
# *The Road to Forecasting Success by means of Classification*

*A Case Study at Wavin*

Master Thesis

***L.M. Noordstar***

***September 2019***

# *The Road to Forecasting Success by means of Classification*

*A Case Study at Wavin*

**Leon Noordstar**

*September 2019*

*Enschede, The Netherlands*

MASTER THESIS

INDUSTRIAL ENGINEERING & MANAGEMENT

PRODUCTION & LOGISTICS MANAGEMENT

**University Supervisors**

*Dr. I. Seyran Topan*

*Dr. E. Topan*

UNIVERSITY OF TWENTE.

**Company Supervisor**

*E. Breeuwsma*

wavin

CONNECT TO BETTER

# Management summary

In this research we have examined the forecasting process of Wavin. For the more than twenty countries in Europe, there needs to be made monthly forecasts. Currently, first, the forecasts are generated by the Demand Managers using SAP APO by fitting different statistical methods on the historical data. Then, during the forecast meeting the Demand Manager discusses with Sales & Marketing how to adjust the statistical forecasts with the qualitative information, including the market events, upcoming projects and open orders.

Wavin has a product portfolio of over 30.000 SKUs and even though the products are grouped, making the forecasts requires split focus in the limited time the Demand Managers and Sales & Marketing have. This results that the forecasts are not always very accurate. Moreover, this research has indicated that the extra effort for adjusting the statistical forecasts with the qualitative information sometimes has a damaging effect on the forecasts resulting in a decrease of the forecast accuracy. This is of course not desirable. Therefore, with this research, we classified the products based on both importance and forecastability for the markets of Country A, Country B and Country C and we explain the different forecasting approaches for the classes. The main question we answer with this research is as follows:

*HOW CAN WAVIN IMPROVE ITS FORECAST ACCURACY FOR DIFFERENT TYPES OF PRODUCTS FOR DIFFERENT MARKETS BY PUTTING THE RIGHT FOCUS ON QUANTITATIVE AND QUALITATIVE METHODS?*

By making the classification, both the Demand Managers (statistics) as well as Sales & Marketing ( qualitative information) can focus on the aspects where they can make a difference, instead of losing time on the aspects where they have only a very limited impact.

For making the classification on importance (ABC) we used revenues as the parameter, applying the Pareto rule. For the forecastability (XYZ) we used thresholds of the statistical forecast accuracy (defined by 1 - wMAPE). Since the market in Country A is much more stable than the markets in Country B and Country C, we set different thresholds. For Country A we set thresholds of respectively 80% and 65% for class X and Y. For Country B and Country C we set thresholds of 65% for class X and 50% for class Y. All the remaining products are classified as class Z.

For each class a different approach can be applied. While for the more important and more difficult to forecast products (e.g. class AZ) the most Sales & Marketing input is needed, for the easier to forecast and with lower importance products (e.g. class CX), the least time should be spend and it can be more automated using merely the statistical forecasts. In Figure 1 the approaches per class are illustrated.

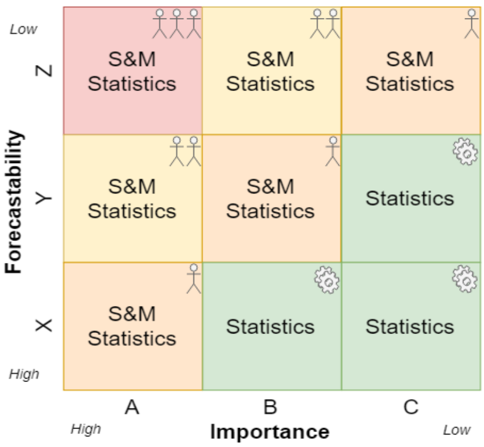


*Figure 1: Strategies per class*

Another valuable outcome of the classification is that different targets can be set for the classes. While a forecast accuracy of 60% can already be a big accomplishment for class AZ products, a forecast accuracy of 75% can be achieved by simple statistical methods for class AX products for example. Moreover, the forecasts can be less accurate for the products which are less important for the business.

A common occurrence, as is also the case for Wavin, is that adjusting the statistical forecasts with the qualitative information of Sales & Marketing leads to over-forecasting. Sales & Marketing benefits from the higher availability of the products and rather make sure there are too many products produced than too few. Therefore, we also researched what the best balance between the statistics and the qualitative information would be per class and per country. The possible increases of the forecast accuracies are shown in Table 1 when optimal (depending per class and country) and equal weights are used. We calculated the forecast accuracy when the forecast was calculated by the formula: $w_1$*statistical forecast + $w_2$*final forecast (with $w_1+w_2=1$

*Table 1: Forecast accuracy increase when optimal or equal weights are used*

| | Optimal weights | Equal weights |
|---|---|---|
| Country A | 3,5% | 2,1% |
| Country B | 5,2% | 4,9% |
| Country C | 1,2% | 0,1% |

With this research we also identify how the focus on quantitative information and how the focus on qualitative information can be done and improved. For making the statistical forecasts, good outlier correction and error handling needs to be carried out for the products which are in general the easiest to forecast with the statistics. Focusing on these easier to forecast products which are in class X makes most sense because especially for the more difficult to forecast products which are in class Z, the demand patterns are more random and spending more time on the statistics will (most likely) not improve the forecasts.

We also carried out our own forecasts using RStudio. We fitted the methods that are available in SAP APO, together with some other widely used methods. We used the first six months of 2019 to compare our forecasts with the forecasts generated by Wavin. We made the forecasts both with and without outlier correction. Moreover, we also researched whether forecast combination improves the forecast accuracy. We tested the results of the forecast combination with using one combination for all classes and when applied the best forecast combination per class. The results are shown in Table 2.

*Table 2: Forecast accuracies of experiments with RStudio*

| | Current forecast in SAP APO | Rstudio (wihtout outlier correction) | Rstudio (with outlier correction) | One combination for all classes | Optimal combi-nation per class |
|---|---|---|---|---|---|
| Country A | 84,8% | 83,8% | 84,7% | 87,2% | 88,0% |
| Country B | 74,8% | 77,7% | 76,8% | 78,9% | 80,5% |
| Country C | 66,3% | 67,9% | 66,2% | 72,1% | 72,8% |

Adding the extra methods using RStudio increases the forecast accuracies for Country B and Country C. For these countries, outlier correction did not improve the forecasts. For all countries combining multiple methods further increases the forecast accuracy up to almost 5%. The most common methods for making the forecasts were the seasonal naïve, seasonal regression and the mean method.

Using literature and the best practices of Wavin we also created a list of seventeen qualitative factors that can have influence on the sales which are not predictable by only considering the historical data. This checklist can be used for Sales & Marketing for giving their input during the forecast meetings.

Moreover, we implemented and tested the classification with the suggested approaches as in Figure 1 (iteration 1), together with the weights for combining the statistical and final forecast and the statistical forecast (combination) of RStudio (iteration 2) for Country B in July and August of 2019. We compared these results with the performance of whole 2018 and July/August of 2018. The results of the first iteration give better results compared to whole 2018, but not to July/August of 2018. However, the classification with the different approaches helps the Demand Managers and Sales & Marketing for making the forecasts and the increases in forecast accuracies are more likely to happen in future, when everyone is used to the new way of working. The results of the second iteration where more promising, where it increased the final forecast accuracy with about 1,7% compared to the first iteration.

# Preface

This master thesis concludes the five years that I have been studying Industrial Engineering & Management at the University of Twente in Enschede. I had the honor to do this thesis in corporation with the international company Wavin, located with its' head office in Zwolle.

My period at the company of Wavin is like the by Coca-Cola's chosen promotional anthem for the 2010 FIFA World Cup in South Africa; Wavin' flag by K'naan. "When I get older, I will be stronger" are the recurring lyrics of the song. Doing my thesis at Wavin showed me how businesses are run, it made me learn some of the hard and soft skills, and above all, spending all the hours at the desk in the office in Zwolle, it made me stronger.

For this, I want to express my thankfulness to Erik Breeuwsma, my supervisor of the company. He never missed an opportunity to explain me how the company of Wavin is structured and how the business is run. I thank him for all the constructive feedback he gave me. He not only offered many times to read the report, looking at it with fresh eyes, but also took time every week to discuss the progress, giving new insights. I could not wish a better and more dedicated supervisor than him. Without him, sitting next to me in the "closet room" at the fifth floor, the research was not the same and I could not have been as motivated as I have been now.

Besides, many thanks go out to my supervisors Engin and Ipek Topan from the University of Twente. Although their overly busy schedules, they accepted the challenge with me to carry out this research to good endings. They always found time to discuss the progress, giving their inputs and ideas about the topic. The always friendly but good feedback, giving me much freedom how to do the research, helped me to experience to the fullest how it is carry out a six-month research.

Last but not least (and not 'least but not last' as an Italian program director once misspoken herself, when I was present at a graduation in Milan), I want to thank my family and friends for all the support they gave me. They helped me to keep on persisting, taking the most out of it and giving moral support.

All in all, I can look back on a period where almost all things went smoothly, having no setbacks or delays in the process of doing the research. I can say I learned a lot and it helped me to prepare and to live to the moment that the phase of studying is over, and where a new quite exciting period is laying ahead of me; the working life. I hope and I am convinced I will, is to get most out of my career, pursuing the passions that I still partly need to discover, but always seeing it in the perspective of the broader sense of life. Work is a big part of life, but family, friends and even more God are the things that matter most for me.

Leon Noordstar
*September 2019*

# Table of Contents

# 1. Introduction

This first chapter of the thesis introduces the research. This chapter will help to understand what led to the research and includes the problem approach. We think it is important to think about the future, which activities we will carry out, even though Albert Einstein's quote might suggest something else:

*"I never think of the future — it comes soon enough."*

The chapter starts in Section 1.1 with an introduction of the company Wavin by which the research is commissioned. Then, in Section 1.2, we describe the case background. We examine the core problem, using a problem cluster in Section 1.3. In the same section, we also explain who the stakeholders are. In Section 1.4, we set the scope and the research goal, after which we pose the main research question, together with the sub-questions in Section 1.5. In this section we also give the research framework.

## 1.1 About Wavin

The foundations of Wavin started when in the early 1950s, the local water utility company WMO encountered serious problems of pipe corrosion and leakage for distributing drinking water. The company founder and director Johan Keller found it necessary to do something about it urgently. He started in a small workshop in Zwolle to produce the first large diameter plastic pressure pipe for potable water. Soon, this solution attracted attention both nationally as internationally, and the government organization was unable to cope with the increasing demand. A new independent company had to be created, which happened in August 1955. It got the name Wavin, derived from WAter and VINyl.

Today, Wavin is a global leader in the supply of plastic pipe systems and solutions above (45%) and below ground (55%). It has a total production portfolio of over 30.000 SKUs of which the implementations are illustrated in Figure 2.

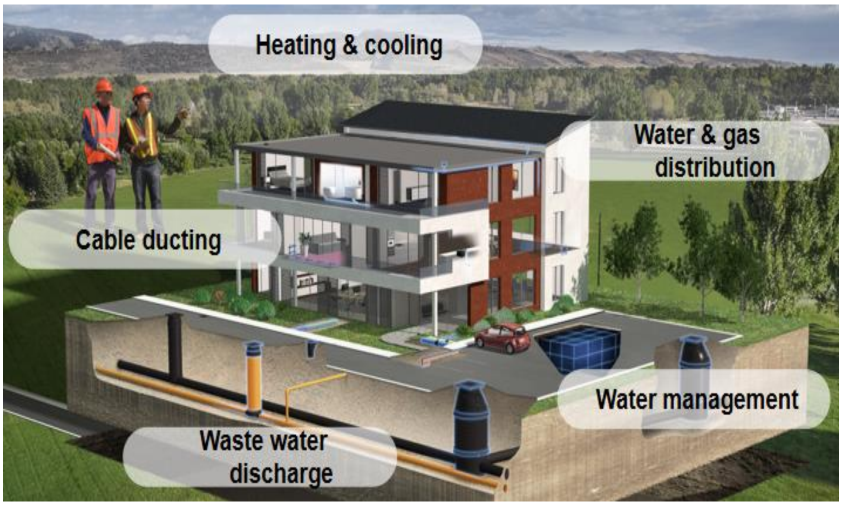



*Figure 2: Implementations of Wavin products*

Wavin is involved in both major prestige construction programs and small domestic installations and refurbishments in more than twenty countries. Most of these countries are in Europe, where it is the market leader. Besides, Wavin has 29 production sites (e.g. Hardenberg; The Netherlands, Twist; Germany), with its head office in Zwolle. It employs about 5000 employees and has yearly revenues fluctuating around $1,2 billion. In 2012, Wavin became part of the conglomerate Mexichem, a Latin American company and world leader in pipe systems and active in chemicals and materials (Wavin, 2019).

## 1.2 Case background

At Wavin there is a current program ongoing to implement Sales & Operation Planning (S&OP), across the twenty countries, which is financial and scenario driven and aligned with their current processes. S&OP is a business process that helps companies to balance demand and supply (on aggregated level). It ensures a close cooperation between Operations, Sales, Finance and Product Development. Moreover, it links the strategic plans with the business plan. Some widely acknowledge benefits are a higher customer service level and lower finished goods inventory (Wallace, 2004). This S&OP, also known as Integrated Business Planning (IBP), consists of 5 monthly processes which are illustrated in Figure 3.

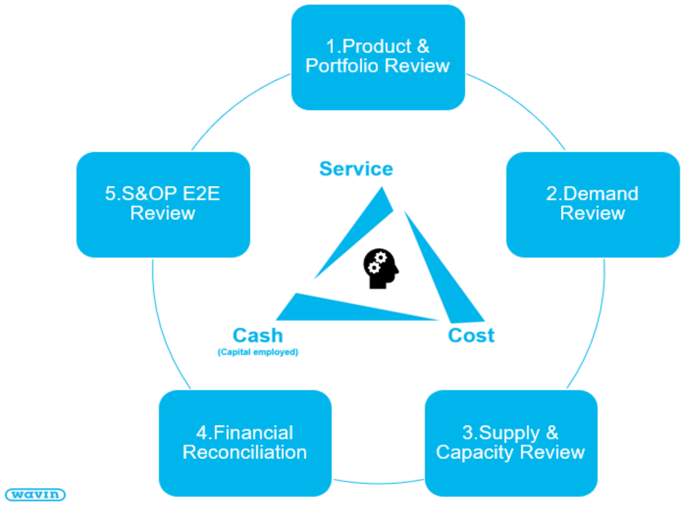



*Figure 3: Steps of S&OP*

One of the areas where there is space for improvement is the Demand Review process step (step 2). Currently, the forecasts for the products are done with a so-called 'one-size-fits-all' forecasting methodology. This means that although different statistical forecasts are used, like exponential smoothing or moving average, the approach for making the forecasts is almost the same for all products. There is only a basic ABC classification which makes a distinction between high and low impact products (e.g. based on revenues and/or orderliness). We refer to Section 3.3 for a broader description of this classification explained by literature. There are seven Demand Managers, who make the forecasts for all the countries. How these countries are divided among the Demand Managers is listed in Table 37 in Appendix A.

Besides this ABC classification, there is not taken into account the predictability of the products. Currently, for all products first a statistical forecast is carried out. This is done on aggregated level, using planning groups (we explain the grouping in 2.2). Then, on product hierarchy, qualitative data, like product promotions or price changes, are added. The statistical forecasts are done with a bit more focus for the A products than for the other products. However, there are no clear guidelines and it is up to the Demand Managers how to do this.

Besides, since adding this qualitative data needs to be evaluated for all product(families), it results for split focus. However, by classifying the products according to predictability, there can be given more attention on the more difficult to predict products. Besides, for the products which are more difficult to forecast, the human adjustments can be done on lower aggregate levels for example. Moreover, classifying enables for setting different targets for different groups. Currently, Wavin uses just one KPI, which is the forecast accuracy for A products and fluctuates around 70%. This accuracy is calculated by 1 minus the 'Weighted Mean Absolute Percent Error' (WMAPE) over the last month (see Section 3.6

for the explanation of the WMAPE). By setting different targets for different groups, based also on predictability, will help to evaluate the forecasts better. Class A products which are easy to forecast should have a higher target than B products with less predictability, for example.

By better focusing and better evaluating, our expectation is that the overall forecast accuracy can be increased. This is very beneficial business wide, which we will explain shortly in the next section. See Section 3.8 for a more elaborate explanation of the implications of better forecasting.

### 1.2.1 Implications of a higher forecast accuracy

The impact of forecast accuracy can be explained by the Supply Chain Triangle (DeSmet, 2018) also known as the 'devil's triangle', as which the middle part of Figure 3 in the previous section shows. Organizations have three main focus points; service, cash and cost. Ideally the service, with among others the target service level, is as high as possible. Besides, the cash, which is the working capital (e.g. inventory and accounts payable/receivable) and fixed assets is desired to be as low as possible. However, in order to have a high service level and low working capital, the costs will inevitably be high. The production needs to be highly flexible in order to meet the service level and to have a low inventory. This is just one example that when improving one or two areas will inevitably mean a decline of another. Therefore, trade-offs should be made and analyses about what is the best division of the areas are necessary.

