F_{t \mid Y-1} + \alpha * E_{t \mid Y-1}$

In this formula t | Y stands for period t in year Y.

The mathematical elaboration would be a periodically distributed forecast for 2016 and the following years would be computed with the formula above.

Second possibility:

$F_{t \mid Y} = P_{t \mid Y} + \alpha * E_{t \mid Y-1}$

where $F_{t \mid Y}$ = Forecast of period t in year Y
$P_{t \mid Y}$ = Periodically distributed forecast of period t in year Y
α = Smoothing factor between 0 and 1
$E_{t \mid Y-1}$ = Forecast error of period t in the year before

The elaboration of this possibility is different than the first possibility since this method is compensating for the forecast error on top of the periodically distributed model and not on forecast of the year before.

Both possibilities started with a periodically distributed model. The addition of exponential smoothing makes the model more flexible for changes in seasonality.

In order to find the optimal value of alpha the solver tool is used. The solver tool analyses the different values of alpha in order to minimize the MSE. For both possibilities, regardless of the used periodically distributed model, the value of alpha results in a value of zero. Several other experiments with the solver tool, like minimizing the Bias, MAD or MAPE, result in a value of alpha of zero.

4.4.1 Example to concretize the theory

What is the forecast for week 25 in 2019?

Assume:
Alpha = 0,5
Periodically distributed forecast for week 25 in 2018 = 65 deliveries
Realized number of deliveries in week 25 of 2018 = 75 deliveries
Periodically distributed forecast for week 25 in 2019 = 67

First possibility:

$F_{t|Y} = F_{t|Y-1} + \alpha * E_{t|Y-1} = \quad 65 + 0{,}5 * (75 - 65) = 65 + 0{,}5 * 10 = 70$

In this possibility the forecast for week 25 in 2019 would be 67.

Second possibility:

$F_{t|Y} = P_{t|Y} + \alpha * E_{t|Y-1} = \quad 67 + 0{,}5 * (75 - 65) = 67 + 0{,}5 * 10 = 72$

In this case the forecast is increased from 67 to 72.

Conclusion: In the first possibility exponential smoothing is based on the forecast error of the previous year* and in the second possibility exponential smoothing is compensating for an error that occurred in the forecast of the periodically distributed model a year ago.

*And the forecast of the first year of the model is computed with a periodically distributed model.

4.5 Validation of periodically distributed models

In the first paragraph the basic periodically distributed model and the periodically distributed model with a weighted seasonal factor are validated and compared with the static planning. The second paragraph makes the conclusion of a periodically distributed model with exponential smoothing

4.5.1 Periodically distributed models

The graphs that represent the forecast for the number of deliveries per week and the static planning compared with the actual number of deliveries between 2016 and November 2018 can be found in appendix D to G. In order to make the validation measurable several KPI's, Key Performance Indicators, are chosen:

- Bias: The bias shows a statistical difference and indicates if there is a structural error in the forecast.
- MAD: The Mean Absolute Deviation shows the average of the absolute deviations. It is a KPI to measure the variability.
- Percentage Error: This shows the relative difference of the forecast for every week. It might be interesting to count the number of percentage errors of a model above a certain value, since it could occur that the average of the percentage errors is quite acceptable, while there are a lot of unpredicted peaks. The purpose of the research is indeed to forecast these peaks.
  As agreed with the improvement coach, a percentage error beneath the 10% is assumed preferable and a percentage error above the 15% is assumed as a possible risk. Therefore the number of percentage errors above 10% and 15% are counted. A forecast can never be 100% accurate and the forecast will sometimes be insufficient but it is an important KPI to see how often the forecast would be insufficient.

- MAPE: Mean Average Percentage Error shows the relative difference of all the time periods.
- MSE: Mean Squared Error: It measures the average of the squares of the errors. It measures the quality of the forecast. The lower the value, the more accurate the forecast is.
- TS: Tracking Signal is computed by dividing the Bias by the MAD. It provides a certain value which can be used to determine if there is over forecasting or under forecasting in the model. A value above 6 would indicate over forecasting and a value beneath -6 indicates under forecasting.

These KPI's are used in order to validate the different periodically distributed models. Before that, the current static planning is validated, assuming the forecast is the average number of deliveries per week, see table 3.

Table 3: Validation static model

Validation	Bias	MAD	MSE	MAPE	Percentage error >15%	Percentage error > 10%	TS
Static planning	0	6,80	76,08	9,71	41	60	0,00

Due to the assumption that a static planning means a forecast on the average number of deliveries, the bias is of obviously zero. The value of the Tracking is zero as well since there is both under and over forecasting, which neutralize each other.

The fact that 60 weeks, which is 41,7% of the total weeks in the Cognos data, have a percentage error above 10% means that the static planning is really insufficient.

The realized numbers of the OD about the number of deliveries are only available on a monthly level. Therefore the forecasts per week are converted to a monthly level.

For example: March 2018 exist of: 4 days of week 9 + Week 10, 11, 12 + 6 days week 13
This results in: Forecast week 9 divided by 7 times 4 + forecast for week 10 till 12 + forecast week 13 divided by 7 times 4.

This way of converting is not very accurate since it assumes that the deliveries are equally distributed over the week. In reality this is not the case because there are more deliveries between Wednesday and Friday. Besides, since several weekly forecast are merged into one forecast for a month it could be that the over forecast of one week is neutralized with the under forecast of the other weeks. Therefore the validation on a monthly level is less accurate.

The different models are based on different values and for this reason there are different possibilities to validate the data. All the models can be validated with the available monthly data of the OD. Besides the models based on 2016 and 2017 can be validated with the Cognos data of 2018. And last, all the KPI's can be tracked during the development process and thus validated with all the data of Cognos. Idealistically the validation is executed on different data than the model is built upon, however, it is still a suitable validation method for periodically distributed model. A clear overview of this can be found in table 4.

Table 4: Ways to validate the different models.

Model	Model based on	Seasonality	Validation with
Periodic	2016 - 2017	Average seasonal factor of 2016 and 2017	1. Monthly number of OD 2. Data of 2018 of Cognos 3. All Cognos data
Periodic	2016 - 2017	Weighted average of various seasonal factors*	1. Monthly number of OD 2. Data of 2018 of Cognos 3. All Cognos data
Periodic	2016 - 2017-Nov. 2018	Average seasonal factor of 2016, 2017 and 2018	1. Monthly number of OD 2. All Cognos data
Periodic	2016 - 2017-Nov. 2018	Weighted average of various seasonal factors*	1. Monthly number of OD 2. All Cognos data

Table 5 on page 30 shows a comprehensive overview of the outcome of the different KPI's per model per possible data validation. All the KPI's have a colour:

- Red Value of KPI is considered sufficient compared to the values of the other models.
- Orange: Value of KPI is comparable with the values of the other model.
- Green: Value of KPI is considered insufficient compared to the values of the other models.

In paragraph 1.5.1.1 till 1.5.1.4 these KPI's are used to validate the developed models.

Table 5: Validation of the different models with the different data.

	KPI	Cognos		Cognos - 2018		2019
	PE>..%	Aantal	percentage	Aantal	Percentage	Aantal
Based on 2016 - 2017 average seasonal factor	Boven 10	38	26,4%	29	72,5%	0
	Boven 15	13	9,0%	4	10,0%	0
	MAPE	7,60%		11,6%		4,9%
	Bias	-46,5		-47,0		18,9
	MAD	5,3		7,5		14,5
	MSE	50,3		10,2		397,7
	TS	-8,7		8,8		1,3
	KPI	Cognos		Cognos - 2018		2019
	PE>..%	Aantal	percentage	Aantal	Percentage	Aantal
Based on 2016 - 2017 Weighted average seasonal factor	Boven 10	41	28,5%	18	45,0%	0
	Boven 15	17	11,8%	1	2,5%	0
	MAPE	7,50%		10,4%		5,1%
	Bias	9,7		26,8		-71,1
	MAD	5,2		7,1		15,5
	MSE	55,6		10,3		404,2
	TS	1,85		3,8		-4,6
	KPI	Cognos		Cognos - 2018		2019
	PE>..%	Aantal	percentage	Aantal	Percentage	Aantal
Based on 2016 - November 2018 average seasonal factor	Boven 10	38	26,4%	4	10,0%	0
	Boven 15	11	7,6%	1	2,5%	0
	MAPE	7,20%		7,7%		4,8%
	Bias	-0,1		-31,3		21,8
	MAD	5,0		5,0		14,5
	MSE	40,3		6,8		331,9
	TS	-0,02		-0,4		1,5
	KPI	Cognos		Cognos - 2018		2019
	PE>..%	Aantal	percentage	Aantal	Percentage	Aantal
Based on 2016 - November 2018 Weighted average seasonal factor	Boven 10	40	27,8%	0	0,0%	0
	Boven 15	21	14,6%	0	0,0%	0
	MAPE	7,60%		3,6%		4,9%
	Bias	119,3		8,8		55,5
	MAD	5,2		2,3		14,6
	MSE	53,0		3,1		301,4
	TS	22,95		4,5		3,8

4.5.1.1 Based on 2016 – November 2018 + weighted average seasonal factor

The KPI's for this model computed with the Cognos data of 2018 are all sufficient. This can be solved with the results of X, Y and Z, found by the solver:

X = 0,154 (2016)
Y = 0,152 (2017)
Z = 0,702 (2018)

The weight of the seasonal factors of 2018 are relatively high and therefore it is obvious that the KPI's for 2018 are decent. However, the KPI's measuring the performance over the whole Cognos data as well as the data for 2019 represent a less accurate model compared with the other models. Therefore this model is not sufficient.

4.5.1.2 Based on 2016 – 2017 + weighted average seasonal factor

The results for the weighted values of X and Y are not in line with the hypothesis. Since the values of X (2016) and Y (2017) are respectively 0,74 and 0,26, the seasonal factors of 2016 are considered more important than the seasonal factors of 2017, while it would be more logic that the latest year is considered as more important in order to have incremental seasonal changes.

Besides, the model does not have KPI's that are considered better than the other models, while the two remaining models do have a significantly better KPI's.

4.5.1.3 Based on 2016 – 2017 + average seasonal factor

By counting the percentage errors above 15% and 10% based on the Cognos data, this model decreased both numbers with respectively 68,2% and 36,7% to the values of 13 and 38. However, the TS measuring the Cognos data shows there is a structural over forecast. This over forecast can not be seen in 2019.

The value of the Means Square Error concerning every dataset is significantly higher than the model based on 2016 – November 2018.

For this model, it is hard to compare the KPI's for only the Cognos data of 2018 with the model that is based on 2016 till November 2018 with an average seasonal factor, since this second model has better KPI's because it is partly based on data from 2018.

4.5.1.4 Based on 2016 – November 2018 + average seasonal factor

The tracking Signals of this model do not show structural over or under forecast. The KPI's for the percentage errors for this model are lower than the percentage errors of the model based on 2016 and 2017. The MSE and MAD are significantly better than those of the model based on 2016 and 2017 as well. The other KPI's are almost equal and do not make a major difference.

As stated before, this model is partly based on the data from 2018 and therefore it is hard to compare both models with an average seasonal factor.

4.5.2 Conclusion

Since the KPI's of the remaining two models do not make a significant difference, the question is what model is most suitable.

Therefore two other aspects are taken into consideration. The first aspect is by making a graph of the forecast. These graphs can be found in appendix D, F and H.

In appendix D, the model based on 2016 - 2017 with an average seasonal factor, the forecast of 2016 and 2017 seems to have the same seasonal pattern as the actual number of deliveries. However in 2018 the fluctuations can differ from comparable to different. At some moments the forecast and the real number of deliveries are both increasing or decreasing and on other moments both lines have different peaks. Besides there is a high peak in week 38 in 2018. This peak is in 2016 and 2017 a week earlier, in week 37 and therefore this peak is not predicted in 2018. Furthermore, the peak is extremely high in 2018. It has a percentage error of 32%. This peak could be identified as the Christmas and New Year's Evening peak.

Comparing the graphs of the model based on 2016 – November 2018, displayed in appendix F, with the model based on 2016 – 2017, the forecasts for 2016 and 2017 appear to be less accurate. However, the results of 2018 are more similar to the realized number of deliveries and the peak in week 37 and 38 has a more realistic prognose to predict the Christmas and New Year's Evening peak.

The Christmas and New year's Evening peak take place either in week 37 or week 38. Therefore the model based on 2016 - November 2018 fit better than the model based on 2016 – 2017.

According to appendix H, the graphs of both models seem similar. However, table 6 shows that on a weekly level there are some differences, although these differences are small in most of the weeks. Nevertheless, week 38 shows a big difference between both forecasts.

Table 6 – Forecast for week 30 till 40 in 2019.

Period	Based on		Difference
	2016-2017	2016 - November 2018	
2019|30	78,4	78,6	0,2
2019|31	78,4	81,3	-2,9
2019|32	82,5	77,0	5,6
2019|33	71,2	72,1	-0,9
2019|34	80,5	78,7	1,9
2019|35	61,9	65,6	-3,7
2019|36	82,6	80,0	2,6
2019|37	83,1	82,0	1,0
2019|38	72,7	84,3	-11,5
2019|39	90,3	90,2	0,0
2019|40	77,9	78,6	-0,7

The last aspect that could make a difference is that the model based on 2016 - November 2018 is based on more data than the model based on 2016 - 2017 and therefore this model is considered as more reliable.