Currently, since the forecast accuracy is around 70%, the safety stock needs to be relatively high, since the actual demand can variate from the forecasted demand. On the other hand, the service level will not be that high either, since when the safety stock is not sufficient, the order will not be completely met, and backorders occur. This in result, can cause penalties or a loose of customers. An increase in the forecast accuracy will make it possible to increase all the three areas of the devil's triangle at the same time, which can be explained as follows. A higher forecast accuracy means that demand can be predicted better. As a result, the safety stock can be lower, which means lower cash. Besides, the orders can be met more often, which means a higher service level. Third, the costs will be lower as a result of more stable production and less administration of backorders.

## 1.3 Problem identification and stakeholders

As we made clear, the problem is that there are still some improvements possible regarding the forecast accuracy, which may result in lower service levels, higher working capital and higher costs. In order to find the root of the problem, which is the core problem, and to illustrate the consequences of it, a problem cluster is shown Figure 4.

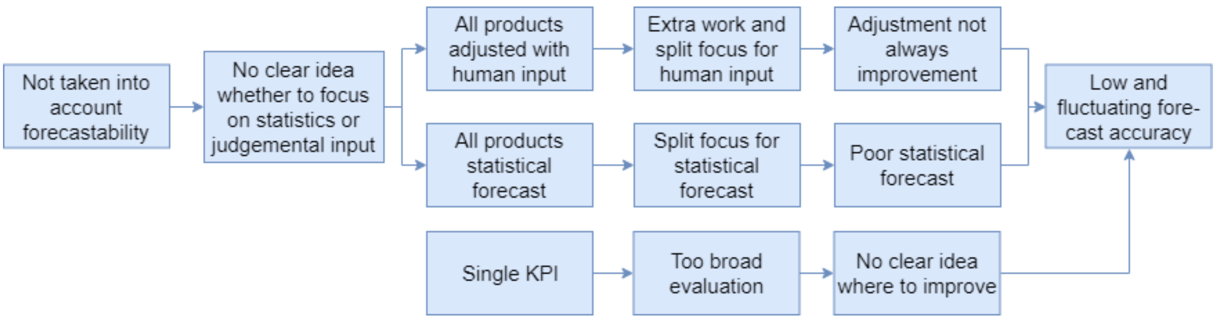



*Figure 4: Problem cluster*

The core problem, as we tackle with this research, is that for making the forecasts there is not taken into account the forecastability of the products. The result is that all products get the same attention for making the statistical forecasts and qualitative input. This requires split focus to handle the (ten)thousands of SKUs even though they are grouped. The result is that for certain products the human adjustments are often not an improvement over the statistical forecast. Besides, the statistical forecast could have done better by putting more effort for certain products.

Although for some products it may be easy to forecast and statistics may give an adequate forecast, other products are fluctuating extensively and may be only be forecasted accurately by means of qualitative input. Classifying the products in a smart and data-driven way will contribute to this better focus. Moreover, as a result of the single KPI strategy it is difficult to evaluate the forecasts carefully. KPIs based also on the predictability is therefore desired. This helps to evaluate the results more specific. We limit the problem cluster with being the end problem the relatively low forecast accuracy. For the consequences of a high or low forecast accuracy we refer to Section 1.2.1 and Section 3.8.

This problem touches many stakeholders at Wavin, since it has far reaching consequences. The research is commissioned by the supply chain management and focuses on the demand forecasts. The direct stakeholders are therefore the Demand Managers, who are the people who make the monthly forecasts. Besides, the forecasts are also based on qualitative data, which is provided by the Sales & Marketing. Therefore, these departments are also important stakeholders, since they make the judgmental adjustments to the statistical forecasts. The project will also have influence on the production planning, since better forecasts lead to a more stable production and less backorders. The last important group of stakeholders we name here, is the procurement department. They purchase the needed raw materials for the production, based on the forecasts.

## 1.4 Research scope and goal

As we explain in Section 1.2, the research is about the improvement of the second step of the S&OP (Figure 3), which is about demand planning and more precisely about the increase of the forecast accuracy. This ongoing implementation of S&OP consists of all the countries where Wavin is active. However, to investigate all these markets is beyond limits in the time the research can be done. Therefore, we will focus on the product sales in three countries: Country A, Country B and Country C. While Country C and Country A have about 4.000 SKUs, Country B has over 7.000 SKUs. We chose these countries, since they have different market characteristics and varying forecast accuracies.

There are three different types of markets. The first is the so-called ‘over-the-counter’ (OTC) sales. These are the sales to the merchants, which consecutively sell the products to installers. The installers install the products for the end user. The second group consists of the project-based sales. These project-based sales are in general the sales to major projects executed by e.g. the government or water authorities. These sales are often dependent on whether a tender will be won. The third category is the export and consists of the sales to countries other than where Wavin is situated. These are also often sales to major projects commissioned by the government or water authorities.

For the research we need to examine what the differences between these types of markets are and what impact it has on the forecastability. Besides, we need to analyze what types of markets are dominant in each country and what effect it has on the forecast accuracy. The forecasts need to be put in context. In this way, we can clarify the differences between the forecast accuracies. Moreover, by analyzing different markets, it makes it possible to give better recommendations for how to use the model in other contexts.

For this research we also test statistical forecast methods for the product groups to give guidelines for Wavin which method to use in which case. To limit the scope, we only take a selected number of the statistical forecast methods that are present in literature. Research about trying to find the best statistical fit to the data has already been going on for decades, which resulted in dozens of different methods. We will limit ourselves by using rather simple extrapolation methods that are widely used in practice. This includes the methods available in the ERP module SAP IBP, which Wavin wants to use in future. Not included in SAP IBP, but which we include in this research is the naïve with drift and Theta methods. The first is easy simple to implement, understand and use. The Theta method showed to perform very well in the M-competition (Makridakis & Hibon, 2000). These M-competitions are a series of open competitions, intended to compare and evaluate the accuracy of different forecasting methods. In Section 3.4, we explain all these methods by literature.

Using other, more complex forecasting methods, will not be beneficial. Literature (e.g. Rasmussen; 2004) argues that complex forecasting methods often result in overfitting. Moreover, Green & Armstrong (2015) argue that the forecasting methods should be understandable by the user.

To be able to design the differentiated forecasting process, it is necessary to make a good product segmentation, with respective to the markets. Upon next we need to determine which strategy is best for which categorization. This helps to get the focus right for choosing between the quantitative and qualitative data for forecasting. Then, for both the statistical (quantitative) forecast and the judgmental (qualitative) forecast we need to deep dive to see how these forecasts can be improved. The research objective can therefore be stated as follows:

***Research objective****: By means of product and market segmentation, designing a differentiated forecasting and demand planning strategy based on both quantitative and qualitative data for improving the forecast accuracy.*

## 1.5 Research questions

We can translate the formulated research objective into a knowledge problem, which is the main research question. We formulate the main research question as follows:

***Main Research question:*** *How can Wavin improve its forecast accuracy for different types of products for different markets by putting the right focus on quantitative and qualitative methods?*

When we give a comprehensive answer to this question, proved with a quantitative analysis, the research can be considered as successful. In order to give structure to the report and to divide this main research question, we set up multiple sub-questions. We answer each of these questions by a chapter. We state the sub-questions as follows:

1. How is the current forecasting process designed? (Chapter 2)
    a. How are the statistical forecasts made?
    b. How is the qualitative input added?
    c. What classification and grouping are used?
    d. What is the current forecasting performance?
2. What is written in literature about forecasting? (Chapter 3)
    a. How is a forecasting process designed?
    b. Which classification methods can be used?
    c. Which forecast methods are suitable?
        i. Which statistical methods for which situation?
        ii. Which qualitative data is needed and how to implement?
        iii. How can the forecasts be measured and validated?
        iv. How to choose the best forecast method?

   d. What are the implications of better forecasts?
3. How to classify the products based on which parameters? (Chapter 4)
   a. Which products are in which category depending on which parameters?
   b. Which class should have which accuracy target?
4. How to design the forecasting process according to the classification? (Chapter 5)
   a. Which forecast methods (quantitative/qualitative) to use for which categorization and for which products?
   b. Which adjustments to make in the current forecasting process in order to incorporate the new classification method?
      i. How to do the focus on statistical forecasting?
      ii. How to do the focus on judgmental forecasting?
5. What are the results and improvements of the new forecasting process? (Chapter 6)
   a. What are the improvements of the new designed forecast model?
   b. How to implement it to the other countries?

Answering these five questions leads up to give a thorough and structured answer of the main research question. We answer the main research question in the conclusion (Chapter 7) of the thesis. In this final chapter we also explain how the research contributes to literature and practice and we give recommendations for further research. The research framework, illustrated in Figure 5, summarizes the steps that are made for answering the five research questions.

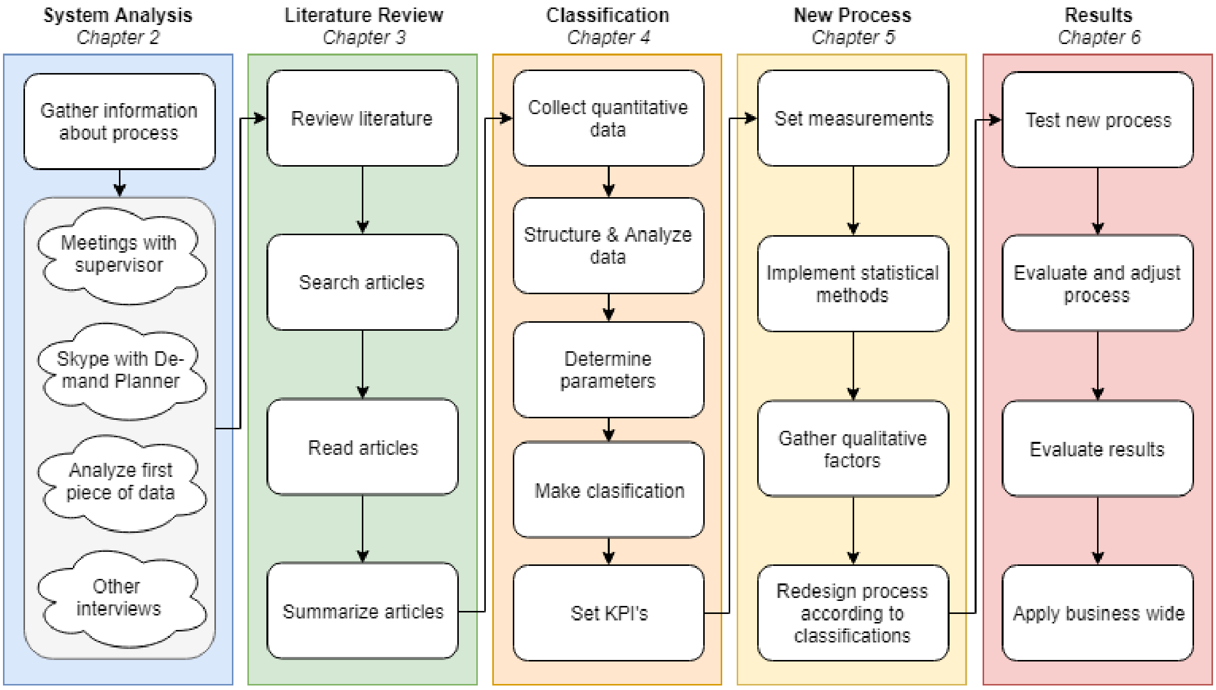



*Figure 5: Research framework*

# 2. Current forecasting process

Although there are many guidelines about how to make the forecasts, each company needs to design their own forecasting process that fits into their business. This makes the forecasting process dependent on the company. Therefore, it is key to describe the current forecasting process at Wavin, in order to know where improvements can be made and how to implement these improvements. The way the forecasts are made, logically influences highly the performance of the forecasts as we quote Warren Buffet:

*"Forecasts usually tell us more of the forecaster than of the forecast."*

Even though forecasts may tell us something about the forecaster as the quote illustrates, it is still necessary to describe the current forecasting process at Wavin. In Section 2.1, we explain the overall forecasting process with the monthly activities. In Section 2.2, we explain the different groupings Wavin uses for handling the thousands of SKUs. In Section 2.3 and Section 2.4, we explain respectively the quantitative and qualitative forecasts in more detail. Section 2.5 is about the current ABC classification Wavin uses. Section 2.6 begins with some market characteristics of Country A, Country B and Country C, after which we discuss the current forecasting performance of the three markets. The chapter ends in Section 2.7 with the conclusion.

## 2.1 Overall process

Every month, new forecasts need to be made. At Wavin, they currently only make medium-term forecasts with a horizon of 18 months, having monthly time-buckets. The seven Demand Managers are responsible for making the forecasts for all the countries. Throughout the past years there is tried to give more and more guidelines for how to do these forecasts in order to have a more generalized forecasting methodology throughout the countries. However, because of different market characteristics or cultural differences these methodologies may differ depending on the country. We discuss the overall processes which apply for all countries, with some more market characteristics for the countries Country A, Country B and Country C, which are the countries in the scope of this research.

### 2.1.1 Monthly schedule

There are eight monthly steps for making the forecasts as shown in Table 3. Steps 1-3 are the Pre-demand phase, step 4 is the Forecast meeting phase, steps 5&6 are the Demand meeting phase and steps 7&8 are the Post-demand phase. The Demand Manager is responsible for each step, but share its responsibility with Sales & Marketing in step 3, 4 and 6.

*Table 3: Monthly activities for making the forecast*

| Step | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Phase | Pre-demand | Pre-demand | Pre-demand | Forecast meeting | Demand meeting | Demand meeting | Post-demand | Post-demand |
| Responsible | Demand Manager | Demand Manager | Demand Manager/Sales & Marketing | Demand Manager/Sales & Marketing | Demand Manager | Demand Manager/Sales & Marketing | Demand Manager | Demand Manager |
| Process step | Run statistical forecast | Align forecast | Preperation of base forecasting file and market intelligence | Agree new forecast with Sales & Marketing | Adjust demand plan based on forecast meeting | Run demand review meeting | Generate consensus demand plan | Finish demand review process and prepare for next month |
| TimeLine | 1st day of month | 3rd day of month | 5 days before Demand Review | 4 days before Demand Review | 3 days before Demand Review | Friday closest to 15th of month | Friday closest to 15th of month | After forecast release |

During the Pre-demand phase the base statistical forecast should be generated by the Demand Manager. The first task is to run the statistical forecast in SAP. This is done on aggregated level using planning groups (see Section 2.2). Country A makes the statistical forecasts also on SKU level for the A items. Then this forecast needs to be aligned. This means that the forecast needs to be evaluated when errors in the system occur, e.g. when there is not enough historical data available, the forecasted items are less than zero etc. Besides, the forecast accuracy report for last month needs to be made. The last step of the Pre-demand phase for the Demand Manager is to prepare the forecasting file for the forecast meeting, containing the statistical forecasts. This file is used for adding the qualitative data. Sales & Marketing should gather the market intelligence for the input for the forecast meeting. This market intelligence consists of e.g. marketing events and open orders (see Section 2.4).

During the forecast meeting, the base forecast needs to be adjusted according to the market intelligence. This is done on customer group level and product hierarchy level 4/7/8/9 (we explain these groupings in Section 2.2). After the forecasts are adjusted, the Demand Manager analyses the significant changes compared to last month, identifies potential demand issues or opportunities, and reviews the forecast accuracy. He/she brings this information to the Demand meeting, where he/she shares his/her findings. During this meeting the performance of the forecast accuracy is reviewed as well.

After the final adjustments have been made and there is agreed on the final aggregated forecast, the forecasts on SKU level can be generated. This is based on historical sales proportions of the past six months. This is called the top-down approach (see e.g. Fliedner, 1999; Grunfeld & Griliches, 1960). This should be done the latest on the last Friday before the 15$^{th}$ of the month. The final forecast on SKU level is input for the 29 production plants throughout Europe. Each production plant has his own product portfolio. Therefore, it may happen that the production plants in a country does not produce all products that will be sold in the country. Then the forecasted sales of this country will become an intercompany demand for a production plant in another neighboring country.

The remaining weeks of the month can be used for activities like history cleaning, alerts cleaning, product life cycle and hierarchy review, planning group maintenance and statistical model selection. These activities ensure that from the first day of the next month, the forecasts can be generated again.

This research touches every step of the forecasting process at least to some extent. For easy to predict products, the focus should be on statistics (step 1, 2, 8) but for more difficult products the focus should be on qualitative information (step 3-5). There should be made a good consensus between these two types of forecasts (step 6, 7). However, a complete redesign of the steps will not be the goal. In Chapter 5 we discuss what the implications of the classifications will be for making the forecasts. The ultimate goal is to give the forecasters a better guidance and focus for making the forecasts resulting in a higher forecast accuracy. This forecast accuracy and also the experiences of the forecasters will therefore be reviewed in order to determine how the classification contributes for making the forecasts.

## 2.2 Grouping

Since the forecasts need to be done for thousands of SKUs, Wavin uses different types of grouping. Some groupings are used for making the statistical forecasts, while other groupings are used for implementing the market intelligence. We discuss respectively the product hierarchy, planning groups and customer groups.

### 2.2.1 Product hierarchy

The product hierarchy is the basis of the groupings. A product hierarchy is a model of the hierarchical relationships between the products in a tree structure. The lower the level, the more basic is the description and the more products fall under this level. The product hierarchy of Wavin consists of nine levels as shown in Figure 6.


*Figure 6: Product hierarchy*

This product hierarchy enables the grouping of products and defines the relationships between products and groups at different hierarchy levels. A SKU is defined by all levels of the hierarchy. When a single level is chosen, for example level 5, the products will be grouped according to the material it is made of. Multiple levels can be chosen at the same time as well. For example, level 7/9 means that the products are grouped based on the assortment/brand with the division of pipes and fittings. For the forecasts mainly (combinations of) the levels 4, 7, 8 and 9 are used. When selecting one or multiple levels, the products are grouped and aggregated decisions can be made specific for the group. The level should be chosen such that the decision has effect on all products in the group. Take for example the level 7 assortment *floorheating*. It may happen that the Marketing department plans to make a promotion for floorheating, and where they expect an increase of 10% in the next three months. Then the adjustments can be made on this level. However, there are two sub-assortments of floorheating; *standard* and *manifold* which is on level 8. It may be the case that there is only a promotion happening for the manifold version of the floorheating. Then the adjustment should be made not only on level 7, but together with level 8. This makes analyzing the products on different levels of detail possible.

### 2.2.2 Planning groups

The planning groups are designed based on level 7/8/9 of the product hierarchy and is country specific. However, some level 7/8/9 groups are put together into one group or are split into several groups. The Demand Manager makes these planning groups. The groups are made in such a way that all the SKUs in a group have somewhat the same sales pattern. As a result, one statistical method can be used for one planning group. It is important that the Demand Manager keeps track of these planning groups to see whether the sales of the SKUs are still acting in the same way. The aggregated planning groups determine which statistical methods are chosen, and when the SKUs in a group behave in varying ways, the forecasts can become inaccurate.

### 2.2.3 Customer groups

Customer groups are mainly used for adjusting the statistical forecasts with the market intelligence during the forecast meeting in step 4, where the sales representative is present. Often these sales representatives have information specific about a customer. A customer may indicate that they will buy significantly more or less products in future due to several factors. The forecasts of the concerning customer can be adjusted accordingly.

## 2.3 Statistical forecasts

Currently, the statistical forecasts are made with the ERP system SAP APO (Advanced Planning and Optimization). However, in the near future, Wavin's intention is to implement SAP IBP (Integrated Business Planning), which is a newer module of SAP with new features and which should make the forecasting easier. A feature that SAP IBP enables is making the classification not only on importance (ABC), but also on forecastability (XYZ). This is also where the idea of the research began.

In SAP APO and also in SAP IBP, there are different statistical methods to use for making the forecasts. See Table 38 in Appendix B for the methods available in SAP IBP. The same methods are available in SAP APO except for the (seasonal) naïve, ARIMA methods and Brown's linear regression. For choosing the best method for a planning group the 'pick best' feature in SAP is used. This method fits all methods to the time series and picks the method with the lowest errors (using MAE, see Section 3.6), as the best method. This method is then set as default for making the forecasts every month. The guideline is to carry out the 'pick best' method every quarter in order to determine if still the same statistical method gives an accurate forecast for the time series. Demand patterns may change due to a shift of stage of the production life cycle for example. In the past, the 'pick best' method was used every month for generating the forecasts. Wavin changed this approach, since this resulted that for some groups the selected forecasting method changed from month to month. For one month it could happen that single exponential smoothing was chosen as the best pick, while the next month triple exponential smoothing gave the best fit. Another reason was that it is a sort of 'black box' solution. The Demand Managers run the statistical forecasts and get some outcomes, without knowing how it is done precisely.

Although the statistical forecast can be generated automatically, it is important to clean the history before running the statistical method. It can be the case that because of market events, some months have considerably higher or lower sales than normal. For a good fit of a statistical model, these outliers should first be corrected. By reviewing these outliers and comparing it with the causes of these outliers the historical data can be cleaned and made ready for running the statistical forecasts.

## 2.4 Qualitative input

During the forecast meeting, the statistical forecasts are adjusted with qualitative information. This is done for the customer group or product hierarchy level 4/7/8/9, depending what is desirable. Five types of qualitative data are considered. The first are the open orders. It may happen that some orders may not yet be fulfilled. The prospects whether the orders will be closed before the end of the month or remain open for the coming months needs to be evaluated.

Second are the open projects. Project-based sales can be a big part of the total sales, although they are difficult to forecast. These projects can be occurring very occasionally and it often depends whether a tender will be won. The open quotes of the Customer Relationships Management (CRM) needs therefore be considered. However, the certainty of whether the project will be won or not has a high influence on the sales. Therefore, only projects with a high probability of happening will be considered in the forecast.

The third type of qualitative information that is considered is the wholesaler trend. Sell-out data of the wholesaler can give prospects if there is an increasing or decreasing trend of the sales. This should, therefore, be considered when adjusting the forecast.

The fourth is the market events. Wavin uses a checklist for this, which is given in Figure 31 in Appendix C. This checklist includes sales price management, promotions, marketing campaign, conditional bonus performance, product launch and competitive information. The market events update is done for the next six months.

The last is the sourcing events. Wavin uses a checklist for this as well, which is given in Figure 32 in Appendix C. This checklist includes (raw) material availability, overall raw material price development, phase-in/phase-out, procurement bonus, third party supplier and intercompany supplier. Like the market events, the sourcing events update is done for the next six months.

The Sales and Marketing needs to bring their forecasts according to this qualitative information to the forecast meeting. They discuss during these meetings with the Demand Manager what the final forecast should be, comparing the qualitative forecast with the statistical forecast.

## 2.5 ABC classification

Wavin uses the ABC classification, which classifies the products based on importance. The classification is based on two parameters which are net turnover (NTO) and number of picks (also known as orderliness). The highest 80% of the SKUs of the NTO (also known as revenues) and orderliness is assigned to class A. The next 15% to class B and the remaining 5% to class C. The guideline is that the Demand Managers evaluate the ABC classification twice a year. Sales, and consequently revenues and orderliness, are subject to change and therefore it is important to update the ABC classification now and then. The classification helps the Demand Managers, as well as the production planners to know where to put the focus. For example, for the A products, there can be put more effort for cleaning the data, like outlier correction, making a better statistical forecast. Besides, Country A also runs separate statistical forecasts for the A products and not only on the planning group level. Moreover, for all countries the targets and evaluation of the forecasts are only done for class A products.

## 2.6 Performance

Before we give the details of the current forecasting performance of the three countries, it is good to give some general information of the countries, which is shown in Table 4. Wavin sells many different products having relatively low revenues per SKU. Of the three countries, Country A has the biggest market. Country B is the country with the highest number of different SKUs and planning groups.

*Table 4: Information per country in 2018*

| Country | Sales | Revenues | SKUs | Planning groups | Level 7/8/9 |
|---|---|---|---|---|---|
| Country A | | | 3794 | 33 | 251 |
| Country B | Confidential | | 7055 | 71 | 139 |
| Country C | | | 4147 | 43 | 139 |

The performance of the forecasts at Wavin are evaluated by using the wMAPE. The MAPE as we also explain in Section 3.6, calculates the difference of the actual sales with the forecasted sales relative to the actual sales. The outcome is the percentage of the error. The wMAPE also takes into account the total sales of all products, which makes that the value is calculated relative to the sales size of a certain group. In this way, the MAPE of products with high sales have a bigger impact than the MAPE of products with low sales. The forecast accuracy as Wavin uses is 1 minus the wMAPE. However, it may happen when the sales are relatively low and the forecast relatively high that the outcome gives a minus value. A minus accuracy is not possible and the accuracy is then set to zero. The following formula is used:

$$Forecast\ accuracy = Max\left(\left(1 - \left(\frac{\sum_{t=1}^{N} |x_t - \hat{x}_t|}{\sum_{t=1}^{N} x_t} 100\%\right)\right); 0\right)$$

In Table 5, the forecast accuracies for the three countries are given. Although Wavin only evaluates their A products, we calculated the forecast accuracies over all the products. We believe that all products are of value to evaluate and therefore, with this research, we take all products into account. The final forecast accuracy is when the statistical forecasts are adjusted according to the qualitative information. Only for Country C these qualitative adjustments result in an improvement of the forecast accuracy averaged over 2018. The final forecast accuracy is 2,87% higher than the statistical forecast accuracy. For Country A, adjusting the statistical forecast decreased the accuracy with almost 1%, while for Country B this was a decrease of 0,31%.

*Table 5: Forecast accuracies in 2018*

| Country | Statistical forecast accuracy | Final forecast accuracy | Difference |
|---|---|---|---|
| Country A | 76,34% | 75,35% | -0,99% |
| Country B | 49,90% | 49,59% | -0,31% |
| Country C | 65,85% | 68,72% | 2,87% |

Country A has the highest forecast accuracy with about 75%. This is the result of many 'over-the-counter' sales, which are more stable and easier to predict than project-based and export products which are a higher percentage in especially Country B and to some extent in Country C.

The forecast accuracies per month for the three countries are illustrated using graphs in Figure 33, Figure 34 and Figure 35 in Appendix D. From these figures we can derive that adjusting the statistical forecasts have varying results. There are many fluctuations. For better analyzing we list in Table 6 the minimum and maximum values together with the standard deviation of the differences between the statistical and final forecast accuracy of the 12 months in 2018 for each country.

*Table 6: Min, max and standard deviation of the improvement of the final forecast accuracy of the 12 months in 2018*

| Country | Min (month) | Max (month) | St. dev. |
|---|---|---|---|
| Country A | -9,96% (1) | +13,23% (4) | 6,04% |
| Country B | -19,1% (1) | +22,6% (12) | 10,00% |
| Country C | -6,70% (9) | +46,30% (8) | 13,81% |

A decrease of around 7% for Country C, 10% for Country A and almost 20% for Country B shows that the qualitative input can have significant negative effects on the forecast accuracy. However, for other months this market intelligence means an increase of over 13% for Country A, over 22% for Country B and over 46% for Country C. There needs to be noted that for Country C in August, the factories close, which resulted in the low statistical forecast accuracy and the high improvement of the final forecast accuracy of this month.

Moreover, analyzing the standard deviation, we can conclude that the increase in forecast accuracy fluctuates much when adding the qualitative information. This is not desirable, since then it is more based on luck, rather than of consistent improvements. This makes clear that this research can be of much added value in order to know for which products to focus on statistics and for which on qualitative input. As a result, the sometimes damaging effect of adding the qualitative information can be avoided.

## 2.7 Conclusion

The forecasting process of Wavin consists of eight monthly steps for making the statistical forecasts and adjusting it with the qualitative input. The statistical forecasts are made by the Demand Managers using SAP APO on aggregated level using planning groups. Then, using customer groups and different hierarchical levels Sales and Marketing give input for adjusting the statistical forecasts with market intelligence. Using an ABC classifications helps to give some focus for making the statistical forecasts. Besides, only the forecasts for class A products are taken into account for evaluating.

The forecast accuracies of the three countries show relatively large differences. This is the result of different market characteristics, where there are more difficult to predict project-based sales in Country B for example. Besides, although there is put much effort for adjusting the statistical forecasts, it is not always the case that it means an improvement of the statistical forecast. Given that the standard deviation of the differences between the statistical and final forecasts of the months in 2018 is high, shows that the improvements fluctuate much. By implementing the classification we expect that both the statistical and qualitative forecasts can be increased. This is a result of the fact that the classification makes different approaches and foci possible.

# 3. Literature study

In order to anticipate to the future and for the production to know what to produce, forecasting is the bedrock. However, forecasting is often not easy and straightforward as the Danish physicist Niels Bohr quoted:

*"Prediction is very difficult, especially if it's about the future."*

Forecasts are usually based on historical data, extrapolating or translating it to the future. However, demand is often not very stable and it is subjective to many factors, of which product promotions and seasonality are just some of them. Therefore, forecasts should be based on quantitative as well as qualitative data. In this chapter we give the literature fundament of the research. In Section 3.1 we discuss what factors and criteria are necessary for making a forecast. In Section 3.2 we write about the forecasting process. In Section 3.3 we describe how to classify the products. In Section 3.4 and 3.5, we describe respectively the qualitative and quantitative forecasting methods. Section 3.6 is about measuring the forecast accuracy and Section 3.7 about choosing the right model. In Section 3.8 we put everything in a bigger context and we write about the implications of forecasting. We explain how this research contributes to literature in Section 3.9 and we conclude the chapter in Section 3.10.

## 3.1 Forecasting in general

As the quote of Niels Bohr illustrated, forecasting can be very complex. However, forecasts are an important aid for effective and efficient planning and it is therefore essential to make accurate forecasts. The predictability of an event and therefore inherently the possibility for an accurate forecast is depended on three factors (Hyndman & Athanasopoulos, 2018):

1. Whether and to what extent the factors influencing the events are known
2. The availability of the data
3. Whether the forecasts can affect the actual event

When these three factors can be met to at least a certain extent, forecasting is useful. This is the case for Wavin. Maybe not all, but at least some factors that influence the events are known, like seasons and price changes. Besides, the historical data is available. For new products this may not be the case, but by experience and using the sales of comparable products the sales can be estimated. The last factor is also met, since the forecasts will not influence the actual sales.

For making these forecasts a good forecast model is needed which can be judged by the following criteria (Duffuaa & Raouf, 2015): (1) accuracy, (2) simplicity of calculation, (3) data needed for model and storage requirements, and (4) flexibility.

Accuracy is whether the future events are in line with the forecasts and different methods can be used to calculate the accuracies which we explain in Section 3.6.

Although there has been an increased interest for complex forecasting for the past decades, Green & Armstrong (2015) argue that simplicity in forecasting is the way to do it. In order to make accurate forecasts, the forecasting process needs to be understandable to the user. The user should be able to explain the used forecasting methods, how the model represents prior knowledge, how the different parts are related and how the forecast from the model can help him to make better decisions.

Forecasts can be either short-, medium-, or long-term orientated, with each having their own data requirements (Duffuaa & Raouf, 2015; Hyndman & Athanasopoulos, 2018). Short term is for scheduling personnel, production and transportation and is often on a day-to-day basis with a timehorizon of below 1 month. Medium term is for determining future resource requirements, like purchasing materials, hiring personnel, or buying machines and equipment and is often on weekly or monthly basis with a maximum time horizon of a couple of years. Long term is for strategic planning. These forecasts cover, among others, market opportunities and environmental factors and have in general a time horizon of at least a couple of years. This research is about the mid-term forecasting, where the forecast accuracy is evaluated per month.

The last criteria, which is flexibility, describes the ability to adjust to changes of the forecasting model. It is the degree of robustness of the forecasting model (Duffuaa & Raouf, 2015).

## 3.2 Forecasting process

Silver et al. (2017) created a very useful framework for forecasting, which is shown in Figure 7. Based on historical data, a mathematical model of the forecast can be made, after a model is selected and initiated based on this data. This statistical forecast should be adjusted by human (judgmental) input, whereafter a final forecast of the demand can be created. In order to examine wheter this forecast is a good resemblens of the reality, forecast errors can be calculated. Based on these errors, the mathematical model can be modified accordingly or the human input can be re-examined. This forecasting process is also the basic forecasting process at Wavin as we describ in Section 2.1.

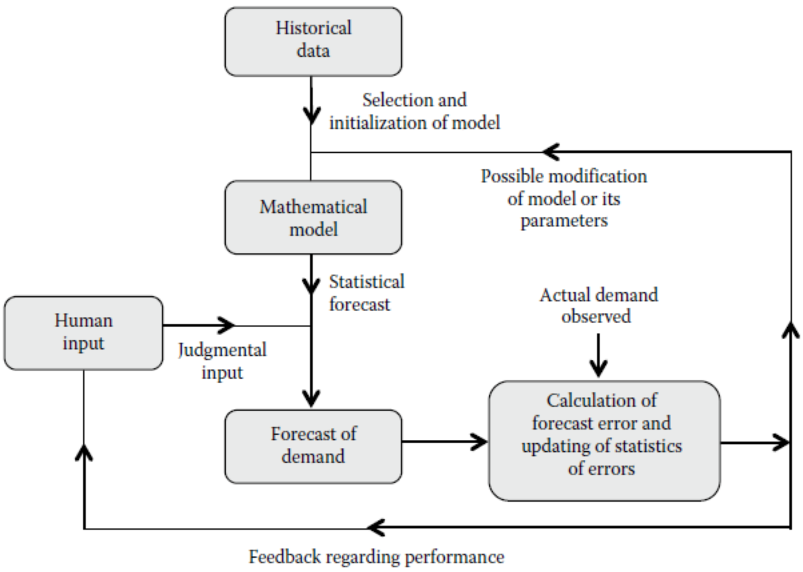



*Figure 7: Process of forecasting (Silver et al., 2017)*

Armstrong et al. (2015) adds on this by describing 28 guidelines to underline the so-called Golden Rule of forecasting: *Be conservative by adhering to cumulative knowledge about the situation and about forecasting methods*. This checklist consists of six subareas which are: (1) Problem formulation, (2) judgemental methods, (3) extrapolation methods, (4) causal methods, (5) combine forecasts from diverse evidence-based methods and (6) avoid unstructured judgemental adjustments to forecasts. The whole checklist is given in Figure 36 in Appendix E. Following this checklist helps forecasters to avoid common mistakes by indentifying and using all relevant knowledge of the problem.