Taking all the aspects into account: the KPI's, the graphs, the Christmas and New Year's Evening peak and the amount of data used to develop the model, the model based on 2016 - November 2018 with an average seasonal factor is most suitable.

4.5.3 Conclusion of the addition of exponential smoothing

Since all the values of alpha resulted in zero, the exponential smoothing on top of a periodically distributed model has no added value because there is no compensation for the forecast error. It shows that the forecast error is based on randomness and not based on structural changes in seasonality.

It confirms that a periodically distributed model is suited for the situation at the department and that there is no need to diverge from the periodically forecast.

5 Convolution model

The convolution model is based on the theory described in chapter 3.3.2. The model is based on two aspects:

1. Average percentage of women giving birth at the OD
2. Distribution of giving birth e.g. in week 36 till 42 of the pregnancy

5.1 Analysis

The first aspect is analysed in appendix A. The average percentage of women giving birth at Isala is 83% with a standard deviation of 6,8%.

Assuming the 428 pregnant women with a parturition date in October could result in different number of deliveries according to the average and standard deviation, see table 7.

Table 7 – percentage and deviation of women giving birth at the OD results in #deliveriers.

Percentage women giving birth at the OD	(μ - σ) 76,2%	(μ) 83%	(μ + σ) 94%
Number of deliveries related to average percentage women giving birth at OD	326,1	355	384,3

Both the standard deviation and the table shows that there is a large variability in the average percentage of women giving birth in the hospital. This variability can result in a difference of 384,3 – 326,1 = 58,2 deliveries per month.

After predicting how many women are going to give birth in the hospital, The forecast of the week these women will give birth in should be made. This forecast is made with the due date of the pregnancy and the distribution of giving birth in e.g. week 36 till 42 of the pregnancy.

The distribution of giving birth before and after the due date is shown in figure 2 and is investigated by an internal research of Isala. Most of the women giving birth at Isala give birth 14 days before their parturition date (Isala, 2018), which is approximately 23% of the women. By analysing all the due dates and the dates of giving birth, the distribution of giving birth in each week of the pregnancy can be determined. This research is executed on a daily basis, but it can be transferred to a weekly level, which could be useful for this research. Also this factor includes some uncertainty.

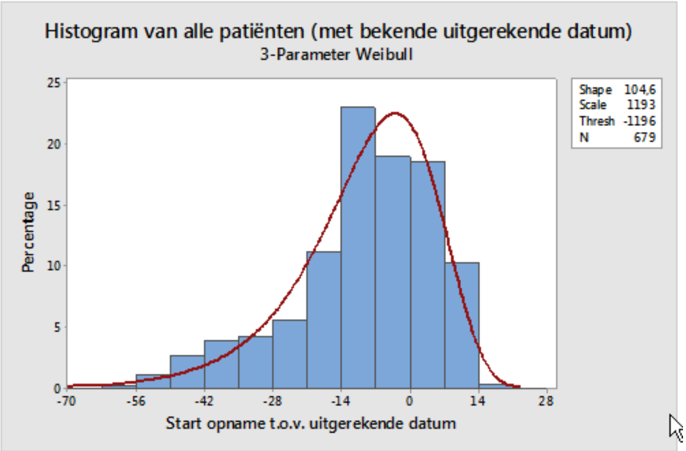


Figure 2: Distribution function giving birth per week - Bron Isala.

As described in chapter 3.3.2, the outcome of this model is a boxplot that contains:

- The average expected number of deliveries for that week
- The interval for the number of deliveries for that week. These intervals could be formulated as followed:
  - [μ - σ ; μ + σ], which shows the 68% interval
  - [μ - 2σ ; μ + 2σ], which shows the 95% interval
  - [μ - 3σ ; μ + 3σ], which shows the 99,7% interval

The choice which interval is most suitable, would be deliberated with the obstetrics department. Since the employees of the obstetrics department have no background in statistics or mathematics this discussion is only interesting if the results are available in an understandable way.

5.2 Conclusion

There are different uncertainty factors in this model. The uncertainty in the average percentage of women giving birth in the hospital can make a difference of 58,2 deliveries per month, which is equal to 13,1 deliveries per week. Besides, the uncertainty for giving birth in each week needs to be added to this uncertainty. Comparing this with the variability in the actual number of deliveries, as stated in table 1 on page 17, which is 8,8 deliveries per week, the uncertainty in this model exceeds the variability in the actual number of deliveries. In this research the choice is made to stop developing this model and to focus on the periodically distributed model.

6 Conclusion, recommendations & discussion

The research results in the conclusion described in section 6.1 and the recommendations can be found in section 6.2. The discussion is stated in section 6.3. An overview of the characteristics of both the periodically distributed model and the convolution model can be found in table 2 on page 22.

6.1 Conclusion

The obstetrics department has the problem of not always being able to provide enough and sufficient care for its patients due to the use of a static planning and large fluctuations in the number of patients. Besides, the misbalance leads to less employee satisfaction and less flexibility.

There is a static planning because the department has no input for the static planning. The core problem is the need for a model that can predict the number of deliveries at the department for the upcoming week in order to scale up or down on the personnel planning. So the purpose of the research is to develop such a model.

Based on the literature two possible models are analysed. The convolution model is considered as unreliable since the uncertainty is considered too large due to the large variability in both the average number of women giving birth at Isala and the probability distribution of giving birth per week of the pregnancy.

The periodically distributed model results in a reliable forecast. Due to the KPI's, the graphs and the amount of data the model is based on, the model based on 2016 – November 2018 with an average computed seasonal factor is considered most reliable. So in this case, the forecast for 2019 is based on the average seasonal factors of 2016 up and until November 2018.

The periodically distributed model is based on the Cognos data. As described in appendix A there is an external change, which causes more women going to Isala. Appendix B is describes that the forecast should be 1,4% higher in order to compensate for an external factor.

With this model the forecast for the number of deliveries per week can be made for every week of 2019 and further on. The forecast for every week in 2019 can be found in appendix I.

The research has accomplished its goal to develop a model that predicts the number of deliveries for the upcoming week. With this model the core problem is solved and on its turn, it can solve the action problem.

6.2 Recommendations

6.2.1 Static planning

The research accomplished its goal to develop a model that predicts the number of deliveries. This model should be used as an input for the personnel planning. The forecast can be found in appendix I – Forecast 2019. The head and planners of the obstetrics department should investigate how the forecast will affect the personnel planning. Therefore the question arises when the department should shift up or down on its personnel planning.

6.2.2 Developed model

A template of the developed model will be sent to the obstetrics department, including the instructions. The instructions can be found in appendix J and is written in Dutch in order to serve the employees of the obstetrics department. Screenshots of the developed model can be found in appendix K.

In order to use the model, the number of deliveries for at least two years need to be filled in as an input. In order to get the regression line, a simple button needs to pushed. The optimization of the value of alpha for exponential smoothing is made as easy as possible and the instructions for this are described step by step. In step four a factor can be added in order to increase or decrease the forecast. This could be used if there should be compensated for external changes, like the situation described in appendix A and analysed in appendix B. On the last page of the model, the forecast for the number of deliveries for the next few years can be seen.

In the next section a recommendation for the collection of data is stated. Due to the transfer to HiX new data should be collected. When the HiX data is available, the template should be used to make a new forecast.

The exponential smoothing part will automatically compensate for changes in seasonality. However, if there are too many changes in seasonality, the forecast will be less accurate and it could be that the exponential smoothing part is not compensating enough. If this situation occurs, the data of the last two years should be used as an input for a the next periodically distributed model. So the template should be used again.

6.2.3 Data

The first recommendation is about the data that Isala uses. Apart from the fact that the Cognos data included data from other departments, the data from the obstetrics department was complex. Neither the people at the obstetrics department, nor the improvement coaches and other data specialist could tell what every label, category, code, type, etc. meant. The goal is to use the data and not to collect the data.

In the new system, called HiX, there is a large increase in the data of records without a label, category or something that indicates what the record means. It takes the business intelligence employees in the hospital a lot of effort to make the new system work sufficiently. However, I would advise the hospital to only keep track of useful data. For the logistical part the following things are important:

- Inpatient / outpatient
- A label that indicates a natural delivery, a delivery with medication, a caesarean, or a primed delivery.
- Function types of staff helping at the delivery.
- Room number of the delivery

Keeping track of these data will help the obstetrics department with possible research about the logistical part of the department in the future.

The obstetrics employees should think about what is important and useful to measure and to keep track of, from both a logistical as a health-care point of view.

6.2.4 Christmas and New Year's evening peak

The obstetrics department should be aware that the forecasts for week 38 and 39 have a deviating pattern. The model is based on 2016 up and until 2018 and in these years Christmas and New Year's evening fall in different weeks. Therefore there is no clear pattern that can predict these peaks.

6.2.5 Inpatient <-> outpatient

In this research there is no distinction made between inpatient and outpatient. The developed periodically distributed model predicts the total number of deliveries at the hospital. It provides a good indication for when the peaks take place, but it is not stated how many deliveries are inpatient and outpatient, while both deliveries use different employees. For the obstetrics department that is not that important since they treat the inpatient deliveries and facilitate the outpatient deliveries at the department. However for the personnel planning it could make a difference.

6.2.6 Convolution

The convolution model can still be interesting. With the current available data, the variability of the average percentage error is considered to large compared with the variability at the department. In the future, if more KCL data, the data with the pregnant women, is available, a more accurate analysis can be executed. Besides, it could be that there are seasonal patterns in the percentages of women giving birth at the hospital.

6.3 Discussion about using exponential smoothing

The counterpart of using the exponential smoothing on top of a periodically distributed model is that exponential smoothing is normally based on the former period and not on the same period in the previous year. The fact that the exponential smoothing in this way is based on a year ago could indicate some kind of seasonality. As stated in the beginning of the section about exponential smoothing: "Exponential smoothing is a forecasting method with no clear trend or seasonal factor". Therefore it could be concluded that exponential smoothing on top of a periodically distributed model is not suitable.

On the other hand it could be seen as a new creation of a series of numbers. Normally, the exponential smoothing method is built upon consecutive periods. This time the exponential smoothing is built upon a week number of the different years. In this way exponential smoothing is investigating if there is randomness or if there is a structural change over the years in seasonality. It sounds fair to build an exponential smoothing model on top of a periodically distributed model and with an optimal value of alpha of zero, the basic periodically distributed model is suitable.

Summarizing, the exponential smoothing on top of a periodically distributed system is not an added value and it confirms the accuracy of the periodically distributed model.

7 Sources

Ash-Edmunds, S. (2018). The Difference Between Qualitative and Quantitative Forecasting Techniques. Retrieved from https://bizfluent.com/info-12042327-differences-between-qualitative-quantitative-forecasting-techniques.html

Cambridge dictionary (n.d.). Forecasting. Retrieved from https://dictionary.cambridge.org/dictionary/english/forecasting

CBS (2014). Seizoensfluctuaties in geboorten: veranderende patronen door planning? Retrieved from https://www.cbs.nl/nl-nl/achtergrond/2004/50/seizoensfluctuaties-in-geboorten-veranderende-patronen-door-planning-

CBS (2014). Geboorten: Vooral midden in de week. Retrieved from https://www.cbs.nl/nl-nl/nieuws/2004/14/geboorten-vooral-midden-in-de-week

CBS (2012). Babyboomers. Indrukken vanuit de statistiek. Retreived from https://www.cbs.nl/nl-nl/publicatie/2012/17/babyboomers-indrukken-vanuit-de-statistiek

Chambers, C. J., Mullick, S. K., Smith, D. D. (1971). How to choose the right Forecasting Technique. Retrieved from https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique

Cheng, X., Zhang, R., Zhou, J., Xu, W. (2018). DeepTransport: Learning Spatial-Temporal Dependency for Traffic Condition Forecasting. Retrieved from https://ieeexplore.ieee.org/abstract/document/8489600

Fall, D. (2017). Powerpoint from Clark University. Sums and Convolution math 217 Probability and Statistics. Retrieved from https://mathcs.clarku.edu/~djoyce/ma217/convolution.pdf

Hans, E. W., Van Houdenhoven, M., Hulshof, P. J. H. (2011). A Framework for Health Care Planning and Control. Retrieved from https://pdfs.semanticscholar.org/9db5/a03a223a5940735a46f7a4b5c215fc424b3f.pdf

Heerkens, H., & Winden, A. v. (2012). Geen probleem. Buren: Business School Nederland.

Isala (2018). Verloskunde aanpak personele planning. Internal research.

Laeven, J. (2010). Published in NRC. Worden er meer baby's geboren met volle maan? Retrieved from https://www.nrc.nl/nieuws/2010/01/25/worden-er-meer-babys-geboren-met-volle-maan-11841404-a593260

Meijer, D. (2016). Kansrekening voor TBK. Study material University of Twente.

Nugus, S. (2009). Smoothing Techniques. Retrieved from https://www.sciencedirect.com/topics/economics-econometrics-and-finance/smoothing-technique

Omroep Flevoland (2018). Definitieve doorstart voor ziekenhuis Lelystad. Retreived from https://www.omroepflevoland.nl/nieuws/165822/definitieve-doorstart-voor-ziekenhuis-lelystad

Remmers, F. (2017). Published in AD. Culemborg: babyboom na stroomstoring feit of fabel? Retrieved from https://www.ad.nl/utrecht/culemborg-babyboom-na-stroomstoring-feit-of-fabel~ae699e93/

Rhoades, D. A., Schorlemmer, D., Gerstenberg, M. C., Christophersen , A., Zechar, D., Imoto, M. (2011). Efficient testing of earthquake forecasting models. Retrieved from https://link.springer.com/article/10.2478/s11600-011-0013-5

Rtv Drenthe (2018). Laatste weekend verloskunde in Bethesda: 'We zijn naar dit moment toegegroeid'. Retrieved from https://www.rtvdrenthe.nl/nieuws/139727/Laatste-weekend-verloskunde-in-Bethesda-We-zijn-naar-dit-moment-toegegroeid

Soyiri, I. N., Reidpath, D.D. (2012). An overview of health forecasting. Retrieved from https://environhealthprevmed.biomedcentral.com/articles/10.1007/s12199-012-0294-6

V&VN (2017). Personeelstekorten in de zorg | Oplossingen van de werkvloer. Retrieved from https://www.venvn.nl/Portals/1/Downloads/Personeelstekorten-zorg-oplossingen-van-de-werkvloer.pdf

Wei, C., Hung, C. W., Cheng, K. S. (2006). A multi-spectral spatial convolution approach of rainfall forecasting using weather satellite imagery. Retrieved from https://www.sciencedirect.com/science/article/pii/S0273117705010495

Witte, I., Ter Horst, N. (2016). Medicalisering komt door zwangere zelf. Retrieved from https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/medicalisering-komt-door-zwangere-zelf.htm

Appendix A – Data validation

This section describes the different data sets that have been used in the research. The datasets contained a lot of data and needed to be analysed, filtered and validated before it could be used.

Cognos data

The obstetrics department has a precise overview of the number of deliveries per month, but for the research it is important to know every delivery and its date. Therefore the Cognos data is used.

The Cognos data contain all the information from all the medical treatments in the 'vrouw-kindcentrum' from January 2016 till 2019, including the treatments at other departments. This consists of the deliveries at the Obstetric department and a lot of other hospitalizations and surgeries of women, partners and children. Therefore it is important to prepare the data, filter the right treatments and validate the remaining data. This process is described in the next paragraph.

Validation dataset

Before the data can be used the data should be prepared. Columns that do not have added value for the research are immediately deleted, like the duration of the deliveries, the location, which was Zwolle in all cases, the obstetric nurse and other columns with unnecessary data.

The next step is to delete the data of deliveries that have more than one track record. For example, a woman who came at the department had five data points for one delivery. Four of these tracks are exactly the same and were labeled as a delivery, while the other record was a specific medical treatment. In order to delete the duplicates, the record number and the label are combined in a following column and the duplicates in this combination are deleted. In a next phase of the data validation this medication record will be disregarded to make the data useful. When the duplicates are deleted only based on the column with the record number it could have occurred that the excel file deleted all the records with the delivery label and that the label which is a medical treatment remained. In that case, the delivery would not exist anymore in the new dataset. Therefore the combination between record number and label is necessary.

This way 40.218 duplicates have been deleted and 23.135 unique values remained. There are still some duplicates. The remaining duplicates have one record for the delivery and one record for medication. In the next phase the medical indication is filtered and only the delivery records remain.

In the following phase, the remaining data is processed in a pivot table. The number of patient numbers is displayed per month and filtered by all the labels that included a delivery label. Since the data is in Dutch, the labels with 'partus' or 'bevalling' are used. However the added number of patient numbers is less than the precise overview of the obstetrics department. So therefore some labels are missing because otherwise the data could not have been less than the realization at the department. However, there were no other labels that had something to do with a delivery. Nevertheless, there exist a label called 'nothing' or 'geen label' in Dutch.

The hypothesis is that some deliveries did not get a delivery label, so some deliveries should occur in the records without a label. This is checked by the improvement coach of Isala in the digital patient dossier and around 80% of the records without label at the gynecology and obstetrics department are a delivery.

Therefore the records without a label are also included in the hospital and to make sure that there were no patients without a label from another department, the filter of the gynecology and obstetrics department is added. By working this way, the given amount of deliveries per month diverged with 2% from the realization that the obstetrics department claims to be true.

This result seems to be close to reality and is therefore considered valid in order to continue. A really important note is that the data is not perfect and includes some records that are not a delivery because of the last added filter. However without this filter, the number of deliveries in the Cognos data would be too low, so some delivery records are missing. It could be said that the missing records and the extra records compensate for each other.

Before the conclusion was made to continue with the dataset, there has been contact with the head of the OD, the planners and the people who have access to the data. However, the datapoints that the head of the OD uses, could not be found or traced.

The conclusion is to continue with the described remaining data, which differs two percentage compared to the data of the OD itself.

HiX

In November 2018 the hospital switched to another ICT system that manages the administrations, dossiers and patient logistics. Since that moment there is an increase in the number of patients without a diagnose label of more than 400% and a decrease in the number of patients with a diagnose label. Since November 2018 around 50% of the data with the filters and validation described above was labelled as a delivery. For this reason the Cognos data since November 2018 is considered as unreliable and not valid.

KCL data

The KCL data is the data from the 'Klinisch Chemisch Labratorium'. This laboratory tests the blood of all the pregnant women when they reach their 12th week of pregnancy. The laboratory does this for all the women in the region of Zwolle and therefore this laboratory has an overview of all the pregnant women in the region. The KCL is part of the Isala hospital but is also active for patients from the general practitioner. Therefore the KCL is not able to provide personal information in the dataset, like names or patient numbers. For this reason it is not possible to combine this data set with the Cognos data to see which patients did and did not come to the hospital.

The dataset is used to compute the average percentage of women that give birth in the hospital compared to the whole population of pregnant women. In order to do this, all the pregnant women in the population are categorized in the month of their parturition date and the monthly numbers of deliveries according to the obstetrics department are compared with these numbers.

Table 8 shows the average percentage of women that give birth in the hospital per month in 2018.

Table 8: Percentage women that give birth in the hospital per month.

	Number of women with a parturition date in this month	Actual number of deliveries	Percentage women giving birth in hospital
January	320	268	84%
February	308	249	81%
March	329	310	94%
April	357	320	90%
May	400	322	81%
June	370	286	77%
July	374	304	81%
August	441	322	73%
September	437	341	78%
October	428	315	74%
November	378	341	90%
December	350	311	89%

The average percentage of women giving birth in the hospital in March is 94%, while this percentage is 74% in October. The average percentage of women giving birth in the hospital is 83%.

External factor

Besides the transfer to the HiX system there is a change outside of the data: At the end of October 2018 the obstetrics department of the Bethesda hospital in Hoogeveen was closed (rtv Drenthe, 2018) and the hospital in Lelystad, the MC IJselmeerziekenhuizen, is declared bankruptcy (Omroep Flevoland, 2018). The pregnant women that should have gone to Hoogeveen are now going to the obstetrics department in Zwolle. The pregnant women in the region of Lelystad are send to the hospital in Harderwijk, but the obstetrics department of Harderwijk is not always able to take care of the pregnant women due to this increase in number of deliveries. Therefore it occurred that Harderwijk is closing its doors for new arriving pregnant woman. This results in extra deliveries at the OD in Zwolle.

Both changes are causing a new flow of pregnant women to the OD of Isala in Zwolle. An estimation of the head of the OD states that approximately 250 extra deliveries will take place in Zwolle, due to the closing of the hospital of Hoogeveen. There is no research available for the situation of Lelystad. Due to the changes, mainly the region of Urk would go to the Isala clinic in Zwolle.

The question is what percentage of the deliveries is caused by these changes and how to deal with the change in the forecast model. A possibility to solve the question could be an assumption that states that the population of pregnant women that give birth in Zwolle is living outside the region of Zwolle, so Isala is not familiar with those women. Therefore the new population of pregnant women should be multiplied with a factor that compensates for the pregnant woman outside the region of Zwolle.

Appendix B – Compensating for external factor

As described in appendix A – data the closing of the Obstetrics department in Lelystad and Hoogeveen are causing more deliveries at Isala. Therefore an analysis about whether there should be a compensation in the forecast is executed.

In chapter 4 is concluded that the periodically distributed model based on 2016 – November 2018 is most suitable for the situation at Isala in Zwolle, therefore this model is the starting point of the analysis.

The external changes are occurring since November 2018. The only data that is available and reliable to use for this analysis are the numbers of the OD on a monthly level. The monthly numbers of January till March of both 2018, before the change in the data, and 2019, after the change in the data are compared in table 9. For these three months there is an average increase of 13,19%. This could be caused by:

- An increase in the number of deliveries over time
- An increase in the population of pregnant women that give birth in Zwolle due to the closing of the hospitals in Lelystad and Hoogeveen
- A coincidental peak in the number of deliveries or no structural change.

Table 9: Number of deliveries per month in 2018 and 2019.

Month	#deliveries 018	#deliveries in 2019	Increase
january	268	315	17,54%
february	249	291	16,87%
March	310	326	5,16%

In order to investigate if the increase is caused by the flows from Lelystad and Hoogeveen the periodically distributed forecasts are transferred to a monthly forecast:

The monthly formula for the forecast of January 2019 =
Forecast week 1, 2, 3, 4 + 4 days of forecast week 5

The weekly forecast already contains the increase in the number of deliveries over time and is only based on the population of pregnant women in the region of Zwolle. In order to investigate if an assumption should be made to increase the population to compensate for the closing in Lelystad and Hoogeveen the formula is changing towards:

The Monthly formula for the forecast of January 2019 =
(*Forecast week 1, 2, 3, 4 + 4 days of forecast week 5*) * *F*
Where F is the factor that indicates the increase in population

The given way of computing the monthly forecasts is also used for the other months.

This analysis is executed in excel and in order to find the optimal value of F the KPI's for Bias, MAPE, MAD and MSE in combination with the solver tool is used. Table 10 shows the optimal value of F giving the solver tool is minimizing the value of the related KPI.

Table 10: Optimal values of F giving the solver tool is minimizing the KPI.

Minimizing KPI	BIAS	MAPE	MAD	MSE
Optimal value of F	0,00	0,014	0,014	0,00

Minimizing the MAPE and MAD by using the solver, the value of F should be 0,014. This indicates that the forecast should assume a population of 1,4% bigger in order to compensate for the closing in Lelystad and Hoogeveen. However when the Bias and MSE are used to analyse the optimal value of F, it should be stated that no compensation should take place. Table 11 shows the value of the KPI's given the value of F.

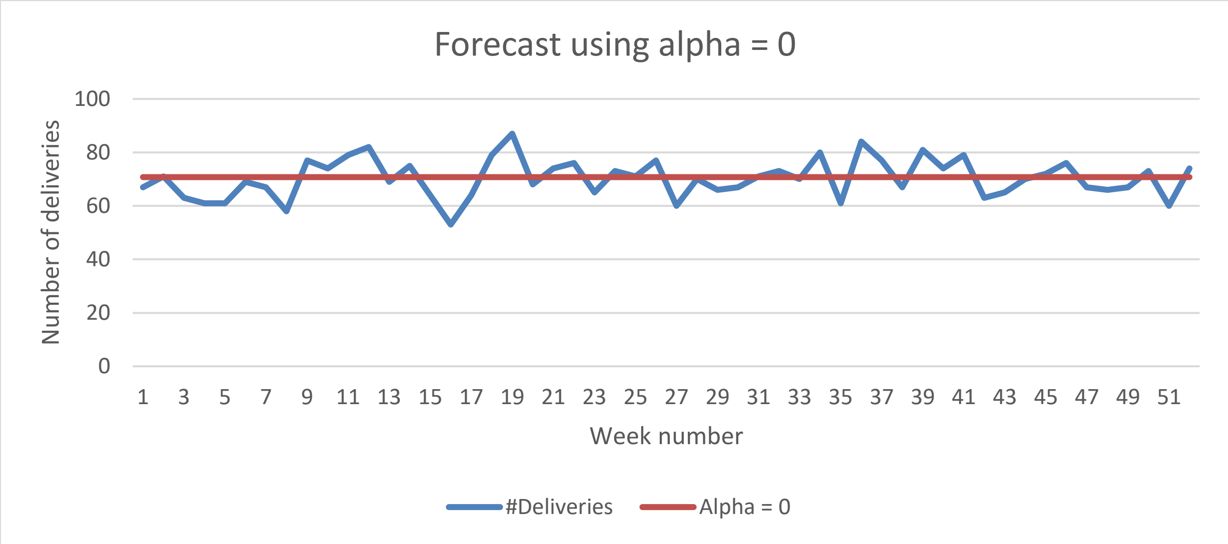
Table 11: Measurements of KPI's given the value of F.