## 3.3 Categorizations

Often, companies are dealing with many individual items. The specific unit which needs to be controlled is called a stock-keeping unit (SKU). It is defined to be a unit which is completely specified according to function, size, color, style and location. This means that many companies would have to control (ten)thousands of different SKUs, which is also the case for Wavin. Family grouping is a common way for dealing with this issue. By grouping them, the decisions can be made on a higher hierarchical level, which requires less decisions. However, it may be useful to have a closer examination on the SKU-level for more important SKUs, instead of these higher hierarchy levels. Besides, although the products are grouped, there still may be many different groups to evaluate. Another classification is therefore desirable. In general, it is the case that a relatively small group of products has a high influence. This principle is first introduced over a hundred years ago by Vilfredo Pareto (1896), by noting the 80/20 connection. Pareto showed that 80% of the land in Country C was owned by only 20% of the people. Others (e.g. Koch, 1997) translated this to the business field and explained that it is also a very valuable principle for companies, handling their products. Using this principle, the SKUs are generally classified into three different groups (A, B, C), with each having different importance. This is also the classification Wavin uses. We discuss this classification after which we describe how to extend this classification in order to also incorporate the forecastability of the products.

### 3.3.1 ABC Classification

The book of Silver et al. (2017) classifies the products according to the total annual dollar usage. This is the value $v$ in dollars per unit multiplied with the annual usage (demand) $D$ for each SKU. First, the SKUs need to be ranked in descending order, starting with the largest value. Then the classification can be made. In class A, which should get the highest priority, the first 5-10% up to 20% of the SKUs should be designated. In general these SKUs account for about 50% of the total annual dollar usage. Class B contains about 50% of the SKUs and account for most of the remaining 50% of the annual dollar usage. This class should also receive some attention, but can be more generalized and automated than class A items. Class C contains the least important SKUs, where all the remaining SKUs should be classified. Decisions for this class need to be kept as simple as possible, and the products should be grouped on high hierarchy levels. Important to note is that the percentages given are just guidelines and are company dependent. For example, Onwubolu & Dube (2006) and Kepczynski et al. (2018) give for the ABC classification, respectively the guideline percentages of SKUs as 20% (80% dollar usage), 30% (15% dollar usage) and 50% (5% dollar usage).

The classification does not need to be made on the annual dollar usage (also known as revenues) alone. Managers may shift some SKUs to a higher or lower class when needed. An example can be that some inexpensive but very crucial SKUs should be assigned to a higher class. A two-digit classification can therefore be useful by also classifying based on the number of customer transactions (Krupp, 1994) or criticality (Flores & Whybark, 1987). These are just a few examples of what is been recognized in literature that the classical ABC analysis alone is not able to provide a good classification in practise (Guvenir & Erel, 1998). For the topic of forecasting, Kepczynski et al. (2018) describe another classification which is useful in combination with ABC. This is called the XYZ classification and is based on the the degree of ability to forecast accurately (based on statistics). We discuss this classification in the next section.

### 3.3.2 XYZ Classification

The XYZ classification is specifically focused on forecasting. It segments the products according to forecast error variation, which describes the stability and predictability of the forecasting process. Using this classification, there can be given different focus, depending on the possibility of extrapolating the historical data by means of a mathematical model. There are different options to make this classification, but Kepczynski et al. (2018) recommends using the wMAPE or the coefficient of variation (CoV). We describe the first measurement more elaborately in section 3.6.2. The CoV is the measure which is defined as the standard deviation to mean ($\sigma/\bar{x}$) and is used in the majority of the literature to make the classification. The thresholds for the XYZ classification are not straightforward to determine and are industry depended. Possible thresholds for the CoV could be for X (< 0,2), Y (0,2-0,3) and Z (> 0,3) (Pandaya & Thakkar, 2016) or also for X (< 0,5), Y (0,5-1) and Z (> 1) (Scholz-Reiter et al., 2012; Sankaran et al., 2019). These values are not very consistent and should therefore be based on field experts. The classification should be made such that the generally stable and predictable products are put in class X. Class Y should consist of the products when this is the case in a lesser extent. For the Z class, forecasting the products with only statistical methods will not work properly, since the demand patterns are more random.

### 3.3.3 Implications of the classifications

Now it is clear what the possible classifications are, we need to describe what the practical implications and use of these classifications are. Each separate class requires a different approach of forecasting, with different people responsible, setting different targets and having different forecasting techniques. Kepczynski et al. (2018) created a figure which summarizes these different aspects and which is shown in Figure 8. As input for the differentiated forecasting, both a statistical forecast and qualitative forecast is necessary, together with the products segmented and defined process measurements.

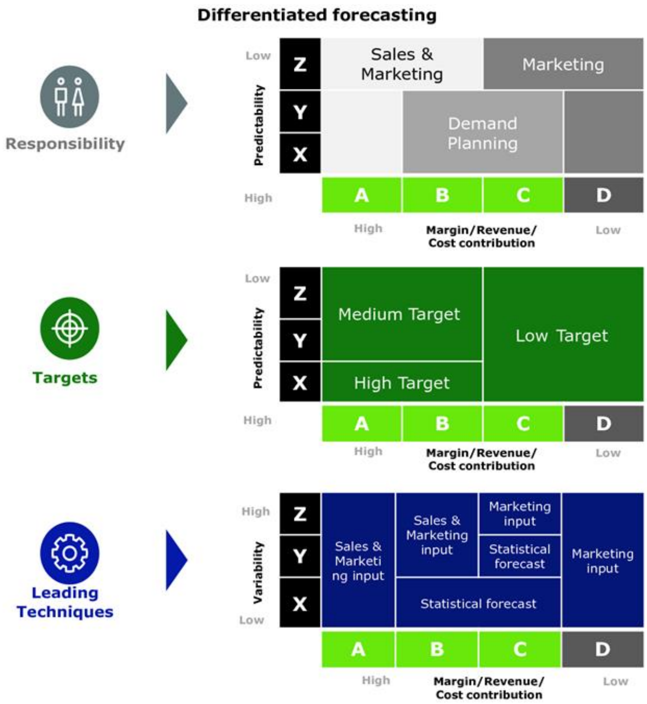



*Figure 8: Differentiated forecasting (Kepczynski et al. 2018)*

#### *3.3.3.1 Leading techniques/Responsibilities*

The basic statistical forecasts are made by the Demand Planning. They have the responsibility to create a valuable forecast of the products. However, they have not the single responsibility which is illustrated in Figure 8. Qualitative data is especially needed for products which are important and where the statistical forecasts fall short. According to Figure 8, statistical forecasts is only sufficient on its own for the product groups BX, CX and CY. These are the products which are relatively easy to forecast and do not have a high impact. This differentiated forecasting method ensures that Sales & Marketing only need to focus on important products, or where the demand sales are more difficult to predict.

#### 3.3.3.2 Forecast accuracy targets

Setting targets is very important in many businesses. By evaluating the reality with the targets, there can be seen if something is not going how it is supposed to be. Having different forecast accuracy targets for different classes, enables a company to evaluate more specifically the outcomes, and therefore the improvements can be made more focused. Products which are important and are relatively easy to forecast should have higher targets than products which are more difficult to forecast and are less important. Currently, Wavin only has a target for the A products, which is 70%. However, this may be that some groups perform really well, while for other groups the forecast accuracy is very low. By setting different forecast accuracy targets, better evaluation is possible.

## 3.4 Quantitative models

The basis of every forecasting methodology are the statistical forecasts, which are based on the mathematical models. Each method considers the data to be time series, which can be either continuous or discrete. Since this research is about demand patterns, we will only discuss discrete time series, with having intervals of months. When plotting time series, one can distinguish up till five characteristics; level ($l$), trend ($b$), seasonality ($s$), cyclical movements ($C$), and irregular random fluctuations ($\epsilon$) (noise) (Silver et al., 2017). The same book describes that using these components, we can formulate a multiplicative model, an additive model or a combination of these two. In the next section we explain the quantitative forecasting methods by combining the methods described by Silver et al. (2017), Duffuaa & Raouf (2015), Hyndman & Athanasopoulos (2018), Markin & Sinha (2018), Hyndman et al. (2008), and Axsäter (2006), unless stated otherwise. We discuss the average, naïve, exponential smoothing, linear regression, ARIMA and Croston methods respectively.

### 3.4.1 Average methods

We discuss four methods based on average, which we explain in this section. These methods assume that the demand can be calculated with the following formula:

$$x_t = l + \epsilon_t$$

This means that the demand is equal to the level with random noise.

The first method is the most simple one which is in the name; *simple average*. This takes the average of a certain number of time series of the historical data and projects it in a constant forecast value. This is applicable for mature products with stable demand. When we define $x_i$ as the historical data for all time periods until the end of period $t$, we can estimate the demand with the simple average as follows:

$$\hat{x}_{t+h} = \frac{\sum_{i=1}^{t} x_i}{t}, for\ any\ h$$

Second is the *weighted average*, which is similar to the simple average, but differs in the way that each observation will be given a different weight $w_i$. In this way, more recent data can be given a higher weight, which then will have a bigger impact on the forecast. The idea behind this, is that recent data are a better estimate for the new events than older data. This method can be used for mature products with stable demand, having some time periods and seasonality. The weights can be determined empirically or based on experience. The formula is as follows:

$$\hat{x}_{t+1} = \sum_{i=1}^{t} w_i x_i \quad with \sum_{i=1}^{t} w_i = 1$$

The *Simple moving average* only differs from the simple average, that when calculating the new events, the time periods change, while for the simple average the time periods are fixed. This method can be used for mature products with stable demand and the formula is as follows:

$$\hat{x}_{t+1} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Here, the average is taken for the last *n* periods. The larger the *n* the more stable will the forecast be and less adaptive to changes. A smaller *n* means the opposite. Typically, the *n* value ranges from 3 to 12 (Silver et al., 2017).

The last average method we describe is the *weighted moving average*, which is a combination of the weighted average and the simple moving average. This method is suitable for mature products with stable demand, having some seasonality. The formula is as follows:

$$\hat{x}_{t+1} = \sum_{i=1}^{n} w_i y_i \quad with \sum_{i=1}^{t} w_i = 1$$

### 3.4.2 Naïve methods

With the naïve method, the forecasts are simply set to the value of the last observation. We can formulate it as follows:

$$\hat{x}_{t+h|t} = x_t$$

This method considers that the data is not seasonal or having a trend. When the data has a seasonal pattern, the seasonal naïve may suit well. This method sets the forecast to be equal to the last observed value from the same season of the previous year. We can formulate this method as follows:

$$\hat{x}_{t+h|t} = x_{t+h-m(k+1)}$$

Where *m* is the seasonal period and *k* the integer part of (h-1)/m. The last naïve method we describe is the drift method. This method allows the forecast to increase or decrease over time. The change over time is set to the average change seen in the historical data. The formula is:

$$\hat{x}_{t+h|t} = x_t + h\left(\frac{x_t - x_1}{t-1}\right)$$

### 3.4.3 Exponential smoothing methods

The beginnings of exponential smoothing methods started as early as in the 1950s, and are untill now still one of the most used forecasting methods in both business and industry. This is because of the fact that these forecasting methods for estimating the parameters are intuitive and easy to understand. The idea of exponential smoothing, what is also in the name,  is that the weights decrease exponentially as the observations get older (Hyndman et al., 2008; Gardner, 2006). This differes from the weigthed (moving) average, where the weights are set by the user. We discuss the single, double and triple exponential smoothing methods. In this section we also discuss the Theta method.

#### *3.4.3.1 Single exponential smoothing*

Single or *simple exponential smoothing* (SES) is the most basic form of exponential smoothing. This method fits well for demand patterns with no clear trend or seasonality. We write the formula as follows:

$$\hat{x}_{t+1} = \alpha x_t + (1-\alpha)\hat{x}_t$$

The forecasted value is based on a weighted recent observation $x_t$ with $\alpha$ and a weighted recent forecast $\hat{x}_t$ with (1 – α). The $\alpha$ is a value between 0 and 1, where typically not higher than 0,5. When $\alpha$ is close to 1, the new forecast will be adjusted substantial for the error in the previous forecast. When close to 0, it has conversely little adjustment. The user has to evaluate what value to choose, but can also be calculated with the following formula to give a good indication (with *n* the moving average period):

$$\alpha = \frac{2}{(n+1)}$$

An extension of this model is called the *adaptive-response-rate-SES.* The idea of this adaptive smoothing is that $\alpha$ can be adjusted overtime when relatively high or low errors occur. This is based on the measurements of MAE and MSE, which we explain in Section 3.6. Although this method is intuitive, multiple research (Chatfield, 1978; Flowers, 1980; Ekern, 1981) found that there is no real proof that adaptive smoothing outperforms the regular smoothing.

#### *3.4.3.2 Double exponential smoothing*

The SES model can be extended when there is a clear trend in the historical data. This method is called *double exponential smoothing* or Holt's linear exponential smoothing to the research of Holt (1957, 2004). With this method, we need to introduce a new constant *β*, which also can take values between 0 and 1, and which is typically not higher than 0,5. This method consists of a forecast equation and two smoothing equiations (level and trend):

$$\begin{aligned} &Level: && l_t = \alpha x_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\ &Trend: && b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\ &Forecast: && \hat{x}_{t+h|t} = l_t + b_t h \end{aligned}$$

When $\alpha$ is equal to *β*, this method is the same as Brown's double exponential smoothing (Brown, 1959), where there is just one smoothing constant. We get the following formulas for level and trend:

$$\begin{aligned} &Level: && l_t = [1-(1-\alpha)^2]x_t + (1-\alpha)^2(l_{t-1} + b_{t-1}) \\ &Trend: && b_t = \left[\frac{\alpha^2}{1-(1-\alpha)^2}\right](l_t - l_{t-1}) + \left[1 - \frac{\alpha^2}{1-(1-\alpha)^2}\right]b_{t-1} \end{aligned}$$

Sometimes it can be the case that when there is a trend, one believes that the increase in sales will not continue on the long run. The exponential smoothing method can be adjusted by applying the damped trend method to dampen the trend as the length of the forecast horizon increases. Gardner Jr. & McKenzie (1985) proposed such modification of the Holt's method. Before we give the equations, we need to introduce the damping parameter $\varphi$, which can take values in the range of 0 till 1. The closer this parameter is to 1, the shorter is the period over which the trend is dampened. When this value is 1, the method is the same as the Holt's linear method. The damped trend model is as follows:

$$\begin{aligned} &Level: && l_t = \alpha x_t + (1-\alpha)(l_{t-1} + \varphi b_{t-1}) \\ &Trend: && b_t = \beta(l_t - l_{t-1}) + (1-\beta)\varphi b_{t-1} \\ &Forecast: && \hat{x}_{t+h|t} = l_t + (\varphi + \varphi^2 + \ldots + \varphi^h)b_t \end{aligned}$$

#### *3.4.3.3 Triple exponential smoothing*

The double exponential method involves both trend and level. However, demand shows often seasonality patterns. Therefore, the method needs to be enhanced with a third smoothing parameter, which the *triple exponential smoothing* method, or better known as the Holt-Winter's method, includes (Winters, 1960). For describing this method, we need to make use of the distinction between the additive and the multiplicative approach. The equations of the multiplicative method are as follows:

$$Level: \quad l_t = \alpha \frac{x_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1})$$

$$Trend: \quad b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$

$$Seasonal: \ s_t = \frac{\gamma x_t}{(l_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}$$

$$Forecast: \ \hat{x}_{t+h|t} = (l_t + b_t h)s_{t-m+h_m^+}$$

The variable *m* stands for the length of seasonality, which are often weeks or months. To improve the quality of the seasonal factor, it is desirable to have data of several seasons. Besides, the *y* parameter is introduced, which has a value between 0 and 1. Finally, it applies that $h_m^+ = [(h-1) \ mod \ m] + 1$.

The additive method differs from this and can be calculated with the following set of equations:

$$Level: \quad l_t = \alpha(x_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$$

$$Trend: \quad b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$

$$Seasonal: \ s_t = \gamma(x_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$$

$$Forecast: \ \hat{x}_{t+h|t} = l_t + b_t h + s_{t-m+h_m^+}$$

As you may have noticed, the formula for calculating the trend is for both methods (additive and multiplicative) equal. The only difference lay in that the seasonal indices are now added and subtracted instead of multiplied and divided. In practice, the multiplicative method is used more, since it normally gives a better fit to the data.

#### *3.4.3.4 Theta method*

We discuss the Theta method also in the section of exponential smoothing, since it is equivalent to the simple exponential smoothing with drift. Hyndman & Billah (2003) showed that when the drift parameter is half the slope of the linear trend fitted to the data, the method performs the same as the Theta method. This method performed very well in the M3-competition (Makridakis & Hibon, 2000), and is therefore of great interest for forecasters.