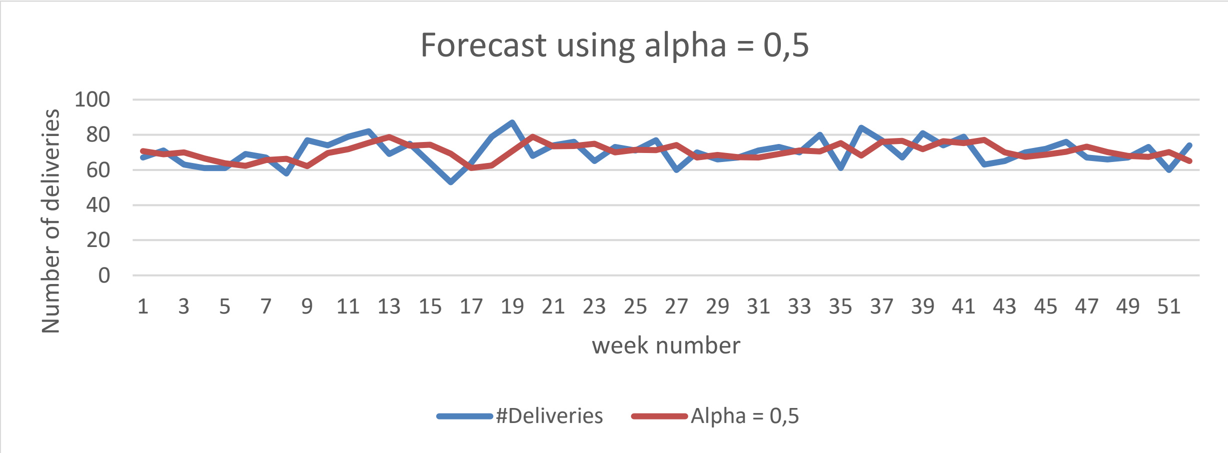
Minimzing KPI	F = 0	F = 0,0138
Bias	46,4	60,0
MAPE	9,3%	9,0%
MSE	1628,0	1840,0
MAD	29,6	28,8

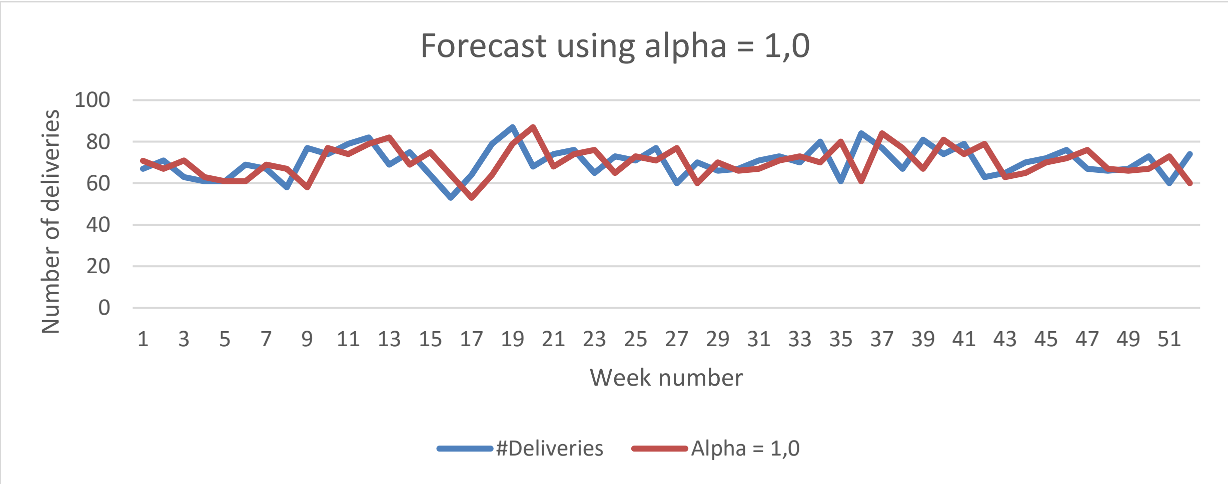
An increase in the weekly periodically distributed forecast of 1,4% is on average 1,001 extra delivery per week. For 2019 the highest increase in the number of deliveries would be an increase of 1,25 deliveries in week 39 and the lowest increase would be an increase of 0,84 in week 8.

The forecast should be able to predict the peaks in the number of deliveries and therefore the percentage error and the deviation are considered more important than the Bias and MSE. For this reason the compensation of 1,4% is of added value and added to the forecast for the obstetrics department of Isala.

Appendix C – Exponential Smoothing

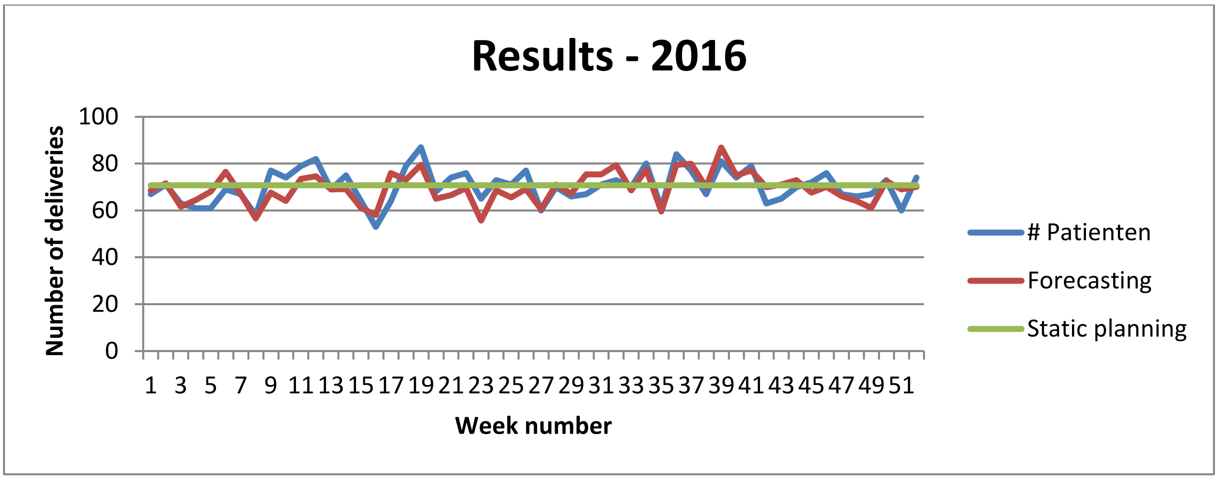


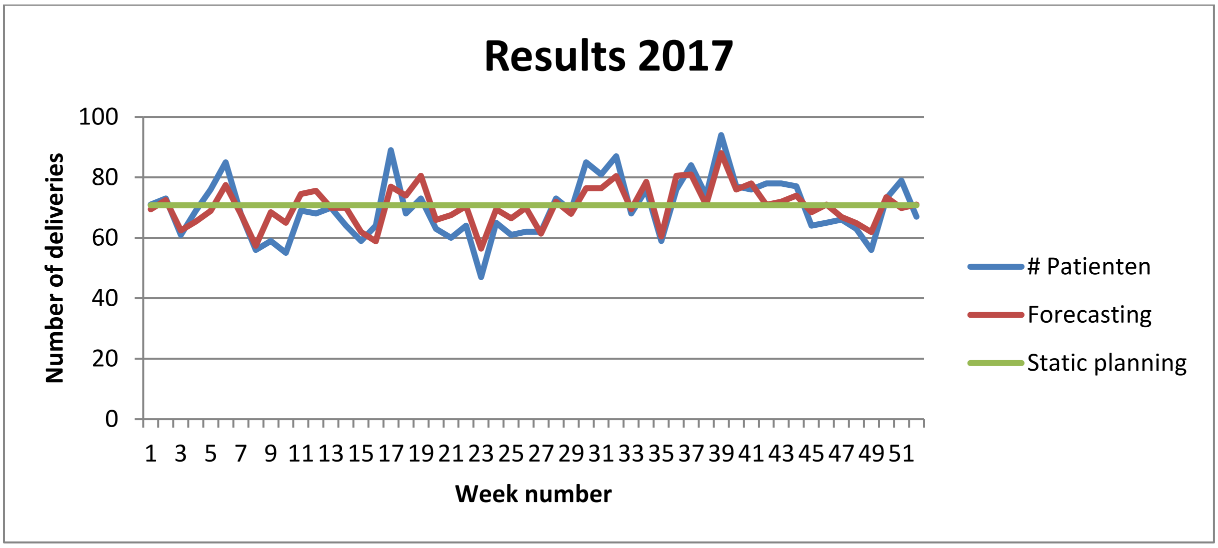


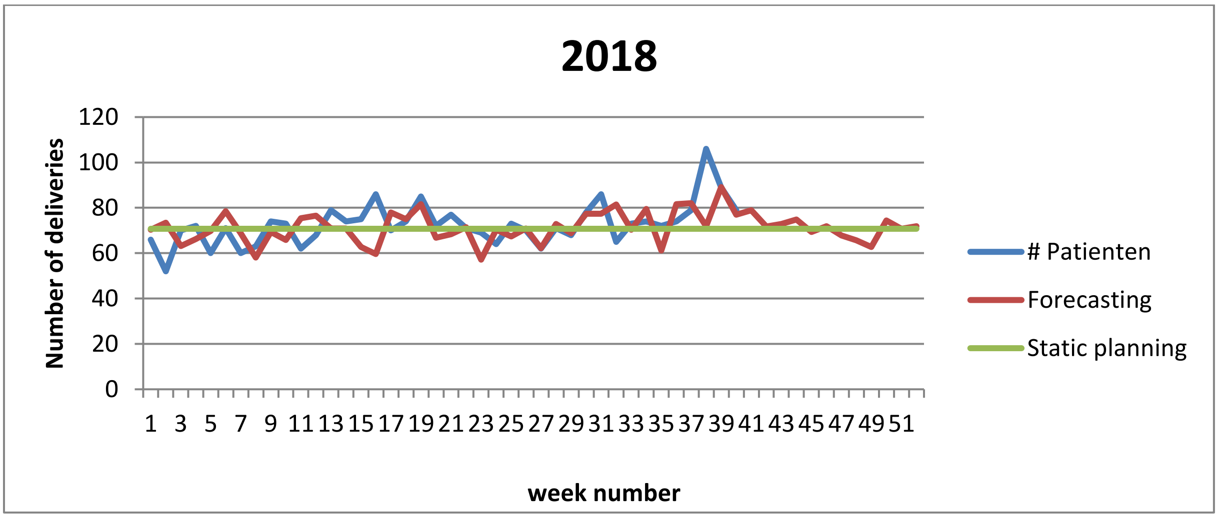


Appendix D – First Periodically distributed model

- Based on 2016, 2017 and an average seasonal factor.

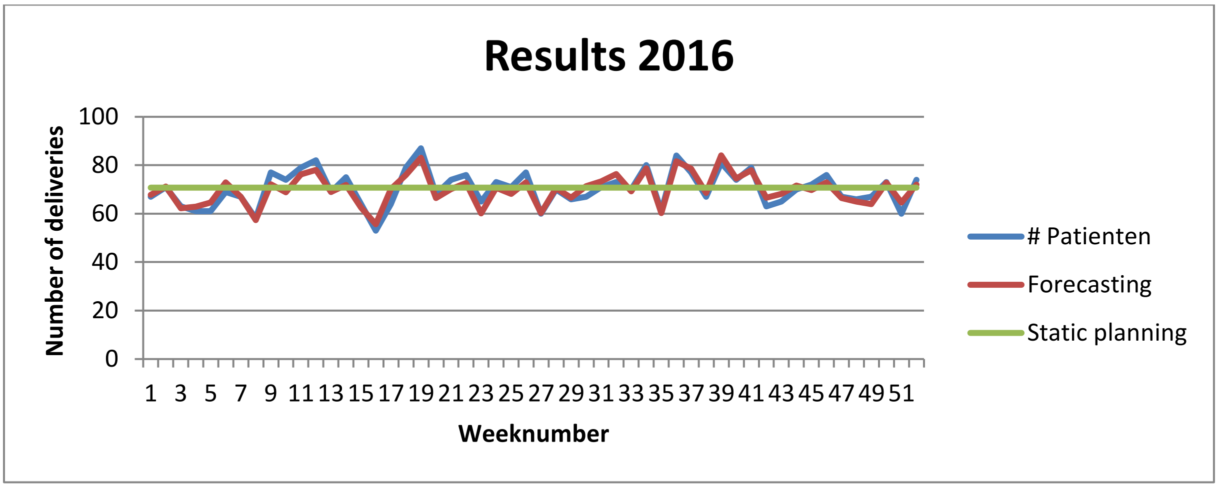


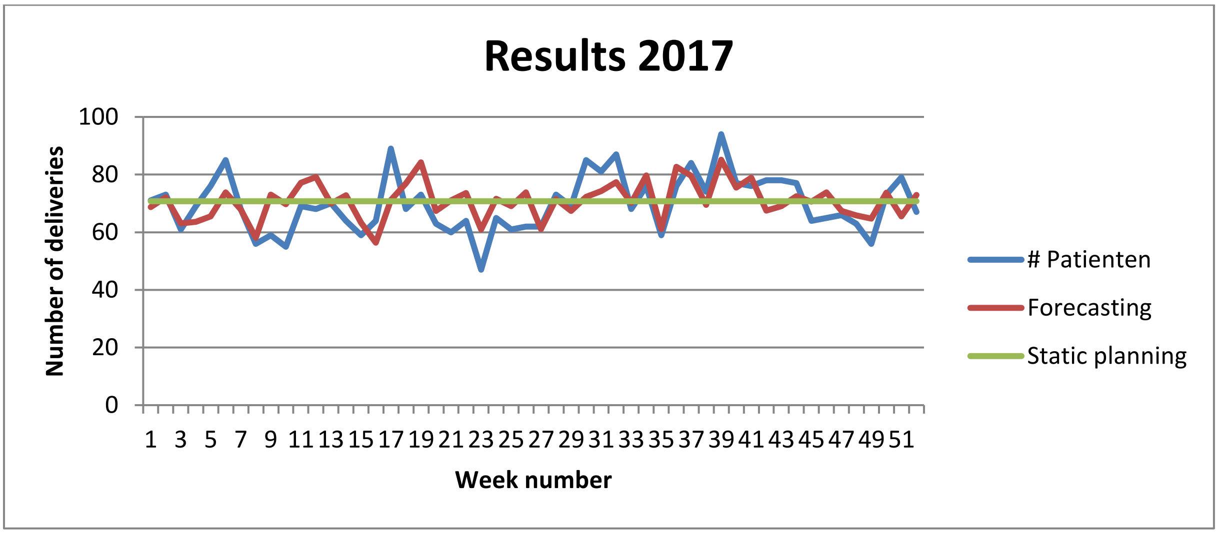


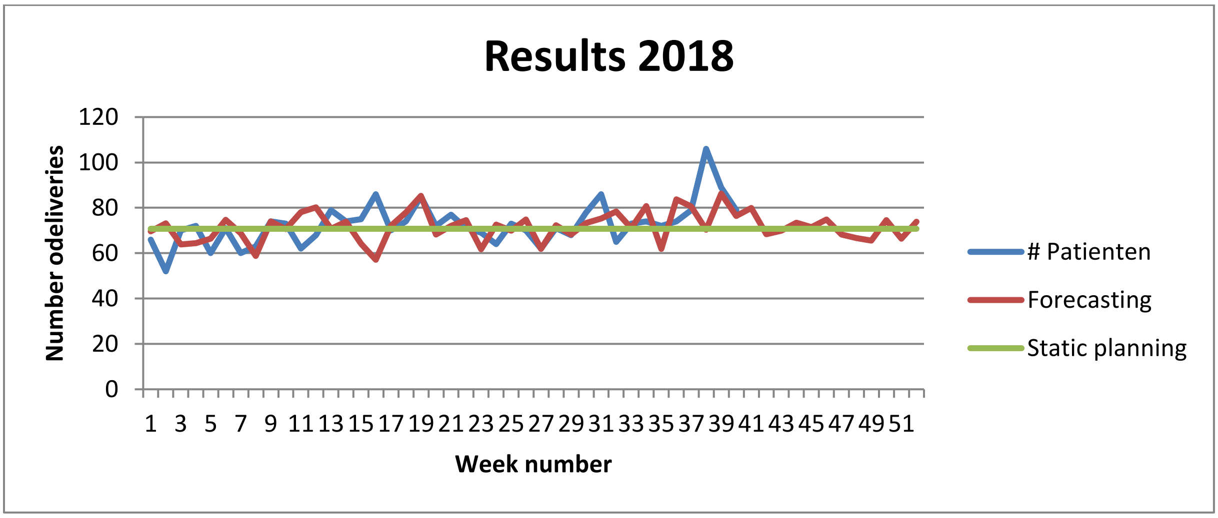


Appendix E – Second periodically distributed model

- Based on 2016 & 2017 and a weighted average seasonal factor.

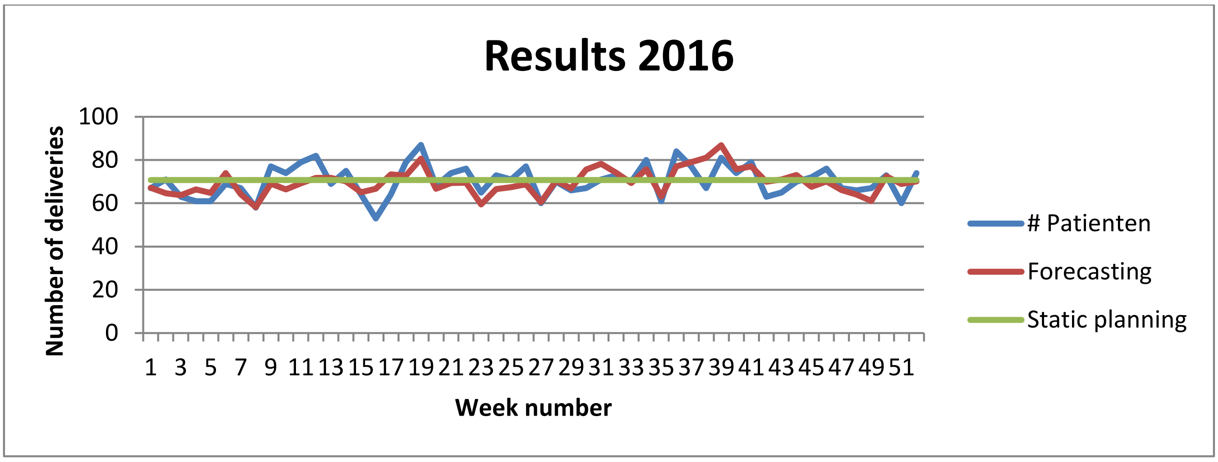


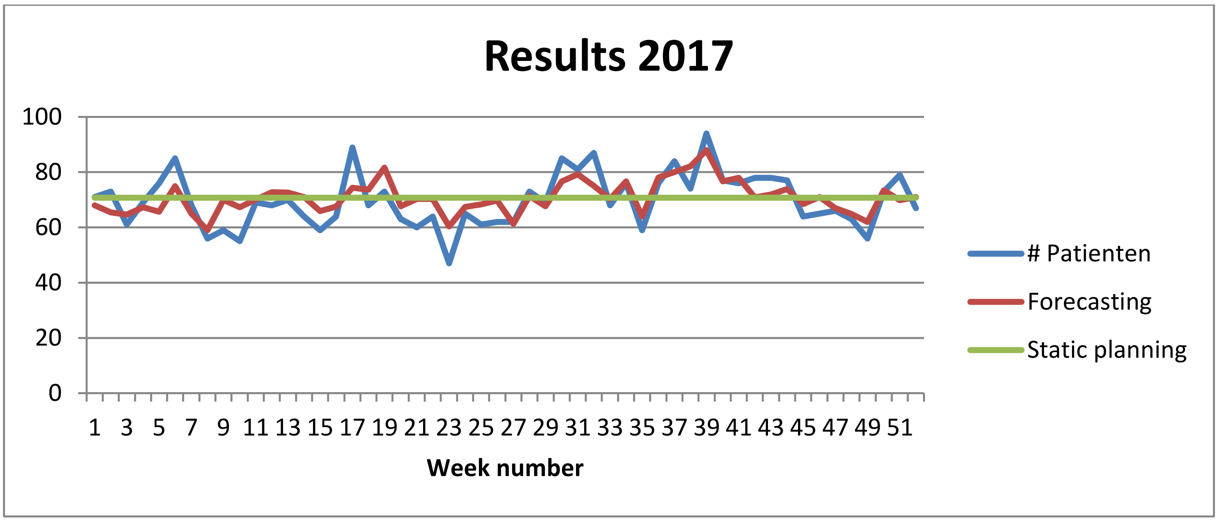


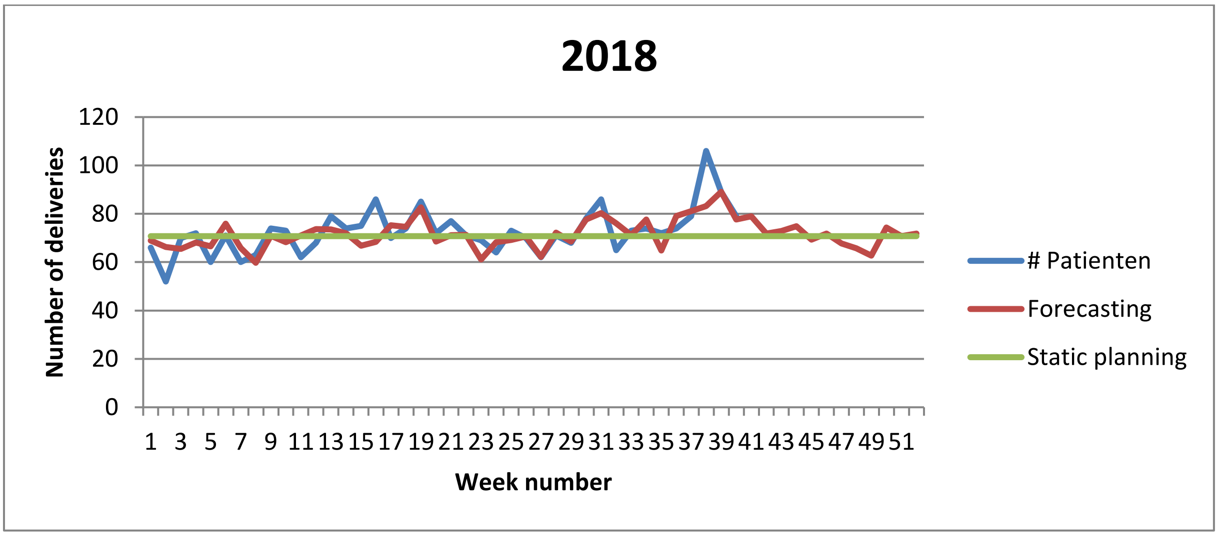


Appendix F – Third periodically distributed model

- Based on 2016 till November 2018 and an average seasonal factor.