#### *3.4.3.5 Estimation of parameters*

For the exponential smoothing, there needs to be used of up to four constants (α, β, ϒ, φ). Although they have a high influence on the outcomes, these parameters need to be determined by the user. This requires understanding of the methods and even then, there are no clear overall optimal values of these parameters. It depends on the data the method is fitted on and using search experiment the optimal values can be determined. However, some guidelines and ranges can be given (Hyndman et al., 2008). These are as follows:

$$0 < \alpha < 1, \quad 0 < \beta < \alpha, \quad 0 < \gamma < 1 - \alpha, \quad 0 < \varphi < 1$$

Silver et al. (2017) also give some guidelines for choosing values for the parameters. They argue that for single parameter model, $\alpha$ is likely to be in the range of 0,01-0,30 with a compromise value of 0,1. In Table 7 the guidelines for the parameters are given for exponential smoothing.

*Table 7: Guidelines for parameters*

| | α | β | γ |
|---|---|---|---|
| Upper end | 0,51 | 0,176 | 0,5 |
| Single value | 0,19 | 0,053 | 0,1 |
| Lower ends | 0,02 | 0,005 | 0,05 |

### 3.4.4 Linear regression models

The linear regression model includes one or multiple factors influencing the outcomes. The simplest form of a regression model is called *simple linear regression* and can be calculated by:

$$\hat{x}_t = \beta_0 + \beta_1 t + \varepsilon_t$$

This can be understood as the trend model, with $\beta_0$ the level and $\beta_1$ the factor that is the trend. These parameters need to be set such that they give the best fitted line through the datapoints. This fit can be determined by minimizing the sum of squares $S$, which can be calculated by:

$$S = \sum_{t=1}^{n} (x_t - \beta_0 + \beta_1 t)^2$$

The simple linear regression model can be extended when multiple factors have impact on the sales. An example of this is including the seasonal patterns.

### 3.4.5 ARIMA

Other types of widely used forecasting models are the *autoregressive integrated moving average* (ARIMA) models. These models can handle correlated stochastic demand variations and have been developed by Box & Jenkins (1970). There is a large variety of these models. The notation ARIMA(p, d, q) is often used and the different parts stand for:

$$AR: \quad p = order\ of\ autoregressive\ part$$

$$I: \quad d = degree\ of\ differencing\ involved$$

$$MA: \quad q = order\ of\ the\ moving\ average\ part$$

We discuss in the next sections each part of an ARIMA model separately.

#### *3.4.5.1 Autoregressive (AR)*

In contrast with multiple regression models where the variable is forecasted using a linear combination of predictors, with autoregression models this is based on past values of the variable. An autoregressive model of order $p$ can be written as follows:

$$x_t = c + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t$$

Where $c$ is the average of the changes between consecutive observations.

Autoregressive models are flexible at handling a wide range of different time series patterns by changing the $\varphi$ parameters. For AR(1) and AR(2) the following constraints are applicable:

AR(1): $-1 < \varphi_1 < 1$

AR(2): $-1 < \varphi_2 < 1$, $\varphi_1 + \varphi_2 < 1$, $\varphi_2 - \varphi_1 < 1$

It gets complicated when the order becomes higher and therefore one should limit the use of a high order.

In order to execute the autoregressions a step needs to be preceded. The data needs to be made stationary, which can be done by means of differencing.

#### *3.4.5.2 Differencing (I)*

An important aspect of ARIMA models is that the time series should be stationary. Stationary means that the time series have no predictable patterns in the long-term. Trend and seasonality should, therefore, be excluded. This is because with ARIMA modeling we assume that the series started up in the infinite past. A way to do this is by means of differencing. This is the concept of computing the differences between consecutive observations to make a non-stationary time series stationary. In this way, it stabilizes the mean of the time series by removing changes in the level of a time series and therefore eliminating trend and seasonality. The first order difference is formulated as follows:

$$x'_t = x_t - x_{t-1}$$

When the differenced series is white noise, we can formulate the so called “random walk” model:

$$x_t = x_{t-1} + \varepsilon_t$$

Sometimes it can be the case that the data does not become stationary after differencing once. The second-order differencing can therefore be used with the following formula:

$$\begin{aligned} x''_t &= x'_t - x'_{t-1} \\ &= x_t - 2x_{t-1} + x_{t-2} \end{aligned}$$

In order to determine objectively whether differencing is necessary, a unit root test can be executed. There are multiple unit root tests, of which KPSS designed by Kwiatkowski et al. (1992) is one of them. This method sets the null hypothesis as trend stationarity, and there needs to be found evidence to reject this hypothesis.

#### *3.4.5.3 Moving Averages models (MA)*

A moving average model uses past forecast errors in a regression-like model. We can formulate the model with parameter *q* as follows:

$$x_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

While subjective to the following constraints:

AR(1): $-1 < \theta_1 < 1$

AR(2): $-1 < \theta_2 < 1$, $\theta_2 + \theta_1 > -1$, $\varphi_1 - \varphi_2 < 1$

With again more complications if the order gets bigger.

#### *3.4.5.4 Bringing it together*

Now we explained each separate aspect of the ARIMA models, it is time to bring them together. For simplicity we give the formulation for the first order differencing model with ARIMA(p, 1, q):

$$x'_t = c + \varphi_1 x'_{t-1} + \varphi_2 x'_{t-2} + \cdots + \varphi_p x'_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

#### *3.4.5.5 Seasonal ARIMA*

The before described ARIMA model does not take into account the seasonality. However, this is often wanted. By including additional seasonal terms we get ARIMA(p, d, q)(P, D, Q)$_m$. In this, the *m* is the number of observations per year. As an example we give the formula for the most commonly used seasonal ARIMA(0, 1, 1)(0, 1, 1)$_{12}$:

$$x_t = x_{t-12} + x_{t-1} - x_{t-13} - \theta_1 \varepsilon_{t-1} - \Theta\varepsilon_{t-12} + \theta_1 \Theta \varepsilon_{t-13}$$

Where we introduce the $\Theta$, which is the coefficient for seasonal moving average (SMA).

### 3.4.6 Croston method

It may happen that an item is demanded vary rarely, even though the quantity may be relatively high. This data which contains many zeros, is called intermittent or sporadic demand. The previous described methods do not fit well with this type of demand. When there is demand occurring, the forecast will increase, while when there is no demand, the forecast will decrease. With high parameters, the demand will be relatively stable, but it also reacts slow to demand changes. Croston (1972) developed a relatively simple forecast method for dealing with such situations. This forecast method seperates the demand time series into two components; the time between two consecutive events and the size of the events. The forecast is updated when there is an event occuring and applies simple exponential smoothing to both components. The point forecast is the ratio of the forecasts of the positive demand and the time gap. Before we give the method we need to define the following:

$$n_t = number\ of\ periods\ since\ the\ preceding\ positive\ demand$$

$$\hat{n}_t = estimated\ average\ of\ the\ number\ of\ periods\ between\ two\ positive\ demands\ at\ the\ end\ of\ period\ t$$

$$\hat{d}_t = estimated\ averge\ of\ the\ size\ of\ a\ positive\ demand\ at\ the\ end\ of\ period\ t$$

As we explained there are two different situations and the variables are updated as follows:

$$(i) \qquad x_t = 0 \qquad there\ is\ no\ demand\ occuring$$

$$\hat{n}_t = \hat{n}_{t-1}$$

$$\hat{d}_t = \hat{d}_{t-1}$$

$$(ii) \qquad x_t > 0 \qquad there\ is\ demand\ occuring$$

$$\hat{n}_t = (1-\alpha)\hat{n}_{t-1} + \alpha n_t$$

$$\hat{d}_t = (1-\alpha)\hat{d}_{t-1} + \alpha x_t$$

The parameter $\alpha$ has the value in the range of 0 and 1. We can calculate the forecasted demand per period as follows:

$$x_{n+h|n} = \frac{\hat{d}_t}{\hat{n}_t}$$

## 3.5 Qualitative forecasting

As we make clear in Section 3.2, where we describe the overall forecasting process and in Section 3.3, where we discuss the different forecasting approaches for different classifications, qualitative or human input is needed for a good forecast. This qualitative data, better known as judgmental input, ensures that market intelligence will be added, which the statistics cannot provide. Fildes & Goodwin (2007) emphasized on the judgmental influence by surveying 149 forecasters. The outcomes of the survey showed that a statistical forecast adjusted by judgmental input is preferred over merely judgment or statistics or an average of these two. Besides, most of the respondents reported an accuracy improvement of 5-10%, due to the qualitative adjustments to the quantitative forecasts. Lawrence et al. (2006) argue that this improvement is possible when the forecaster has important domain knowledge and when he has more timely, up-to-date information. This underlines the importance of the qualitative input for generating the forecasts. In this section we discuss first some key principles for adjusting the statistical forecasts with human input, where after we explain what data is needed. Then we explain the guidelines for how to implement the human input. In the last section we write about the Delphi method, which is a well-known method for making a forecast, based on judgmental input.

### 3.5.1 Key principles

When adjusting the mathematical model with the qualitative data, it is important to take some key principles into account. We discuss some of these principles here including structuring, documenting, grouping, unifying and evaluating.

#### *3.5.1.1 Structuring*

For avoiding common made mistakes, it is important to be aware of the forecasting challenges. Besides, the forecasting tasks need to be clear to everyone. The definitions should be clear to everyone, grasping the whole meaning. Ambiguity or unclear expressions should be avoided at all costs. Checklists are an important mean of categorizing the information and helps for a systematic approach (Hyndman & Athanasopoulos, 2018). Many other researches have also found that structured forecasting processes lead to improvements in accuracy (Lawrence et al., 1985; Vanston, 2003; Harvey, 2007). This is the case in particularly for when judgemental adjustments are required (Marmier & Cheikhrouhou, 2010).

#### *3.5.1.2 Documenting*

For adjusting the statistical forecasts according to the judgmental input, a meeting with the stakeholders is necessary. The demand planner needs to sit around a table with different managers from among others, sales, marketing and production (Silver et al., 2017). It is important that every stakeholder prepares well for these meetings and bring their own forecasts. Important is that the managers justify their forecasts in writing. This ensures that the judgements will be made more carefully and helps the managers to reflect. Moreover, it forces the managers to rethink whether the judgement is appropriate and helps to avoid circumstances where the potential is questionable. Goodwin (2000) found that requiring the managers to justify their judgements in writing decreases the number of unnecessary or damaging judgmental adjustments from 85% to 35%. In result, the median absolute error was reduced from 10% to 3,6% on average.

#### *3.5.1.3 Grouping*

Since a company has often thousands of SKUs, these judgmental forecasts and meetings is only possible by means of family grouping. Besides, using the ABC-XYZ analysis also helps to focus where to use the judgmental input as we explain in Section 3.3, and which is also argued by Petropoulos et al. (2018).

#### *3.5.1.4 Unifying*

During the forecast meetings it is crucial that everybody is heard, and people have the same influence. Different managers may want to influence the forecast in directions that may benefit their own agenda (Shapiro, 1977). A sales manager may try to increase the forecast in order to ensure product availability, since he will be judged according to sales commissions. On the other hand, the operations manager is responsible for managing the supplies, operating capacity and inventories and will try to influence the forecast in a way that smoothens the demand. Therefore, a possible way for unifying is to average these different forecasts (Lawrence et al., 1986). A smarter way to do this is using an algorithm (Armstrong, 2001; Oliva & Watson, 2009). By weighting the sales' forecast more heavily in the short-term and the operations' forecast more heavily in the long-term, the influence of each manager can be made depended on the time horizon.

##### *3.5.1.5 Evaluating*
An import aspect of judgmental forecasting is to review to what extent they improve the statistical base forecasts. Shockingly, only 44,3 percent of respondents of the survey of Fildes & Goodwin (2007) said they did this. Reviewing helps to assess the adjustments made. Organizations can learn which adjustments were beneficial and which should be avoided. Moreover, it gives also insights whether the adjustments were too optimistic or pessimistic, which is valuable information for the future. Considering that the judgmental forecasts are normally made at lengthy meetings involving much time and effort, it is crucial for good evaluation. Using multiple error measures will help to make these assessments.

### 3.5.2 Types of qualitative data
It is very important to identify all the judgmental factors which have an influence on the sales. Literature describes many of these factors. Figure 9 lists the outcomes of the survey, carried out by Fildes & Goodwin (2007) for investigating which reasons are perceived to be the most important for adjusting the forecasts.

| Reason for using judgment | Percent indicating important or extremely important |
|---|---|
| Promotional and advertising activity | 48.3 |
| Price change | 44.3 |
| Holidays | 37.6 |
| Changes in regulations | 32.9 |
| Insufficient inventories | 30.9 |
| Government policy | 28.2 |
| Activity by competitors (promotions, advertising, etc.) | 24.2 |
| Substitute products produced by your company | 18.8 |
| International crises | 18.8 |
| Weather | 18.8 |
| Insufficient inventories of competitors | 17.4 |
| Sporting events | 11.4 |
| Strikes | 9.4 |
| Other | 10.1 |

*Figure 9: Reasons for judgmental adjustment and their importance (Files & Goodwin, 2007)*

From this figure, we can derive that promotional and advertising activity, and price changes are considered to be the most important judgmental factors. However, the whole list should be considered. Although, there needs to be mentioned that some of the factors are sector dependent (Lee et al., 1990). Besides, it is good to give some more remarks on Figure 9. For example, not only a possible international crisis should be monitored, but it is useful to keep track of the general economic situation, focused on the market of the company as well. The expectations of an economical growth or decline should be considered. Holidays can be evaluated for the human adjustments, but it is often already visible in the seasonal patterns of the data, assuming that the holidays are at the same time each year. The government policy includes the regulations like subsidies, pollution restrictions, import duties and quotas and safety standards. An aspect that is not included in the figure but can have an influence, is the engineering changes that improves the reliability of the product, which reduces the demand for service parts (Silver et al., 2017). The same authors distinguish the factors into internal and external factors. Internal factors are influenced by the company itself and include promotions or price change. External factors happen from outside the company and include changes in regulations, government policy or activity by competitors.

Most of the named qualitative methods are also considered for the qualitative adjustments at Wavin. However, the reasons for judgements which are not really considered at Wavin, but which we think can be of added value are change in regulations, government policy and international crisis.

### 3.5.3 Implementing the judgmental factors

When the factors are known for each product(family), it needs to be implemented to the statistical forecast. The first step is to determine the intensity of the influence of the judgmental factor which needs to be done by an expert. Then there needs to be determined what kind of effect the factors have. The judgmental factors can be of one of the four types: transient, transferred impact, quantum jump or trend change (Marmier & Cheikhrouhou, 2010). The transient factors only influence the data during the period an event is happening. When the event is over, the effect is also not lasting anymore. An example of this is a strike and a typical pattern is shown in Figure 10. Transferred impact factors are factors when the impact of an event is transferred from one period to another, without changing the forecasts of the consecutive periods. An example of this is an announced price change where the expected price change is compensated in the following period (Figure 11). When an occurring non-repetitive event results in a permanent down or up change, it is called a quantum jump factors. An example for this situation when a promotion is done (Figure 12). The last type of factor, which is called trend change factor results in a change of the demand by a percentage. A price variation may result in such a trend modification (Figure 13).