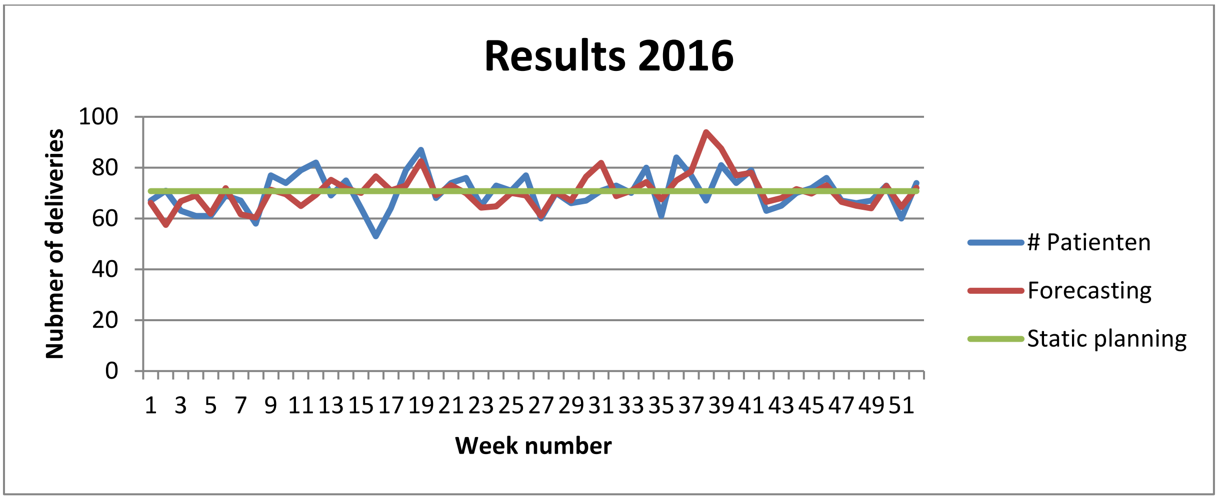


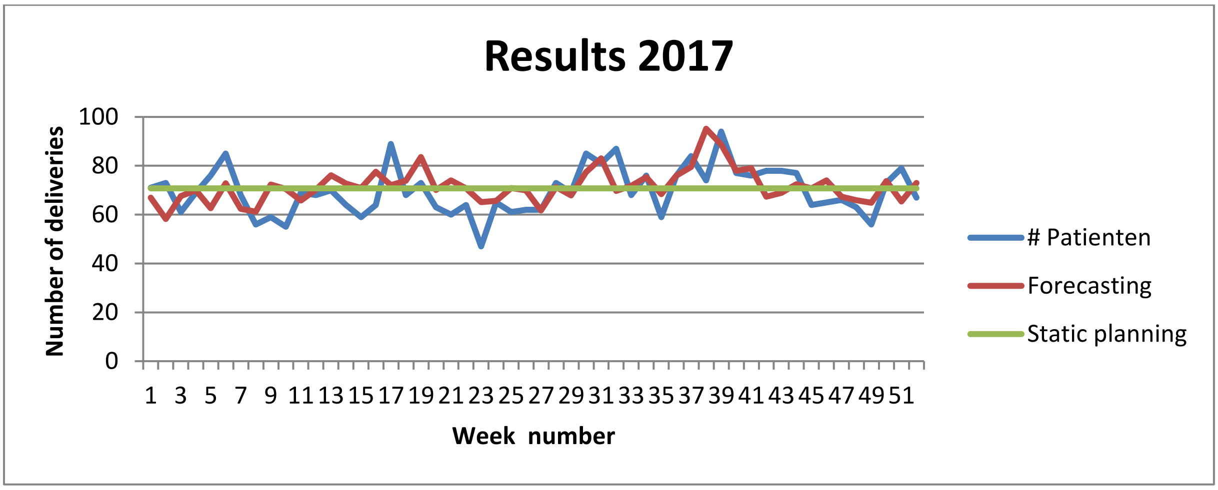


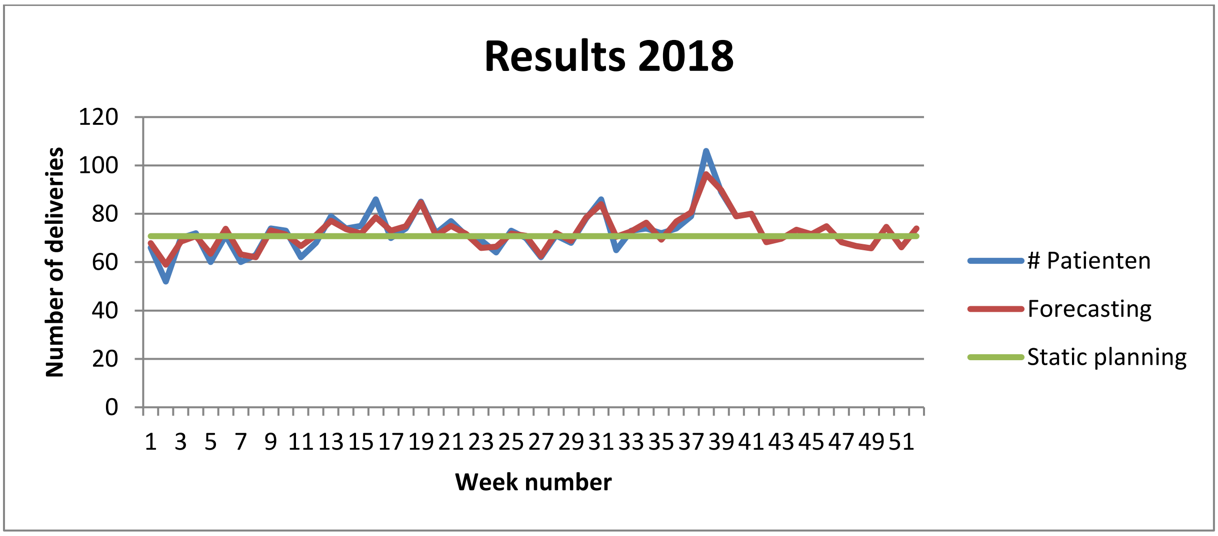


Appendix G – Fourth periodically distributed model

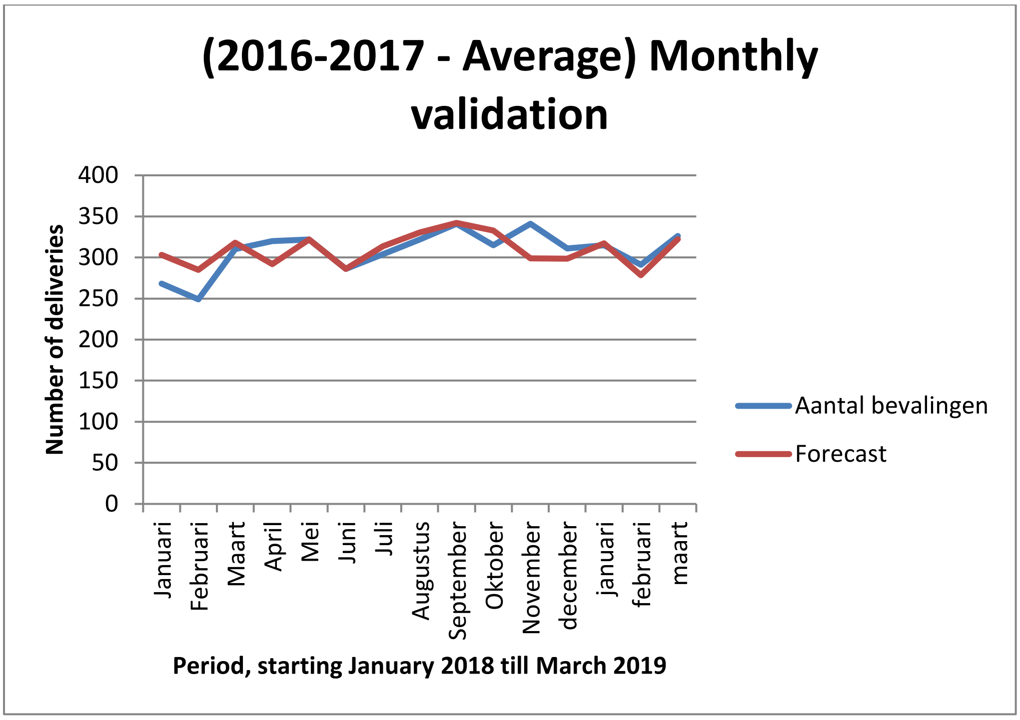
- Based on 2016 – November 2018 and a weighted average seasonal factor.

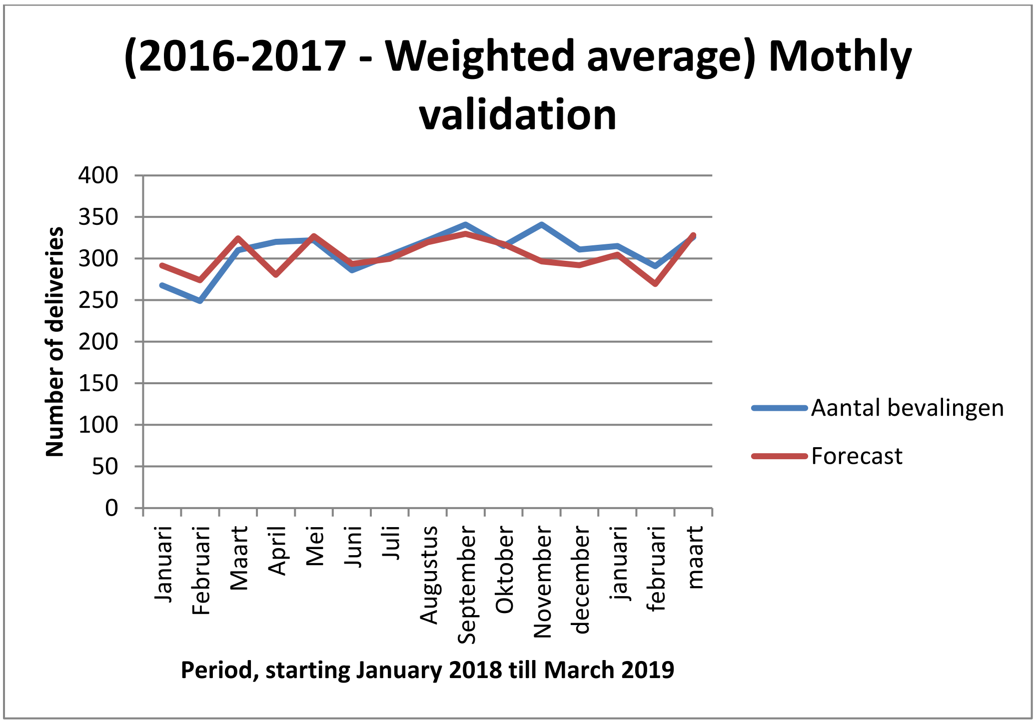


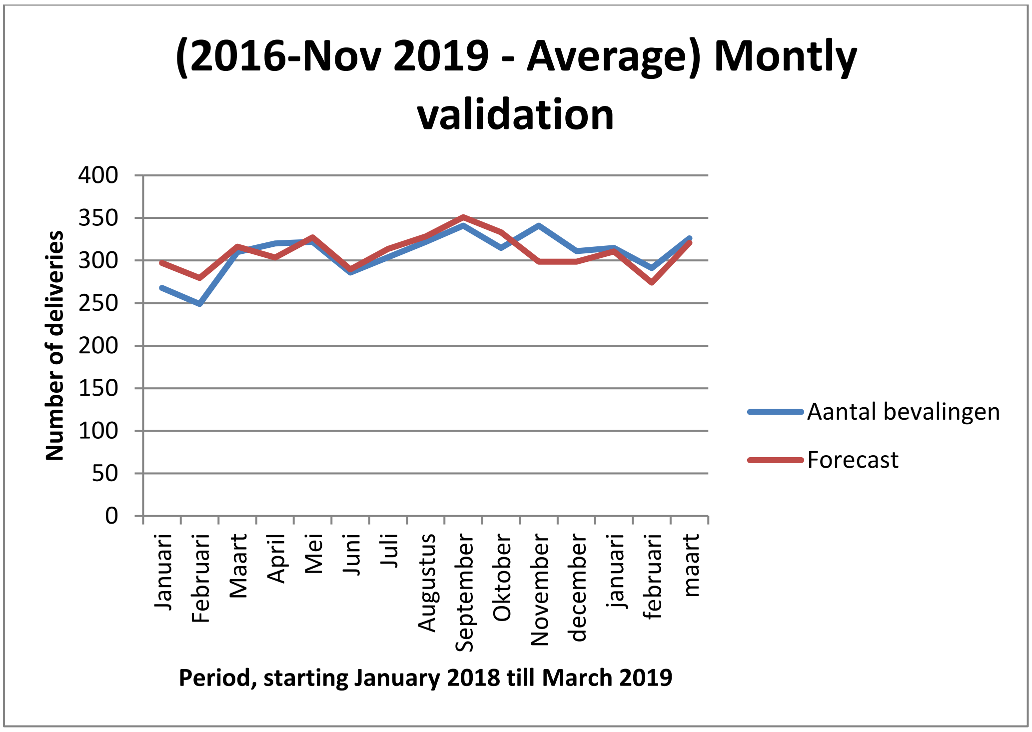




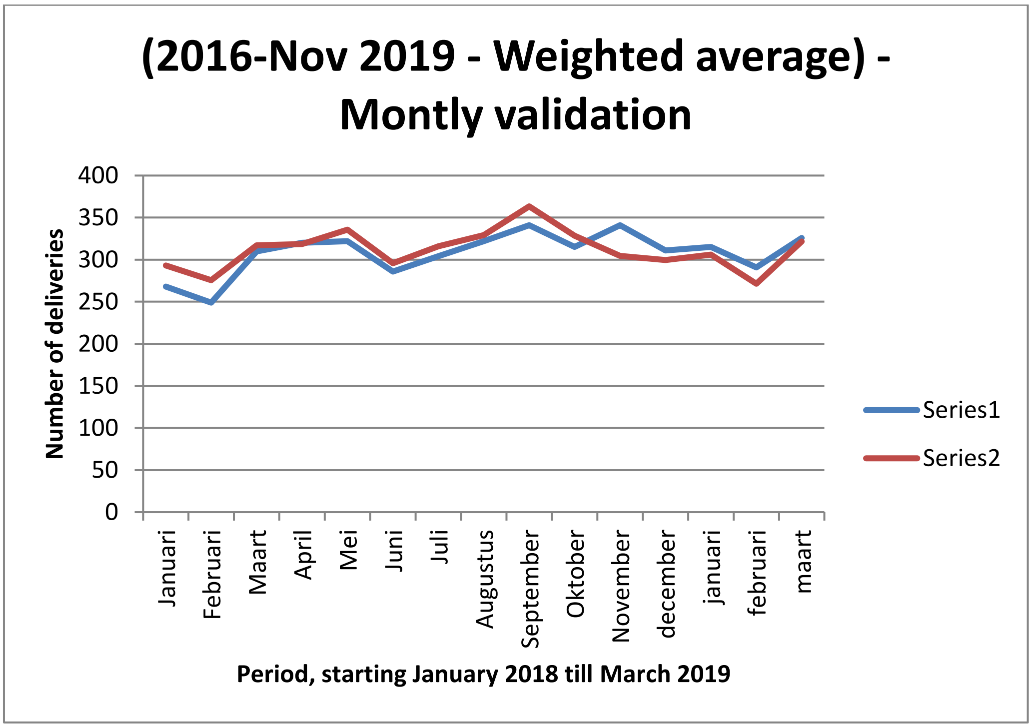
Appendix H – Monthly validation for different models





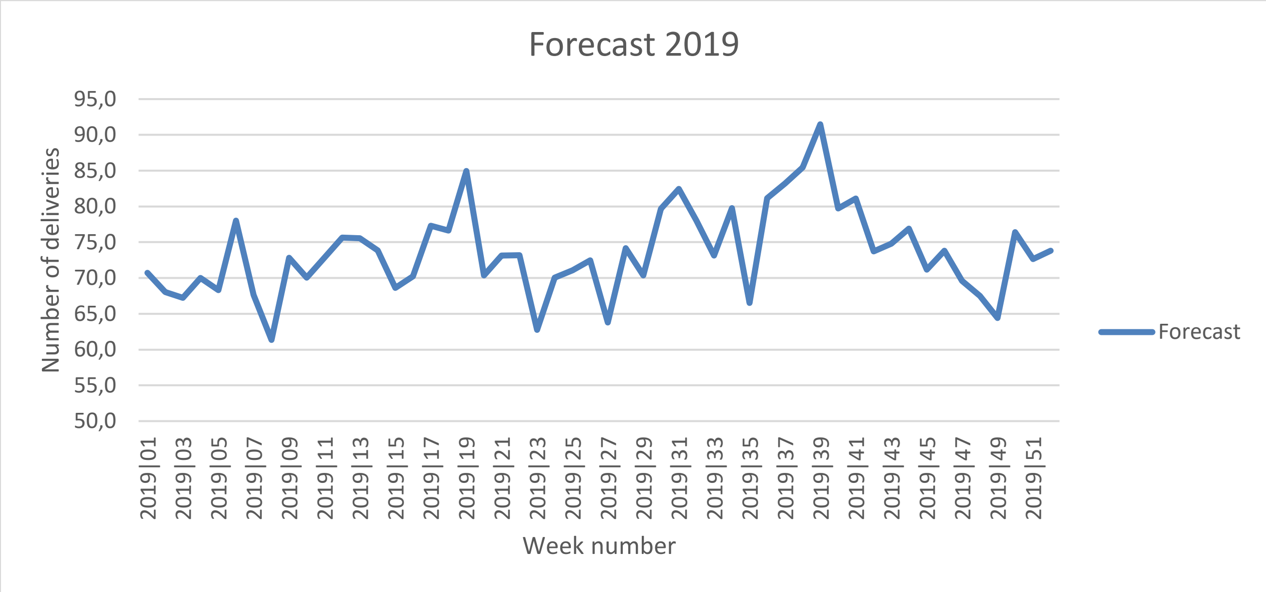


(2016-Nov 2019 - Average) Montly validation
Number of deliveries
400
350
300
250
200
150
100
50
0
Januari
Februari
Maart
April
Mei
Juni
Juli
Augustus
September
Oktober
November
december
januari
februari
maart
Period, starting January 2018 till March 2019
Aantal bevalingen
Forecast



(2016-Nov 2019 - Weighted average) - Montly validation
Number of deliveries
400
350
300
250
200
150
100
50
0
Januari
Februari
Maart
April
Mei
Juni
Juli
Augustus
September
Oktober
November
december
januari
februari
maart
Period, starting January 2018 till March 2019
Series1
Series2

Appendix I – Forecast 2019



Period	Forecast		
2019|01	70,7	2019|27	63,8
2019|02	68,0	2019|28	74,2
2019|03	67,2	2019|29	70,4
2019|04	70,0	2019|30	79,7
2019|05	68,3	2019|31	82,4
2019|06	78,0	2019|32	78,0
2019|07	67,6	2019|33	73,1
2019|08	61,3	2019|34	79,8
2019|09	72,8	2019|35	66,5
2019|10	70,0	2019|36	81,2
2019|11	72,9	2019|37	83,2
2019|12	75,6	2019|38	85,5
2019|13	75,5	2019|39	91,5
2019|14	73,9	2019|40	79,7
2019|15	68,6	2019|41	81,1
2019|16	70,2	2019|42	73,7
2019|17	77,3	2019|43	74,8
2019|18	76,6	2019|44	76,9
2019|19	85,0	2019|45	71,2
2019|20	70,4	2019|46	73,8
2019|21	73,1	2019|47	69,6
2019|22	73,2	2019|48	67,5
2019|23	62,7	2019|49	64,4
2019|24	70,1	2019|50	76,4
2019|25	71,1	2019|51	72,7
2019|26	72,5	2019|52	73,8

Appendix J – Instructions developed model

Since the obstetrics department is the user of the model, therefore the instructions are written in Dutch.

7.1 Overzicht model

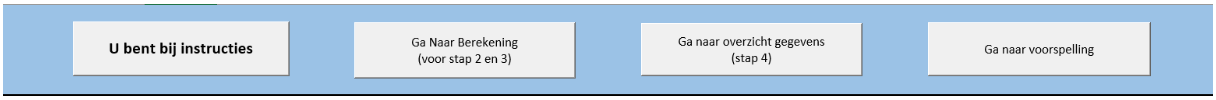
Het voorspel model is gemaakt om het aantal bevallingen per week te voorspellen. Met deze voorspelling kan vroegtijdig op en afgeschaald worden. In dit boekje wordt uitgelegd hoe het model gebruikt moet worden.

Allereerst bestaat het model uit 4 verschillende pagina's.

1. Instructies: Op deze pagina is een overzicht van de stappen te vinden om tot een goede voorspelling te komen. Daarnaast moet op deze model de input voor het model worden ingevoerd.
2. Berekening: Op deze pagina is een grote tabel te vinden. Wanneer er nog geen input is staat er overal in deze tabel "#Waarde!". Zodra er input wordt gegeven en stap twee wordt uitgevoerd, wordt deze tabel automatisch ingevuld. Daarnaast is ook stap 3 op deze pagina te vinden.
3. Overzicht gegevens: Hier zijn alle gegevens van het model te vinden die leiden tot een juiste voorspelling. Hier de mogelijkheid gegeven om in te grijpen in het model in het geval er externe factoren zijn waar rekening mee gehouden moet worden.
4. Voorspelling: Op deze pagina is de voorspelling voor de komende jaren te vinden.

Bovenaan elke pagina is een navigatiebalk te vinden, zie tabel 1. Door op de knop te drukken wordt de desbetreffende pagina geopend.

Tabel 1: Navigatiebalk



7.1.1 Achtergrond model

Er was nog geen wetenschappelijk onderzoek beschikbaar naar geschikte voorspelmodellen voor een verloskunde afdeling. Daarom is er gekozen om een nieuw model te ontwikkelen. Dit model is een 'periodically distributed model'. Dat houdt in dat het model uitgaat van een bepaalde seizoensfluctuaties. De gedachte hierachter is dat het totaal aantal geboorten in Nederland erg bepaald is per seizoen. In Nederland ligt de geboortepiek rond januari en september. Daarom is gekeken of een periodically distributed model ook geschikt is voor de verloskunde afdeling. In mijn onderzoek heb ik het model gevalideerd en blijkt het te werken.

Het model bestaat uit 5 stappen. Stap 1 is het invoeren van een 'input' voor het model. Voor stap 2 zijn enkele instructies toegevoegd in het model. Deze stap zorgt er voor dat het periodieke model wordt gebouwd. Stap 3 checkt of er kleine veranderingen in het seizoenspatroon zijn en compenseert hiervoor. Stap 4 geeft je de mogelijkheid om in te grijpen in het model wanneer er bijvoorbeeld meer vrouwen naar Isala komen door het sluiten van andere ziekenhuizen. Stap 5 is de output. Dit is het aantal bevallingen per week voor de komende jaren.

7.2 Eerste keer gebruik

Ga naar de pagina 'Instructies'. Hier is een overzicht van alle stappen weergegeven. Voor stap 1 moet de input voor het model ingevoerd worden. De input voor het model is het aantal gerealiseerde bevallingen per week, het gaat dus om de historische data van de verloskunde afdeling.

In de template staan de jaren 2016 t/m 2024. De cell met 2016 kan worden aangepast. Vul hier het eerste jaar in waarvan je de data beschikbaar hebt. Stel je gaat de data van 2020 invoeren, dan kan je 2020 invullen in plaats van 2016. Je ziet dat de overige jaren automatisch aanpassen. Vul vervolgens het aantal bevallingen per week voor deze jaren in.

Let op: Zorg er voor dat er minimaal 2 jaren zijn ingevoerd. Bij voorkeur is het meer betrouwbaar om meerdere jaren in te vullen.

Voor stap 2 en 3 navigeer je naar de pagina 'berekening'. Je ziet dat de tabel overal "#Waarde!" weergeeft. De tabel wordt automatisch gevuld na het uitvoeren van stap 2. Stap 2 en 3 worden uitgebreid en stap voor stap toegelicht in het model. Stap 2 is een analyse van de trend in het aantal bevallingen op de afdeling. Deze trend is nodig om te bepalen hoe de seizoensfluctuatie er uit ziet. Deze seizoensfluctuatie wordt automatisch bepaald na het uitvoeren van stap 2.

Stap 3 bevindt zich naast stap 2. Ook deze stap wordt uitgebreid toegelicht in het model. In deze stap wordt de 'oplosser' functie gebruikt om te zoeken naar de waarde van alpha. De theorie hierachter is dat het mogelijk is dat er kleine veranderingen in het seizoenspatroon te vinden zijn. De 'oplosser' tool zorgt er voor dat hiervoor automatisch wordt gecompenseerd.

Stap 4 geeft je de mogelijkheid om in te grijpen in het model. Tijdens mijn onderzoek kwam naar voren dat het ziekenhuis in Hoogeveen onlangs was gesloten. Daarnaast was het ziekenhuis in Lelystad failliet waardoor mensen vanuit de omgeving Urk richting Isala kwamen. Deze veranderingen zijn 'externe' factoren want ze zijn niet te vinden in de historische data en het vergroot de populatie zwangere vrouwen die naar Isala komt. Wanneer er geen externe factoren van invloed zijn kan je deze stap overslaan. Wanneer er wel van zulke externe factoren zijn waardoor er meer vrouwen dan alleen de zwangere vrouwen uit de omgeving van Zwolle naar Isala komen, kan je met deze stap compenseren. De vraag is dus hoeveel procent meer vrouwen je verwacht door deze veranderingen. Het percentage voor deze stap moet naar eigen inschatting worden ingevuld.

Stap 5 bevindt zich op de pagina 'voorspelling'. Hier tref je door voorspelling voor de komende jaren aan. De vraag voor jullie is: Wanneer ga je op of afschalen met de personeelsbezetting.

7.3 Gebruik na eerste keer

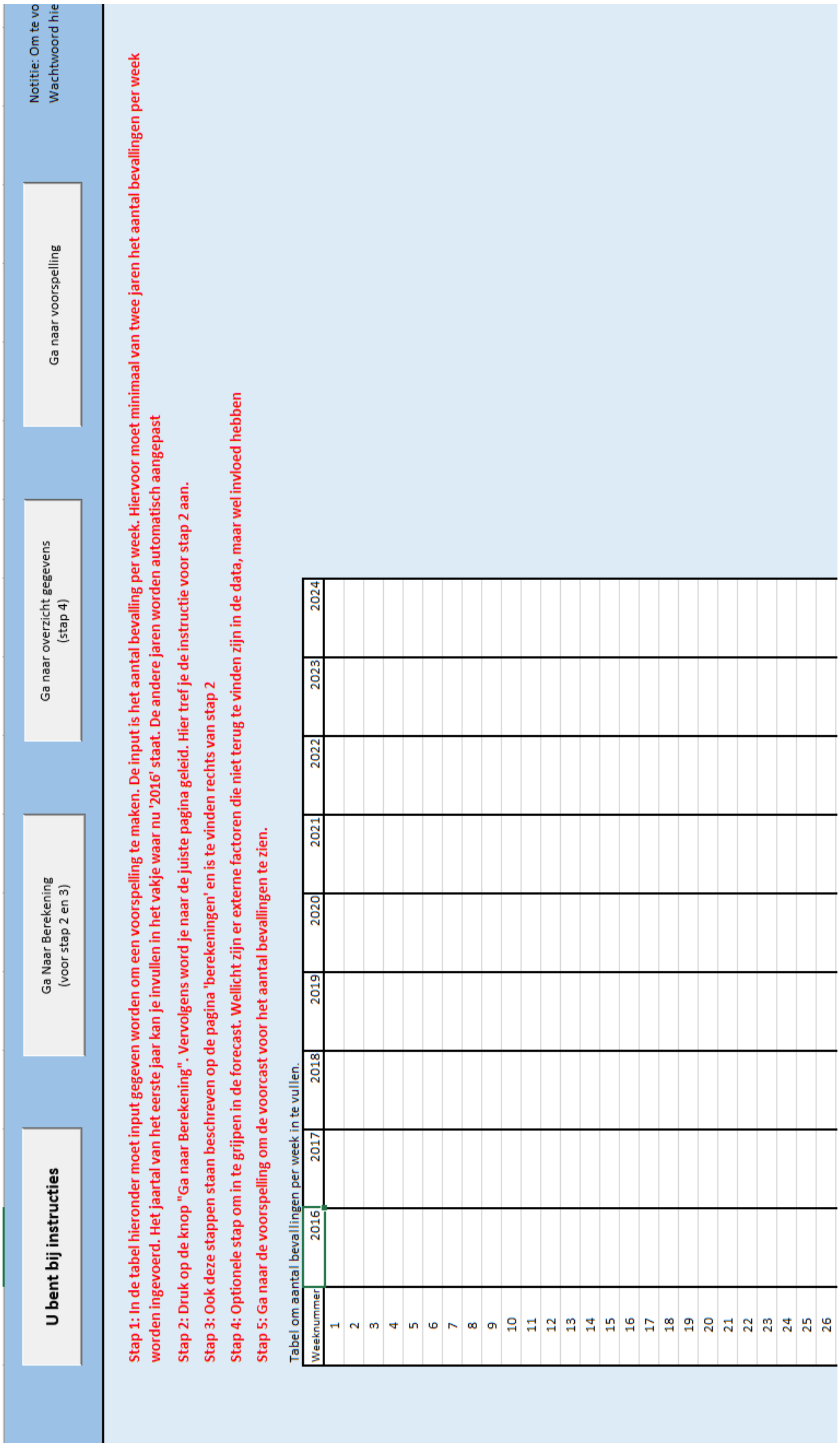
Ik adviseer jullie om het model 1 keer per jaar aan te vullen. Het model is al een keer gebruikt dus op de pagina 'instructies' is al data te vinden. Je kan op deze pagina nieuwe data (aantal gerealiseerde bevallingen) toevoegen. Daarnaast kunnen stap 2 en 3 opnieuw worden gedaan.