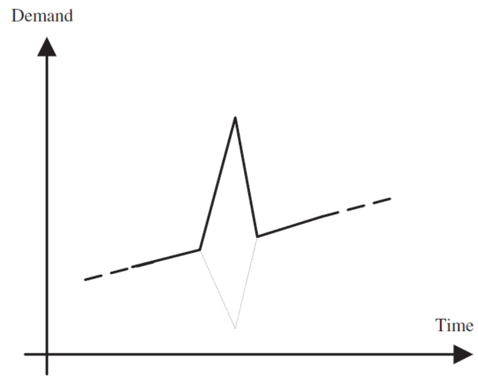



*Figure 10: Transient factor (Marmier & Cheikhrouhou, 2010)*

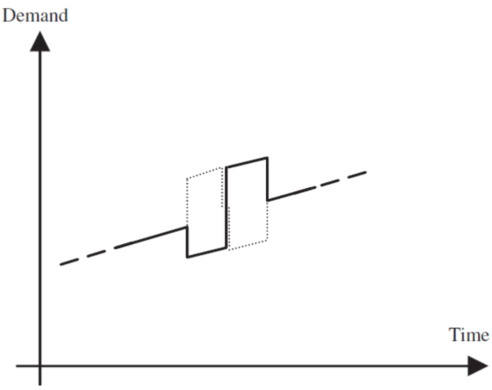



*Figure 11: Transferred impact factor (Marmier & Cheikhrouhou, 2010)*

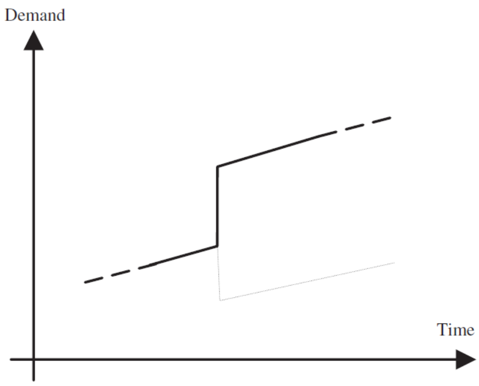



*Figure 12: Quantum jump factor (Marmier & Cheikhrouhou, 2010)*

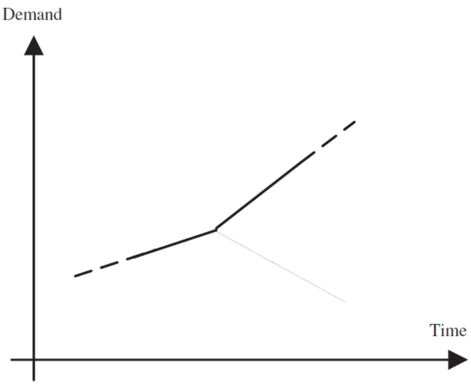



*Figure 13: Trend change factor (Marmier & Cheikhrouhou, 2010)*

### 3.5.4 Delphi method

It can be the case that judgmental forecasting is the only option. There is no historical data available, a new product is being launched, there are completely new and unique market conditions or there is a new (large) competitor entering the market. These can all mean that a statistical forecast will not work, and people need to come up with their own forecast. A widely used method for doing this is the Delphi method (Rowe & Wright, 1999; Rowe, 2007). This method was developed in the 1950s by the RAND Corporation. It is based on the expectation that the forecasts of a group are better than those of individuals. Moreover, it aims to give everyone an equal voice, avoiding the big influence of the dominant individuals. The method is based on a multi-round survey. The first task of the facilitator is to gather a group of experts from different fields who can contribute to the forecast. These experts are asked to generate their own forecasts, with the necessary comments. Then the facilitator collects these forecasts and gives feedback to the experts including a summary of the statistics and the outlines of the qualitative justifications. Based on this feedback, the experts are asked again to make new forecasts. This process can be repeated until a satisfactory forecast is generated. Although this method often results in a convergence of the forecasts of the experts, it also has some limitations. First of all, it is very time consuming, especially in the case when experts have very different views. This may also result in a loose of interest and commitment.

## 3.6 Forecast error measurements

In this report we already mentioned several times the forecast accuracy. The final goal of the research is to increase this forecast accuracy. In Section 3.8 we discuss why it is so important to have a high forecast accuracy. By accuracy, we mean a measure of how far the point forecast is from the actual value. Calculating this may seem reasonable easy, but one value cannot tell the whole truth. It is the challenge to separate the dispersion of the forecast errors with the bias. Dispersion tells something about how widely the errors are spread, while the bias gives the average. The basic formula for a forecast error at period *t* is as follows:

$$e_t = x_t - \hat{x}_t$$

Although being not a valid measure of the effectiveness of a forecasting model, the sum of all the errors is a measure of the bias and can be calculated with the following formula:

$$\sum_{t=1}^{N} e_t = \sum_{t=1}^{N} (x_t - \hat{x}_t)$$

The issue with using this formula is that the positive errors counterbalance with the negative errors. Only when there is a considerable high or low value, this means something. Then the forecast is over- or underestimated which should ring some bells. However, this is not always the case and we are also interested in other error measurements. These are based on taking the absolute or squared values. There are different types of measurement with often made a distinction between scale dependent and scale independent forecasting measurements. We discuss also a third category, which is about the relative error.

### 3.6.1 Scale dependent

Scale dependent measures are suitable for single time-series. With these measurements, one can calculate the errors of the time-series of each SKU. These measurements are useful for comparing different methods applied to the same set of data. The RMSE and MSE are popular scale dependent measurements, with often the preference for RMSE, since it is on the same scale as the data (Hyndman & Koehler, 2006). However, the MAE (also known as MAD) is less sensitive to outliers.

1. Mean absolute error (MAE)

$$MAE = \frac{\sum_{t=1}^{N} |x_t - \hat{x}_t|}{N}$$

2. Mean-squared error (MSE)

$$MSE = \frac{\sum_{t=1}^{N} (x_t - \hat{x}_t)^2}{N}$$

3. Root mean-squared error (RMSE)

$$RMSE = \sqrt{MSE}$$

### 3.6.2 Scale independent

When calculating the overall accuracy of multiple products together, a scaled measurement is needed (Chatfield, 1988). In this way the accuracy will not be highly influenced by products with high demand. Examples of scale independent measurements are MAPE, MSPE and WMAPE. Fildes & Goodwin (2007) and McCarthy et al. (2006) argue that MAPE is the most commonly used measurement in the companies with a percentage of around 50%. However, this MAPE has also some disadvantages. When there are events where there is (almost) no demand, the measure will be highly influenced, since the denominator will be (close to) zero. This method is therefore not suitable for intermittent demand, and these time-series should be omitted for calculating the MAPE. Besides, the measurement has the characteristic that there is a higher penalty on positive errors than on negative errors. This is the case, since there is a lower bound (max 100%) but not an upper bound. Makridakis (1993) came up with the symmetric MAPE, which resolves this problem. Another adjustment to the MAPE is the weighted MAPE. With the WMAPE, the values are weighted according to their demand size. A higher demand means a higher weight. In this way product groups with low average sales have a lower impact on the overall value than product groups which have high average sales. WMAPE is the measure Wavin uses, and it defines the forecast accuracy as 1 minus the WMAPE.

4. Mean absolute percent error (MAPE)

$$MAPE = \frac{100\%}{N} \sum_{t=1}^{N} \left[ \left| \frac{x_t - \hat{x}_t}{x_t} \right| \right]$$

5. Mean-squared percent error (MSPE)

$$MSPE = \frac{100\%}{N} \sum_{t=1}^{N} \left[ \left( \frac{x_t - \hat{x}_t}{x_t} \right)^2 \right]$$

6. Weighted mean absolute percentage error (WMAPE)

$$WMAPE = \frac{\sum_{t=1}^{N} |x_t - \hat{x}_t|}{\sum_{t=1}^{N} x_t} 100\%$$

7. Symmetric mean absolute percentage error (SMAPE)

$$SMAPE = \frac{200\%}{N} \sum_{t=1}^{N} \left[ \left| \frac{x_t - \hat{x}_t}{x_t + \hat{x}_t} \right| \right]$$

### 3.6.3 Scaled error

A more generally applicable measure for forecast accuracy is the MASE. This measure scales the errors based on the in-sample MAE from the naïve forecast method. This naïve method sets the forecast to the value of the last observation. The MASE is independent of the scale of the data. The scaled error is less than one if it is a better forecast than the average one-step naïve forecast, and it is bigger than one if it is worse. Several authors argue that this is the best accuracy measurement (e.g. Hyndman & Koehler, 2006). This is the only measurement that can be used in every circumstance. This is in contrast with the scale dependent (e.g. MAE), which can only be used for single series and the scale independent (e.g. MAPE), which give skewed results for demand close to zero.

8. Mean absolute scaled error (MASE)

$$MASE = \frac{100}{N}\sum\nolimits_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{\sum_{i=1}^{N}\frac{x_i - x_{i-1}}{n-1}}\right|$$

## 3.7 Model selection

Now we described the basic principles of forecasting, there is one crucial step to discuss. We already mentioned this in the previous section, and it is about selecting the model that fits best to the data. The error measures are a good way of evaluating the different fitted forecasting methods. This is the case for Wavin, where they use the 'pick best' method of SAP APO, which selects the best fit according to the MAE (see Section 2.3).

However, there are also other methods to use for choosing the best method. We discuss this in this section. We also discuss whether one method should be used versus a combination of multiple forecasting methods. Before discussing the principles of model selection, we need to make an important remark.

As we explain in Section 3.1, it is important to not get drowned in very complex forecasting methods. Green & Armstrong (2015) argue that the forecasting method should be understandable for the forecaster. This is also underlined by Rasmussen (2004), who argues that just selecting the forecast method that best fits the time series may result in overfitting. These overfitted methods may include the random fluctuations that do not repeat themselves. This makes it clear that we need to be careful for selecting the best method. The goal should not be to get a forecast accuracy of 100%. We should aim for an acceptable forecast accuracy, since forecast is always error (Chambers et al., 1971).

### 3.7.1 Model fitting

Although the error measurements as we explain in Section 3.6 are a good way of measuring the fit of a forecasting method, it is subjective to overfitting. We will discuss two popular Information Criteria (IC), which penalizes the likelihood depending on the number of model parameters. In this way it compensates for the potential of overfitting. This gives a trade-off between complexity and goodness of fit. Both ICs are based on the general form of the information criteria which is formulated as follows:

$$IC = -2\log L_i + Kq$$

Where $L_i$ is the maximum likelihood for the candidate model *i*, *q* is the number of parameters used for the model and *K* the penalty which is depended on the type of IC chosen.

With Akaike's Information Criteria (AIC) (Akaike, 1974) the penalty *K* is set to two. We can, therefore, formulate the formula as follows:

$$AIC = -2\log L_i + 2q$$

For Bayesian Information Criteria (BIC) (Schwarz, 1978) the penalty is set to log (n), with *n* the number of data points (sample size). The formula for BIC is as follows:

$$BIC = -2\log L_i + q\log(n)$$

There has been much discussion about which information criteria should be used (Burnham & Anderson, 2004). However, Kuha (2004) recommends using both criteria together. He showed that both are suitable for identifying good models for the observed data, but failure is on the corner. This is also in line with Armstrong (2001), who advocates for using different methods for comparing forecasts.

### 3.7.2 Model combination

The usual approach for forecasting is to use a single forecasting method that fits the data best. This is also the case for Wavin. However, it is also possible to use several different methods on the same time series. The idea behind this, is that different methods catch different data patterns. It is therefore not always the case that one single forecast method can catch all the patterns in the data. Moreover, because of these aggregated forecasts, the results are not dependent on one single method and is therefore less sensitive on the specific choice of the method. This makes that combined forecasts are less risky.

The combination of forecast models started as early as 50 years ago. Bates & Granger (1969) showed that combining forecasts leads to increased forecast accuracy. There has been much research about combining forecasts from that time on. For example, Clemen (1989) argued that combining forecasts unanimously leads to increased forecast accuracy by simply averaging the forecasts. This is a confirmation what Makridakis et al. (1982) and Kang (1986) wrote, when they argued that taking the average outperforms making the effort of assigning weights based on forecast accuracy, to the methods. The Golden Rule checklist, designed by Armstrong et al. (2015) to help forecasters to make good forecasts, also includes that different forecasts using different methods should be combined. This can be seen in step 3.5 and step 5 of Figure 36 in Appendix E.

The question is to what number of forecast methods should be combined. Literature is not very consistent on this part. Winkler & Makridakis (1983) showed that the forecast accuracy increased as the number of methods increased. However, this increase leveled off the more methods were added. Since the exact quantity is difficult to determine, Silver et al. (2017) recommends to combine only two methods. Armstrong (2001), however, pleads for using five or more methods to get a higher forecast accuracy. He found that by combining multiple forecasts the ex ante error can be reduced by 12,5% on average, ranging from 3% up to 24%.

## 3.8 Implications of good forecasting

Now we discussed all relevant topics for the research by means of literature, it is good to put it in a broader context. Forecasting can be difficult and time-consuming (and therefore costly), especially when the data cannot be extrapolated easily. Important is to know what the outcomes and advantages are of accurate forecasting. Although we will not execute a cost-benefit analysis with this research since it is not in the scope, we touch lightly on the implications of better forecasts in this section.

Many companies use forecasts as input for comprehensive planning processes. Decisions about inventory and production planning almost always involve the allocation of resources based on the demand uncertainty. For future sales, raw materials need to be procured, where financial resources need to be made available for. Moreover, for production planning, capacity must be set in anticipation of the future sales. Besides, when these sales are uncertain, more safety stock is needed, to make sure the orders can be fulfilled when demand is more than expected.

Besides that forecasts are used for distribution of resources (Antle & Eppen, 1985; Stein, 1997), it is also used by managers to provide targets for organizational efforts (Hamel & Prahalad, 1989; Keating et al. 1999) and to integrate the operations management function with sales (Lapide, 2005), marketing (Crittendenet al., 1993; Griffin & Hauser, 1992) and product development (Griffin & Hauser, 1996) departments. Concluded can be that forecast sharing is a widespread practice for aligning capacity and managing supply (Cachon & Lariviere, 2001; Terwiesch et al., 2005).

The importance of forecasting can, therefore, not be overestimated. It touches many aspects of an organization. Forecast errors should, therefore, be avoided as much as possible. Besides the increase in inventory (Syntetos et al. 2009), it has also an impact outside the company. Reduced customer service (Kremer et al., 2011; Oliva, 2001; Oliva & Sterman, 2001), lower shareholders' return on investment (Copeland et al., 1994) and decreased collaboration with external stakeholders (Fildes et al., 2009) are just some of the negative implications of a low forecast accuracy. If not handled quickly, errors made in forecasting are often increased in the supply chain as a result of the so-called bullwhip effect (Lee et al., 1997).

As there has been done much research about the necessity of good forecasting and the implications of it, there is hardly any literature to find which provides quantitative results. The research of Syntetos et al. (2010) stands quite alone in this. They found a relation between forecast accuracy, calculated by MAPE, and inventory reduction, and cycle service levels and fill rates. They showed that a reduction of 1% of the MAPE can be translated into an inventory reduction of about 15-20%. Besides, it means an increase of about 1% for the cycle service levels and fill-rates. These numbers show that higher forecast accuracies result in great improvements. There needs to be mentioned, however, that the research is based on a relatively small dataset and the same outcomes may not result for every organization. But is does not alter the fact that forecast accuracy has an high effect thoughout a company.

## 3.9 Filling the gap

While there is much research to find about the forecasting process and which forecasting methods and error measurements to use, the research falls short on how to deal with the forecastability of the products. For example, the ABC/XYZ classification is not described extensively in literature and no clear guidelines are given for how to make this classification and how to design this process. The CoV thresholds given in literature are not consistent and are industry dependent. More importantly, we have our questions about using solely this measure for making the distinction between the XYZ classes. The coefficient of variation describes the spread of the sales, but tells not particularly how difficult it is to forecast. We illustrate this in Figure 14.

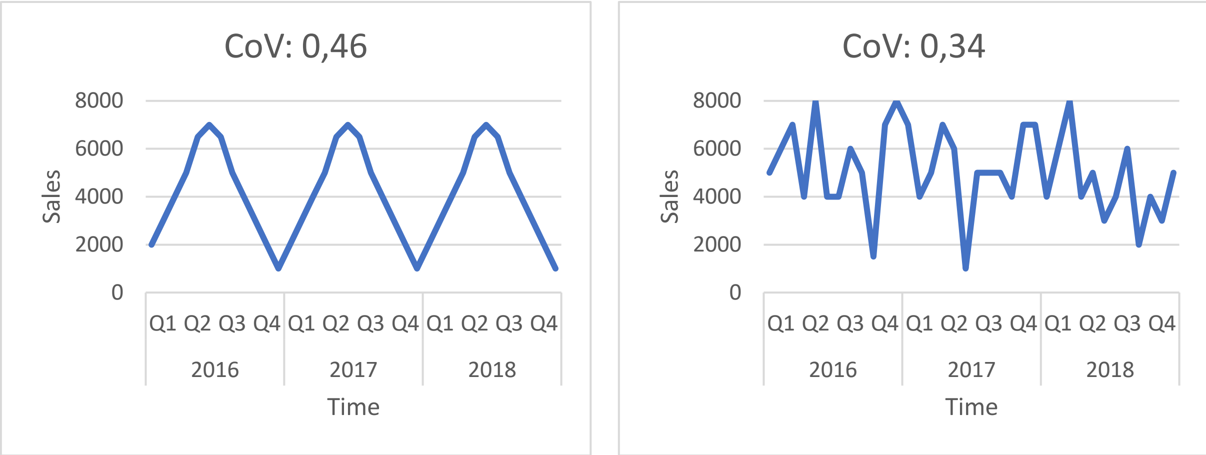



*Figure 14: Example that CoV fails to indicate the forecastability*

The left figure has a relatively high CoV value. However, there can be seen a clear seasonal pattern in the data and using a statistical method suitable for seasonality will make the forecast accurate. The right figure has a lower CoV, but the fluctuations are more random. This makes it more difficult to forecast. This raises the issue that it may not be desirable to make the XYZ classification on the CoV value alone. Distinguishing whether there is a clear trend or season may be a step to be preceded. Differentiating by means of the MAPE can be another possibility. High statistical MAPE, means more difficult to forecast, while lower MAPE indicate more accurate to forecast. This is more directly related to the forecastability of the sales. However, this has as drawback that this MAPE is an output value of the made forecast. The classification is an input, which influences the forecast accuracy. This research, therefore, tries to fill this gap in literature on how to make the right XYZ distinction.

Besides, literature does not explain extensively how to use this classification in practice. Literature describes that more difficult to forecast and important products needs to have more manual (judgmental) input. However, how the process exactly should be designed to incorporate the classification is missing. By using the case of Wavin we try to also fill this gap in literature.