Doordat het model elk jaar wordt aangevuld, kan het zijn dat er veranderingen over de jaren heen zijn. Stap 3 is dus een soort validatie van het model. Zolang de waarde van alpha om en nabij 0 blijft, is het model bruikbaar. Er is geen onderzoek beschikbaar naar welke waarde van alpha acceptabel is. Persoonlijk adviseer ik jullie het volgende:

> Wanneer de waarde van alpha in stap 3 boven de 0,20 komt, dan open je een nieuwe template. De data van het aantal bevallingen van de afgelopen twee jaar kan dienen als input voor een nieuwe voorspelling. Dit betekent niet dat een periodically distributed model niet meer voldoet, maar juist dat er veranderingen in het seizoenspatroon te zijn. Daarom kan een nieuwe template met de laatste twee jaren als input leiden tot een nieuw geschikt periodiek model.

Appendix K – Screenshots developed model

First page:



Second page:

Ga naar instructies | **U bent bij de berekening** | Ga naar overzicht gegevens | Ga naar voorspelling

Notitie: Om te voorkomen d
Wachtwoord hiervoor is 'Is

De tabel hieronder geeft de berekening weer. Deze wordt volledig automatisch ingevuld na het uivoeren van stap 2. Vervolgens moet stap drie worden uitegevoerd, deze tref je aan op deze pagina, rechts van stap 2.

Jaar	Week	Periode nummer	Aantal bevallingen gerealiseerd	Deseasonalised	Regression	Seizoensfactor	Voorspelling	Afwijking	Mean Square Error	Exponential Smoothing	Nieuwe Afwijking	Mean Square Error
2016	1	1			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	2	2			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	3	3			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	4	4			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	5	5			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	6	6			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	7	7			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	8	8			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	9	9			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	10	10			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	11	11			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	12	12			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	13	13			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	14	14			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	15	15			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	16	16			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	17	17			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	18	18			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	19	19			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	20	20			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	21	21			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	22	22			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	23	23			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	24	24			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	25	25			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
	26	26			0	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!

Stap 2 - Regressie

Regressie is nodig voor een eerste analyse

Stap 2.1
Klik op het kopje 'gegevens' (deze vind je in het rijtje r

Stap 2.2
Vervolgens klik je op de tool 'hulpmiddelen voor gege

Stap 2.3
Vervolgens verschijnt er een menu genaamd 'gegevens

Stap 2.4
Vervolgens zie je verschillende velden die ingevuld m
Invoerbereik Y: In de tabel hiernaast link
vanaf week 27. Voor 'Invoerbereik Y' klik
Invoerbereik X: Hier klik je dezelfde rang
periode nummers.
Uitvoeroptiets: Hierbij klik je het bolletje
cel aanklikken. Dit moet Cel R37 zijn.

Klik de cel direct onder dez

Stap 2 - Regressie

Regressie is nodig voor een eerste analyse van het aantal bevallingen en wat voor trend er zich afspreeld

Stap 2.1
Klik op het kopje 'gegevens' (deze vind je in het rijtje met 'start - invoegen - pagina-indeling - formules - Gegevens - Controleren - Beeld - Ontwikkelaars

Stap 2.2
Vervolgens klik je op de tool 'hulpmiddelen voor gegevensanalyse', deze tref je rechts aan, binnen het 'Analyse' menu

Stap 2.3
Vervolgens verschijnt er een menu genaamd 'gegevensanalyse' in beeld. Klik in dit menu 'regressie' aan en vervolgens op 'oké'

Stap 2.4
Vervolgens zie je verschillende velden die ingevuld moeten worden:
Invoerbereik Y: In de tabel hiernaast links zie je een kolom met 'deseasonalised'. Deze kolom krijgt input vanaf week 27. Voor 'Invoerbereik Y" klik je deze eerste stel aan tot en met de laatste cel in
Invoerbereik X: Hier klik je dezelfde range als 'Invoerbereik Y" aan, maar dan met de corresponderende periode nummers.
Uitvoeroptiets: Hierbij klik je het bolletje met 'Uitvoerbereik:' aan. Vervolgens kan je in het lege vakje een cel aanklikken. Dit moet Cel R37 zijn.

Klik de cel direct onder deze aan: In het witte kader linksboven

Stap 2.5
Klik op oké. Er wordt nu een uitput van deze analyse weergegeven en deze werkt automatisch door in het model.

Stap 3 - Exponential Smoothing

In stap twee is de basis voor het 'periodically distributed system'. In deze stap wordt gecheckt of er veel veranderingen in het seizoenspatroon zijn

Waarde van Alpha:
0

Optimaliseren:
#VALUE!

Stap 3.1
Klik op het kopje 'gegevens' (deze vind je in het rijtje met 'start - invoegen - pagina-indeling - formules - Gegevens - Controleren - Beeld - Ontwikkelaars

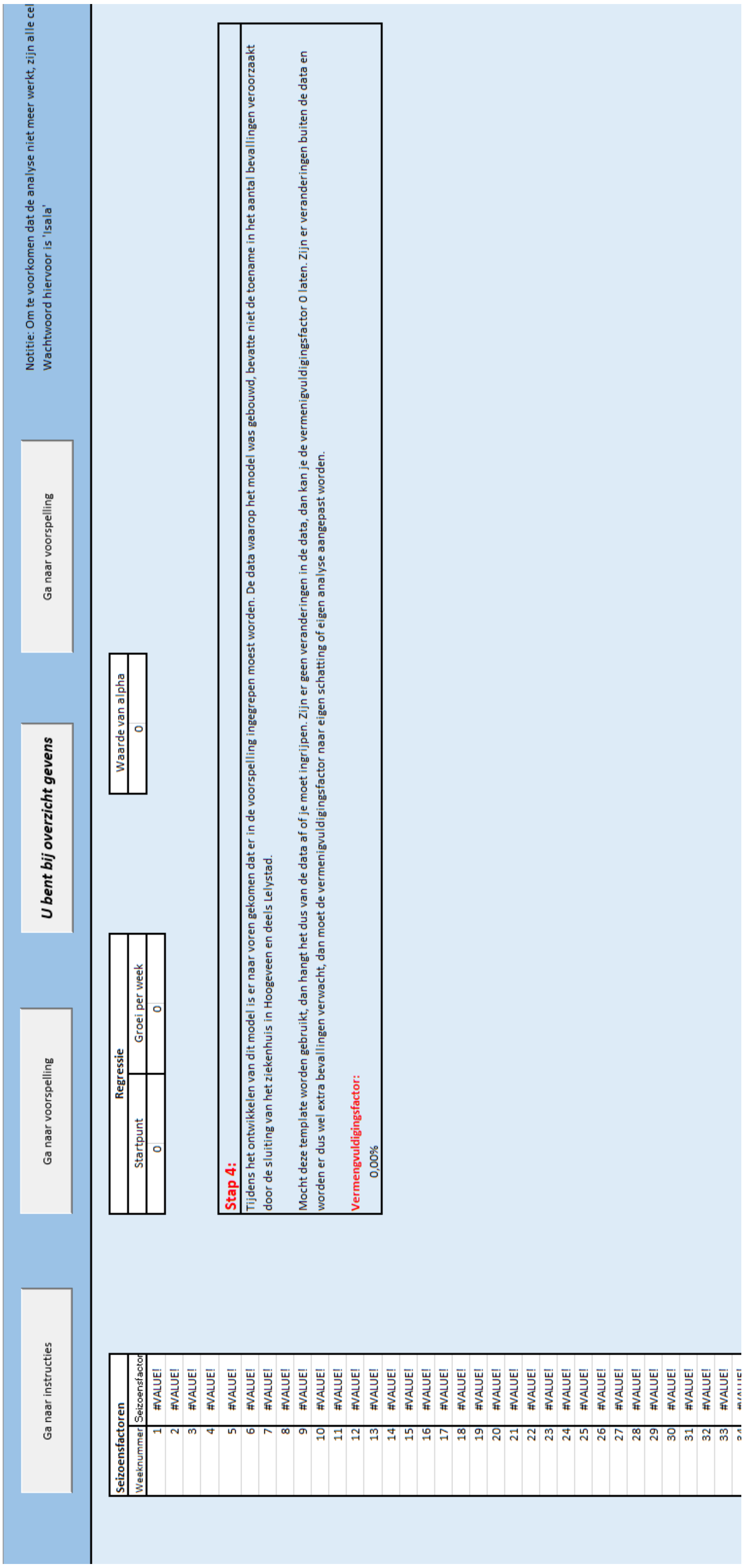
Stap 3.2
Klik op de tool 'oplosser', deze tref je rechts aan, binnen het 'analyse menu'

Stap 3.3: het volgende verschijnt in beeld.
Vervolgens verschijnt er een menu genaamd 'Parameters van Oplosser' in beeld. In dit menu moeten de volgende stappen worden ingevuld:

Doelfunctie bepalen: Klik de cel onder 'Optimaliseren:' aan (zie hierboven binnen stap 3)
Zorg er voor dat bij keuze menu 'Naar:' het bolletje met 'Min' is ingevuld.
Bij 'Door veranderen van variabelecellen:' klik je de cel onder 'Waarde van Alpha' aan (Zie hierboven binnen stap 3)

Stap 3.3
Klik vervolgens op oké. Er verschijnt een nieuw scherm in beeld. Klik op 'oplosser-oplossing behouden'. Hierdoor kan het zijn dat de waarde van Alpha verandert.
Informatie: De waarde van alpha heeft een bereik van 0,0 tot 1,0.
Waarde van alpha = 0,0. Dit betekent dat er geen veranderingen in het seizoenpatronen plaats vinden.
Waarde van Alpha is niet gelijk aan 0. Dit betekent dat er veranderingen in het seizoenpatroon plaats vinden en hiervoor wordt dankzij deze waarde van alpha automatisch gecorrigeerd.
Het is moeilijk te zeggen wanneer de waarde van alpha té groot is, en er dus veel veranderingen in het seizoenpatroon plaats vinden. Wanneer jullie het idee hebben dat de voorspelling in het aantal bevallingen niet meer klopt, adviseer ik om een nieuwe template op te starten en de data van de afgelopen twee jaar te pakken als input voor een nieuw model. Persoonlijk denk ik dat de waarde van Alpha niet boven de 0,2 mag komen, maar dat is een grove schatting.

Third page:



Notitie: Om te voorkomen dat de analyse niet meer werkt, zijn alle cel
Wachtwoord hiervoor is 'Isala'

Ga naar instructies | Ga naar voorspelling | ***U bent bij overzicht gevens*** | Ga naar voorspelling

Regressie	
Startpunt	Groei per week
0	0

Waarde van alpha
0

Stap 4:

Tijdens het ontwikkelen van dit model is er naar voren gekomen dat er in de voorspelling ingegrepen moest worden. De data waarop het model was gebouwd, bevatte niet de toename in het aantal bevallingen veroorzaakt door de sluiting van het ziekenhuis in Hoogeveen en deels Lelystad.

Mocht deze template worden gebruikt, dan hangt het dus van de data af of je moet ingrijpen. Zijn er geen veranderingen in de data, dan kan je de vermenigvuldigingsfactor 0 laten. Zijn er veranderingen buiten de data en worden er dus wel extra bevallingen verwacht, dan moet de vermenigvuldigingsfactor naar eigen schatting of eigen analyse aangepast worden.

Vermengvuldigingsfactor:
0,00%

Seizoensfactoren	
Weeknummer	Seizoensfactor
1	#VALUE!
2	#VALUE!
3	#VALUE!
4	#VALUE!
5	#VALUE!
6	#VALUE!
7	#VALUE!
8	#VALUE!
9	#VALUE!
10	#VALUE!
11	#VALUE!
12	#VALUE!
13	#VALUE!
14	#VALUE!
15	#VALUE!
16	#VALUE!
17	#VALUE!
18	#VALUE!
19	#VALUE!
20	#VALUE!
21	#VALUE!
22	#VALUE!
23	#VALUE!
24	#VALUE!
25	#VALUE!
26	#VALUE!
27	#VALUE!
28	#VALUE!
29	#VALUE!
30	#VALUE!
31	#VALUE!
32	#VALUE!
33	#VALUE!

Fourth page:

Ga naar instructies

Ga naar berekening

Ga naar overzicht gegevens

U bent bij de voorspelling

	Forecast								
Week	2016	2017	2018	2019	2020	2021	2022	2023	2024
1	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
2	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
3	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
4	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
5	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
6	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
7	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
8	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
9	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
10	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
11	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
12	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
13	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
14	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
15	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
16	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
17	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
18	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
19	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
20	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
21	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
22	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
23	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
24	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
25	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
26	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
27	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
28	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
29	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
30	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
31	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
32	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
33	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!