Another interesting aspect of this case study touches the aspect that for the same industry and the same company, different situations in different countries are touched. In this way, the market characteristics need to be taken into account and the classifications need to be made country depending. This will put the research into perspective and emphasizes the dynamic use of the classification.

The fourth gap in literature that we try to fill is to give guidelines which statistical method is most suitable for which class. There is barely anything to find in literature where this is done. Only Errasti et al. (2010) give such recommendations using the ABC/XYZ classification by analyzing a washing machine business unit in Spain. Our analysis contributes to help to choose between the many different statistical methods for the different classes.

Moreover, we also examine whether method combination is indeed an improvement over using a single method. There is already some literature about this topic, but with this research we determine if this is also the case for Wavin. We also examine what the best number of methods to combine is, to get the highest forecast accuracy.

To summarize, the following gaps in literature we will try to fill with this research:

I. How to make the classification based on which parameters and which thresholds
II. Differences of the classification between different markets
III. What the use and implications of the classification are

IV. Guidelines for which statistical methods are most suitable for which class
V. Whether model combination is an improvement over using a single model

## 3.10 Conclusion

The first step of a forecasting process is making the baseline statistical forecasts. For making these statistical forecasts, there are many different methods available catching different demand patterns. Basically, there are four different characteristics that the methods can be fitted on, which are level, trend, season and cycle. Literature argues that using complex methods should be refrained from, because the methods should be understandable by the forecaster.

For choosing the method that has the best fit to the demand pattern, different error measurements can be used. These measurements are also useful for evaluating the forecasts. The MAPE and MASE are the recommended measurements. Besides, literature advises to use multiple measurements in order to catch all the characteristics of the errors. Moreover, AIC and BIC are widely used methods which avoids overfitting by penalizing the number of parameters used. Literature showed that using multiple methods on one time series increases the forecast accuracy.

After the statistical forecasts have been made, it needs to be adjusted by means of qualitative data. Key aspects include that this needs to be done structured, the adjustments need to be documented and in order to not lose focus it should be done on a higher hierarchical level by means of grouping. The judgmental input can be divided into internal (e.g. price changes and promotions) and external (e.g. government policy and competitor activity) factors.

Literature describes different ways and approaches for making the ABC classification. Class A products are the most important and C products the least important. Examples for making the distinction can be done by using revenues. Normally about 20% of the products account for about 80% of the measurement chosen, which are in class A. For forecasting, classifying based on forecastability (XYZ) is very useful in order to know whether to put the focus on statistics or on judgmental input. Besides, different targets can be set. Although literature agrees that the distinction should be based on the coefficient of variation(CoV), it is biased which values to take as thresholds.

Forecasts have a high impact throughout a company. Therefore, a high forecast accuracy is highly desirable. Some results of better forecasting are higher customer service and lower inventory. Besides, it is crucial for allocating the resources for the production as well as finance.

With this research, we will try to fill several gaps in literature. We will give better recommendations how to make the XYZ classification, since the CoV value does not stand one-on-one with the forecastability. Besides, we will design a forecasting process that can incorporate this classification smoothly, implementing it in different markets. We will also give guidelines which statistical method is most useful for which class and whether model combination increases the forecast accuracy.

# 4. Making the Classification

Classifying the products helps to get a better understanding where to put the focus. The quote of the actress Joy Page illustrates where the focus should be on:

*"Instead of focusing on that circumstances that you cannot change – focus strongly and powerfully on the circumstances that you can."*

Focusing on the things that you cannot change, does not help. Instead, focus on the things that you can change and have impact. This is also the case for forecasting. You can forecast some products with statistics. However, other products have a more random sales pattern, and can only be forecasted accurately using qualitative data. Therefore, it is important to know where to focus on for which products. The classification gives guidelines where the focus should be on. In Section 4.1 we explain what data we used for making the classification and why. Section 4.2 contains the discussion on which parameters and corresponding thresholds we used for making the classification. In Section 4.3 we discuss the comparison between two measurements for making the XYZ classification. In Section 4.4 we give the classifications on SKU level, where after we give the classification on aggregated level in Section 4.5. Section 4.6 is about the forecast accuracy targets per class. In Section 4.7, we explain the Excel tool we developed for Wavin for making the classification. Section 4.8, which is the final part of this chapter, contains the conclusion of the chapter.

## 4.1 Setting up the data

For making the classification we used data from 2018. We believe that twelve months of data is sufficient to get a good basis of the characteristics of the data. This is also argued by literature (e.g. Silver et al., 2017; Kepczynski et al., 2018). For this research we made a classification tool in Excel. When the monthly sales, together with the statistical and final forecasts and corresponding accuracies, and revenues are imported to the Excel file, the classification will be made with one push on a button. With the tool, the forecaster can make the classification on different hierarchies (level 4/7/8/9), planning group and SKU level. For this research we give the classification on level 7/8/9. This classifies the products according to the assortment/brand and sub-assortment with the division of pipes or fittings (see Section 2.2 for the explanation). Together with the Demand expert of Wavin, this has been understood as the most useful classification, grouping the products in just enough detail. Moreover, this grouping is also used as the basis for the planning groups and for adding the judgmental input. However, when desired the tool also supports other types of groupings. In Section 4.7, we explain the Excel tool in more detail.

## 4.2 Choosing the parameters

The classification has two dimensions; the importance classified with ABC and the forecastability classified with XYZ. In this section we discuss which parameters we have chosen according to which thresholds. We also give the number of products that are classified in which class as a result of the chosen parameters and thresholds.

### 4.2.1 ABC parameter

As we explain in Section 2.5, Wavin already uses the ABC classification. Currently they make the classification based on revenues and orderliness. Revenues implicate how important a SKU is economically wise. Orderliness is the frequency the SKU is being sold. It can be that an SKU is very cheap, resulting in low revenues, but is sold often which means high orderliness. This classification is not only used by the forecaster but also by the production planner.

The new classification will be on group level, which asks for a new type of classification. The classification will merely be used for making the forecasts. Therefore, because of simplicity it will only be based on the revenues. When products have low revenues, meaning a lower classification, the focus will be less on these products. However, some of these products may have high orderliness, meaning that it is still an important product. Then, simply a higher safety inventory can be used. These products are cheap as a result of relatively low revenues, meaning that it will be a lower investment to have a higher safety inventory for these products than for high revenues products. Moreover, since the larger products of Wavin are stored outside and the smaller products in a warehouse, having a large safety inventory will not result in high warehousing costs (think of building the warehouse, electricity etc.).

#### 4.2.1.1 Thresholds of ABC

For the thresholds of the ABC classification, we used the same method Wavin uses and which is also argued by literature (e.g. Kepczynski et al., 2018). The products accounting for up to 80% of the revenues are grouped in class A, the next 15% are classified as Class B and the products accounting for the final 5% of the revenues is grouped in class C. In Table 8, the number of groups on level 7/8/9, assigned per class and per country are given.

*Table 8: ABC classification*

| | | Country A | | Country B | | Country C | |
|---|---|---|---|---|---|---|---|
| Class | Threshold | Groups (lvl 7/8/9) | % of total | Groups (lvl 7/8/9) | % of total | Groups (lvl 7/8/9) | % of total |
| A | 80% | 27 | 10,8% | 25 | 18,0% | 15 | 10,8% |
| B | 95% | 35 | 13,9% | 22 | 15,8% | 23 | 16,5% |
| C | 100% | 189 | 75,3% | 92 | 66,2% | 101 | 72,7% |
| Total | | 251 | 100,0% | 139 | 100,0% | 139 | 100,0% |

Especially for Country A and Country C there is a small group accounting for a large part of the revenues; only about 10% of the groups accounts for 80% of the revenues. Country B is closer to the original 80/20 pareto rule, with 18% of the products accounting for 80% of the revenues. In Figure 15 the pareto diagram of Country B is shown.

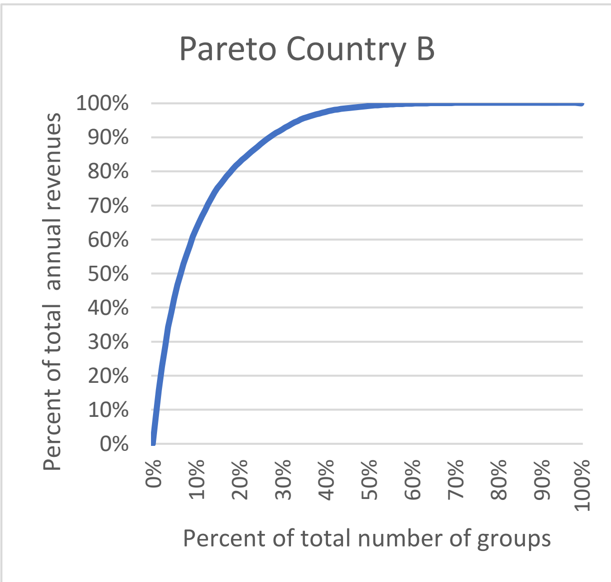



*Figure 15: Pareto diagram of Country B*

### 4.2.2 XYZ parameter

The parameter for the XYZ classification is less straightforward to determine. In literature mainly the coefficient of variation (CoV) of the demand is used. This measurement is defined as the standard deviation divided by the mean. It indicates how much the data fluctuates. However, literature is not consistent which thresholds to use and it is industry dependent. More importantly, we have our questions if this is indeed the right parameter. As we illustrate in Figure 14 in Section 3.9, using the CoV is not always a good measurement to indicate the forecastability. Seasonality increases the CoV value, while one still can forecast accurately with statistical methods including seasonality. For Wavin, the demand patterns often include some seasonality and most of the statistical methods used include seasonality. The CoV value is only useful when the data would be deseasonalized. This results that making the classification is much more complicated and time-consuming. Besides, more data is needed. Instead of the 12 months of data, at least 36 months of data should be analyzed in order to deseasonalize the data. Since the goal of the classification is to be easy to use and understandable by the user, deseasonalizing the demand and therefore the CoV is not a desirable measurement to determine the forecastability for Wavin.

Another option for determining the forecastability is using forecast errors, which are directly related to the non-forecastable parts of a time-series. The question consequently arises: Which error measurement to use? As we explain in Section 3.6, there are quite some measurements available. However, a single error measurement cannot tell the whole truth. On the other hand, using multiple measurements makes classifying more complicated. Therefore, the best would be to use one measurement that catches most of the error behavior and which fits Wavin's businesses. Literature (e.g. Hyndman & Koehler, 2006) argues that the MASE is the most generally applicable measurement for catching the forecast errors. The MASE is based on the comparison of the statistical forecast with the one-step naïve forecast. However, Wavin does not yet use this measurement, which requires an introduction of it. This could be of added value if not another property of this measurement was the case. This measurement does not take into account the seasonality, which may result in skewed results like when using the CoV. Only when the data would be deseasonalized or a seasonal one-step naïve forecast would be used, it will give good outcomes. Therefore, it would not be of any added value to use this measurement, because then we could just use the CoV.

Another widely used and recommended measurement is the MAPE, when comparing multiple time series. There are a couple of options of this MAPE. We choose for the weighted MAPE because this is the measurement Wavin already uses and it is easy to understand. Besides, using the wMAPE solves the problem of that intermittent demand gives skewed results and the wMAPE makes sure it only has a very minor impact on the total forecast accuracy. Moreover, the measurement is directly linked to the ability to forecast the products accurately, since this is the measurement Wavin uses to evaluate the forecasting performance.

However, as we explain in Section 3.6, the wMAPE has some issues. It penalizes positive errors more than negative errors. Let us explain this with an example shown in Table 9.

*Table 9: Illustration of forecast accuracies calculated by wMAPE*

| Actual Sales | Forecasted sales | Forecast accuracy |
|---|---|---|
| 100 | 80 | 80% |
| 80 | 100 | 75% |
| 100 | 250 | -50% (0%) |
| 250 | 100 | 40% |

When the actual sales are 100 and the forecast is 80, the forecast accuracy will be 80%. However, when the sales are 80 and the forecast is 100, the accuracy will be 75%. The difference (20) in sales is the same, but the accuracy varies. A positive error has a bigger impact than a negative error. Another issue is that when the forecasted sales are more than twice of the actual sales, the wMAPE would be higher than 100%. Since the accuracy is calculated by 1 minus the wMAPE, this would get a minus value. A negative value is not possible and this is then corrected to 0%.

Another disadvantage of the wMAPE is that it is an output value, which means that you assume that the statistical forecasts are done correctly. Products with lower statistical forecast accuracies will end up in the Y or Z class, which means less emphasize will be put on the statistics and more on the judgmental input. However, it could be the case that when more work would be spend on the statistical forecast, the product would be in class X. Besides, for products in class X the focus will be on statistics, but for these products the statistical forecasts are already good and may not always be increased when putting more effort on the statistical forecasts.

This is indeed a dilemma, which is not easy to solve. Both the CoV and wMAPE has advantages and disadvantages. However, we still think the wMAPE is the best measurement to use for Wavin. Currently, the statistical forecasts are made with the same attention, with some extra attention for A products. The results are the forecast accuracies. When a product will be classified in the X class, there can be understood that the statistical forecasts are already sufficient. A bit extra focus on these products can probably increase the forecast accuracies at least a little. However, adding the market intelligence to this class is not really needed to reach a good forecast accuracy. Besides, for lower forecast accuracy products, the statistical forecast just seemed not to give a proper forecast in history. In result to reach the target, more qualitative information is needed. The available statistical methods are simply not sufficient for a good forecast. Besides, improving the statistical forecasts is not that easy. With SAP APO, already the best fit of the statistical methods to the demand patterns is used. A way to improve the statistical forecasts is by spending more time on cleaning the history by correcting the outliers. When much outlier correction is needed every month, it means that there are often sales that cannot be predicted with the statistics for which will be the case for class Y and even more for class Z products. Consequently, it means that you can put as much work on outlier correction, in future these outliers will most likely happen again, resulting in a lower ability to forecast.

Kepczynski et al. (2018) also argued that instead of using the CoV, one can use the wMAPE for making the XYZ classification. They give as a remark that also the stability of the wMAPE should be taken into account. If the wMAPE fluctuates much from month to month, then the process is not very consistent and it should be assigned to a lower class. Therefore, for classifying the products we set the requirement that the forecast accuracy should be higher than the threshold for at least 9 out of 12 months. In this way, we expect that at least 75% of the time the statistics can guarantee a good forecast. We chose for 9 out of 12 months, since a higher number of months that the percentage would need to suffice would result in very little groups classified as X or Y for the thresholds we set in the next section. This would especially be the case for Country C, where there are 7 groups in class X when 9 months and only 3 groups in class X when 10 months is chosen . A lower number of months would imply that the forecast accuracies are not very stable. In Table 39, Table 40 and Table 41 in Appendix F, the analysis with the products per class X and Y for different parts of the months for the three countries are given.

#### *4.2.2.1 Thresholds of XYZ*

As we already stated a couple of times, Wavin defines the forecast accuracy by 1 minus the wMAPE (formula in Section 2.6). Using this measurement, we can set the thresholds for the different groups. Preferably the same thresholds can be used for all countries. However, setting thresholds where the majority is in class X or where there are only a couple of products in class X is not desirable. And therefore, unlike the ABC classification where the thresholds are the same for all the countries, we adapt the XYZ thresholds according to the specific country. As we illustrate in Section 2.6, the forecast accuracies can variate much depending on the country. This is due to the market characteristics. For example, in Country A, the sales are relatively stable, in contrast with Country B where there is much more fluctuating project-based demand.

We experimented with some thresholds of which the results are shown in Table 10, Table 11 and Table 12. For Country B and Country C we experimented with thresholds of 60%, 65% and 70% for class X. This is the class where the products in general are the easiest to forecast. For Country A it made more sense to test the thresholds of 75%, 80% and 85% for class X. For class Y we tried thresholds of 10%, 15% and 20% lower of class X. All the remaining products are classified as class C. In the tables, the chosen thresholds are highlighted with green and in the last column the percentages of the total are given.

*Table 10: XYZ classification Country A*

| TH: X<br>TH: Y | 75%<br>55% | 75%<br>60% | 75%<br>65% | 80%<br>60% | 80%<br>65% | 80%<br>70% | 85%<br>65% | 85%<br>70% | 85%<br>75% | % of<br>total |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 43 | 43 | 43 | 27 | 27 | 27 | 11 | 11 | 11 | 10,8% |
| Y | 32 | 26 | 16 | 42 | 32 | 25 | 48 | 41 | 32 | 12,7% |
| Z | 168 | 174 | 184 | 174 | 184 | 191 | 184 | 191 | 200 | 73,3% |
| D | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 3,2% |
| Total | 251 | 251 | 251 | 251 | 251 | 251 | 251 | 251 | 251 | 100,0% |

*Table 11: XYZ Thresholds Country B*

| TH: X<br>TH: Y | 60%<br>40% | 60%<br>45% | 60%<br>50% | 65%<br>45% | 65%<br>50% | 65%<br>55% | 70%<br>50% | 70%<br>55% | 70%<br>60% | % of<br>total |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 13 | 13 | 13 | 7 | 7 | 7 | 2 | 2 | 2 | 5,0% |
| Y | 7 | 3 | 3 | 9 | 9 | 6 | 14 | 11 | 11 | 6,5% |
| Z | 107 | 111 | 111 | 111 | 111 | 114 | 111 | 114 | 114 | 79,9% |
| D | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 8,6% |
| Total | 139 | 139 | 139 | 139 | 139 | 139 | 139 | 139 | 139 | 100,0% |

*Table 12: XYZ thresholds Country C*

| TH: X<br>TH: Y | 60%<br>40% | 60%<br>45% | 60%<br>50% | 65%<br>45% | 65%<br>50% | 65%<br>55% | 70%<br>50% | 70%<br>55% | 70%<br>60% | % of<br>total |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 12 | 12 | 12 | 7 | 7 | 7 | 5 | 5 | 5 | 5,0% |
| Y | 25 | 15 | 13 | 20 | 16 | 15 | 18 | 17 | 8 | 11,5% |
| Z | 90 | 100 | 102 | 100 | 104 | 105 | 104 | 105 | 114 | 74,8% |
| D | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 8,6% |
| Total | 139 | 139 | 139 | 139 | 139 | 139 | 139 | 139 | 139 | 100,0% |

We agreed with the Demand expert of Wavin to set the thresholds for class X for Country B and Country C to 65% and 80% for Country A. In this case at least 5% of the products are in class X. Any higher thresholds for Country B and Country C would not make sense, since (barely) no products would be in class X. However, for the UK the threshold of 65% is a bit too low. Ideally the forecast accuracies are as high as possible, and when the threshold of 65% is chosen, not much attention to increase this percentage will be given. For Country A however, the percentages are in general higher. It is therefore desirable to have for Country A a threshold of 80%.

For the Y class we set the thresholds for Country B and Country C to 50%. In this way at least 5% of all products are classified in this class. For Country C, a threshold of 55% would also be possible, but we agreed that it is better to have one threshold for both countries. For Country A we set the threshold for class Y to 65%, since then class Y has just a bit more products in the class than class X.

As you may have noticed, many products are classified as difficult to forecast, having low forecast accuracies, which are in class Z. This may seem that the classification is not made equally distributed at all. However, many of these products are slow movers, having low revenues. This means that many products are in class CZ as we can see in Section 4.4. This is often intermittent demand, having often demand of (close to) zero. These products are generally difficult to forecast. Moreover, a low forecast accuracy for this class has not a major impact, because for these products a higher safety inventory can be held, which is not a big investment due to the low revenues.

In the tables we also include a fourth class, which is class D. These are the products that had revenues in 2018, but no sales. This can happen since it is among others intercompany demand, equipment or raw materials. These products are not used for selling and do not need to have a forecast. Therefore, they are classified in group D and no attention for forecasting these groups are needed.

## 4.3 wMAPE compared with CoV

Although we make the classification based on the wMAPE, we still find it interesting to compare this measurement when the coefficient of variation would have been used. Especially since this is a measurement used in the majority of literature. For this comparison we deseasonalized the demand, to determine if the CoV of the deseasonalized demand would give an accurate estimation of the forecast accuracy. There are different ways to deseasonalize the demand. We used the rather simple approach of the moving average for deseasonalizing the thousands of SKUs. We used 36 months of data (2016, 2017, 2018) with 12 periods, since we focus on the monthly seasonality. For the CoV values of the product groups on level 7/8/9, we weighted them according to sales. We used this approach, since the forecast accuracy is also calculated using the weights. We compared three different CoV values calculated over the (de)seasonalized demand of 2018. The first is the CoV value when the data is not deseasonalized. We expect that this gives the least accurate classification, since seasonality can increase the CoV value, although using seasonal statistical methods it can still give an accurate forecast. The next CoV value is when all demand is deseasonalized, without considering there is a clear seasonality visible. This simplified approach is used since when no seasonality is occurring the seasonal indexes would also have no major impact on deseasonalizing the data, resulting in a similar CoV value when the data would not be deseasonalized. A drawback of this approach is that it may happen that the CoV value could increase. This is not desirable, because the point of deseasonalizing the data is to lower the impact of seasonality on the CoV value. Therefore, we used the following third approach. We took the minimum CoV value per SKU of the seasonal and deseasonal demand. In this way only a reduction of the CoV value would be considered. Our expectation was that the third approach would give the best results.

Using the XYZ classification based on the forecast accuracy thresholds given in Section 4.2, the optimal CoV thresholds (rounded to two decimals) were determined such that both methods would give the most same classifications as possible. In Table 13, Table 14 and Table 15 the outcomes are shown for the three countries.

*Table 13: CoV values Country A*

| CoV | X | Y | Different class | % of total |
|---|---|---|---|---|
| Seasonal | 0,17 | 0,29 | 32 | 12,7% |
| Deseasonalized | 0,15 | 0,26 | 25 | 10,0% |
| Minimum | 0,16 | 0,23 | 24 | 9,6% |

*Table 14: CoV values Country B*

| CoV | X | Y | Different class | % of total |
|---|---|---|---|---|
| Seasonal | 0,35 | 0,50 | 3 | 2% |
| Deseasonalized | 0,40 | 0,48 | 4 | 3% |
| Minimum | 0,31 | 0,48 | 3 | 2% |

*Table 15: CoV values Country C*

| CoV | X | Y | Different class | % of total |
|---|---|---|---|---|
| Seasonal | 0,31 | 0,49 | 10 | 7,2% |
| Deseasonalized | 0,32 | 0,47 | 8 | 5,8% |
| Minimum | 0,23 | 0,43 | 10 | 7,2% |

For Country B the relation between the CoV and wMAPE is the strongest. With the optimal CoV thresholds, only about 2% of the products is placed different than when using the wMAPE as measurement for the seasonal and minimum CoV. For Country A the percentages are around 10%. This would suggest that the CoV does not always give a good classification for the current statistical forecast accuracy. However, the outcomes suggest that deseasonalizing the demand is an improvement. Country C is between Country B and Country A with percentages of around 7% where the classifications do not correspond with each other. Remarkable is that for Country C the best classification is given when all demand is deseasonalized, while for Country B this gives the worst outcomes.

When we take a closer look to the different classified products for when the minimum CoV is used, it is often the case that the forecast accuracy is close to the threshold, the CoV is close to the threshold or both is the case. This means that when the threshold would be a bit different or when the sales would be a bit different the classification could have been done correct. Giving the fact that the future may give some differing results, it may happen that in the future the classification would have been the same for the CoV and the forecast accuracy. This suggest that the relation between CoV and the wMAPE is relatively strong. In Table 42, Table 43 and Table 44 in Appendix G, the issues where the classes differ when using CoV or forecast accuracy are given. In these tables the difference between the CoV value of the threshold of the other class together with the forecast accuracy difference of the thresholds and the number of months that would needed to change to end up in another class are given. When the CoV would be less than 0,05 different to the threshold we find it close. For the forecast accuracy we assumed it was close to the threshold when 1 or maximum 2 months needed to change in forecast accuracy to be classified to the class of the CoV value.

For Country B, all different placed groups are close to a threshold (CoV and/or accuracy), where for Country C this is only once not the case (marked with orange in Table 44). For product group 36, the CoV value is high, while the accuracy is also relatively high. This means that the forecast is done remarkably well given the (deseasonalized) spread. For Country A there are more issues where the CoV does not correspond with the forecast accuracy. Four groups are placed different where the CoV and forecast accuracy are not close to a threshold (marked with orange in Table 42). All these four groups have a worse forecast accuracy than would be expected using the CoV. Moreover, one of these four groups, it is classified as class X when using CoV, but where according to the forecast accuracy it would be classified with class Z (marked with red in Table 42). For this group, the forecasts seemed to be done bad, given that the (deseasonalized) demand was relatively stable.

Form this analysis we can conclude that especially for Country B and Country C, but also in a bit lower extent for Country A, the forecast accuracy based on the wMAPE is a good alternative for a deseasonalized CoV. Most of the times both measurements would give the same class or are close to getting the same class. The advantage of the wMAPE is that it is already the measurement used by Wavin to determine the forecast accuracy on which the business is evaluated. Therefore, the thresholds can be determined more easily, with better understanding for the user. Besides, using the forecast accuracy as threshold, one can determine more specifically what the targets should be. Moreover, some other aspects that the CoV does not include but the forecast methods do, can have influence on the forecastability. As an example, take intermittent demand, which increases the CoV.

However, sometimes sales can occur very regularly (e.g. every 5th month), making it more easy to forecast using the Croston method (see Section 3.4.6).

## 4.4 The classifications

Now the thresholds are known, we can give the ABC-XYZ classifications of the three countries. For each country we made four diagrams for the classification. The first diagram is the *number of product groups*, which illustrates how many product groups are in each class. In this way, one can easily see how many groups to deal within each class. The second diagram gives the *statistical forecast accuracy* per class. This accuracy is weighted per class. The third is the *final forecast accuracy* and is the accuracy when it is adjusted during the forecast meeting. Ideally, this is an increase of the statistical forecast accuracy. The last diagram is the *difference* between the statistical and final forecast accuracy. In this way, one can see with one glance for which classes the human adjustments are indeed an improvement. Recall that the forecast accuracy of a certain group is calculated by weighting the SKUs based on the sales of the group.

For the diagrams we use conditional formatting of Excel to highlight the high and low values. Except for the diagrams about the *number of product groups*, the highest values are highlighted green, where it will be turned respectively towards yellow, orange and red when the values decrease. An example of the conditional formatting of Excel is shown in Figure 16. We chose for this formatting since high values mean forecasting is done well (green). Low values are formatted red, which intuitively draws the attention that there are improvements possible.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 90 | 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 | 0 |

*Figure 16: Example of conditional formatting of Excel*

For the number of product groups we chose for formatting from blue towards red. A high value does not inherently say that it is a good or bad performance. However, when there are only a few product groups in one class, the forecaster should be careful making too generalized assumptions.

For analyzing the classifications, we use a hypothesis, which we define as follows:

***<u>Hypothesis 1:</u>*** *Products with a lower statistical forecastability e.g. class Y and Z have a higher improvement of the final forecast compared to the statistical forecast.*

If the qualitative adjustments are done correctly, this would have the most impact on more difficult to forecast products. If the statistical forecast is already 80% for example, improving the forecast accuracy by means of qualitative information would be more difficult than when the statistical forecast would be only 30%.

### 4.4.1 Classification of Country A

Figure 17 shows the classification of Country A. As we expected, there are many products in group CZ. These are the least important and more difficult to forecast products. For Class A, most products are in class X, where this is more equally divided for Class B. Even though we set the statistical threshold for class X only at 80%, the weighted averages for these classes are around 85%.

Especially for class B and C, we can conclude that the hypothesis is true. For both classes the final accuracy increases when the forecastability decreases. However, for class AZ, where you would expect an increase, there is a major decrease when the forecast is adjusted with qualitative input. While the statistical forecast accuracy was almost 50%, it decreased to only 27% after adding the qualitative information. This class impacted the total final forecast accuracy that much that the total final forecast accuracy was lower than the statistical forecast accuracy (75% compared to 76%), although for all other classes there was an increase.

**Number of product groups**

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 4 | 13 | 167 | 251 |
| Y | 9 | 13 | 10 | |
| X | 14 | 8 | 5 | |

Importance (High → Low)

**Statistical forecast accuracy**

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 49% | 15% | 9% | 76% |
| Y | 81% | 78% | 76% | |
| X | 87% | 87% | 84% | |

Importance (High → Low)

**Final forecast accuracy**

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 27% | 30% | 26% | 75% |
| Y | 82% | 83% | 81% | |
| X | 87% | 86% | 87% | |

Importance (High → Low)

**Difference**

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | -22% | 15% | 18% | -1,0% |
| Y | 1% | 5% | 5% | |
| X | 1% | 0% | 3% | |

Importance (High → Low)

*Figure 17: Classification Country A*

### 4.4.2 Classification of Country B

With the thresholds of 65% for class X and 50% for class Y, there are no products in class BY as can be seen in Figure 18, showing the classification of Country B. Especially class AX has a relatively high statistical forecast accuracy of 75%. However, class AZ, which consists of 15 products has only a statistical forecast accuracy of 20%, with an increase of 1% for the final accuracy. Also class BZ and class CZ have very low statistical and final forecast accuracies. While for class BZ adding the qualitative information improves the accuracy, for class CZ it means a decrease of 10%. These are the classes that are adjusted most, since for the other classes the difference is only about 1%. The hypothesis is only true for class B, but this is also the class without products in the Y classification.

**Number of product groups**

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 15 | 20 | 76 | 139 |
| Y | 6 | 0 | 3 | |
| X | 4 | 2 | 1 | |

Importance (High → Low)

**Statistical forecast accuracy**

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 20% | 0% | 13% | 50% |
| Y | 68% | N/A | 57% | |
| X | 75% | 70% | 70% | |

Importance (High → Low)

**Final forecast accuracy**

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 21% | 7% | 3% | 50% |
| Y | 67% | N/A | 56% | |
| X | 74% | 71% | 71% | |

Importance (High → Low)

**Difference**

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 1% | 7% | -10% | -0,3% |
| Y | -1% | N/A | -1% | |
| X | 0% | 0% | 1% | |

Importance (High → Low)

*Figure 18: Classification Country B*

### 4.4.3 Classification of Country C

Figure 19 shows that there are no products with very low revenues that are stable for Country C (class CX). As we illustrate in Section 2.6, Country C is the only country where the qualitative input showed to have an overall improvement of the statistical forecast accuracy. This is mainly due to the products in class AZ and BZ, where the final forecast means an increase of respectively 12% and 15%. However, also for Country C there are some classes where adding qualitative information has some damaging effect. Especially for class BX and CY, it means a significant decrease. As we explain in Chapter 5, these are the classes where (almost) only should be focused on statistics. With the new classification, these decreases can therefore be avoided in future. The hypothesis is only correct for class B and C, since for these classes the final forecast accuracy increases when the forecastability decreases.

Number of product groups (Forecastability: Low (Z) to High (X); Importance: High (A) to Low (C))

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z | 4 | 16 | 84 | 139 |
| Y | 6 | 5 | 5 | |
| X | 5 | 2 | 0 | |

Statistical forecast accuracy

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z | 48% | 42% | 14% | 66% |
| Y | 65% | 66% | 65% | |
| X | 77% | 75% | N/A | |

Final forecast accuracy

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z | 59% | 57% | 21% | 69% |
| Y | 63% | 66% | 59% | |
| X | 78% | 68% | N/A | |

Difference

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z | 12% | 15% | 7% | 2,9% |
| Y | -2% | 0% | -6% | |
| X | 0% | -7% | N/A | |

*Figure 19: Classification Country C*

### 4.4.4 Conclusion

The classifications give varying results for the countries. It seems only to be (almost) consistent that for class Z there is a major increase when the statistical forecast is adjusted by Sales & Marketing (except for class C of Country B and class A for Country A). There are even major decreases for some classes of the three countries. This classification made the problem clear, that adding market intelligence is not always done correctly and we expect that using this classification can help to give the right focus. For easy to forecast products, not much qualitative data should be added, since the majority of the time this only has a negative effect. However, for the more difficult to forecast classes, more qualitative input is needed, and here should then also be the focus on. We elaborate on this in Chapter 5.

## 4.5 Aggregated classification

As the previous section made clear, adding the qualitative information is not always done correctly, which results for some classes that there is a decrease in the forecast accuracy. This may be a result of that the qualitative adding is done on aggregated level. During the forecast meetings, the performance of the final accuracy is generally evaluated on this aggregated level. Therefore, it is valuable to evaluate the final accuracy of the classes on aggregated level to see whether the adjustments are indeed an improvement. In the next three sections we give similar figures as in Section 4.4, but now the forecast accuracies are calculated on aggregated level (level 7/8/9). The accuracies are consequently higher in general as a result of the risk pooling effect. The products are still in the same class as we classify them in Section 4.4. For analyzing the figures, we define another hypothesis:

***<u>Hypothesis 2:</u>*** *There are better improvements of the final forecast accuracy compared to the statistics on aggregated level than on SKU level.*

We expect this hypothesis to be true, since the qualitative adding is done on aggregated level.

### 4.5.1 Aggregated classification of Country A

As we expected, the aggregated forecast accuracies shown in Figure 20 are higher than it was the case when it was calculated on SKU level (Figure 17). This is especially the case for class Z products. For the final accuracy for AZ it is 23% higher, for BZ it is 38% and for CZ it is 17% higher. However, the differences between the statistical and final forecast on SKU and aggregated level is somewhat similar. Even on aggregated level, class AZ had a decrease of 18%, when qualitative information is added. However, this is better than the 22% decrease on SKU level. Class CZ is the other class where the improvement of the final forecast is higher on aggregated level than on SKU level. However, these are the only two classes (AZ and CZ) where this is the case. Moreover, even on aggregated level the overall statistical forecast accuracy is higher than the final forecast accuracy. This is a disturbing fact, since with all the extra effort is taken for adjusting the forecasts, it only resulted in a decrease of the accuracy.

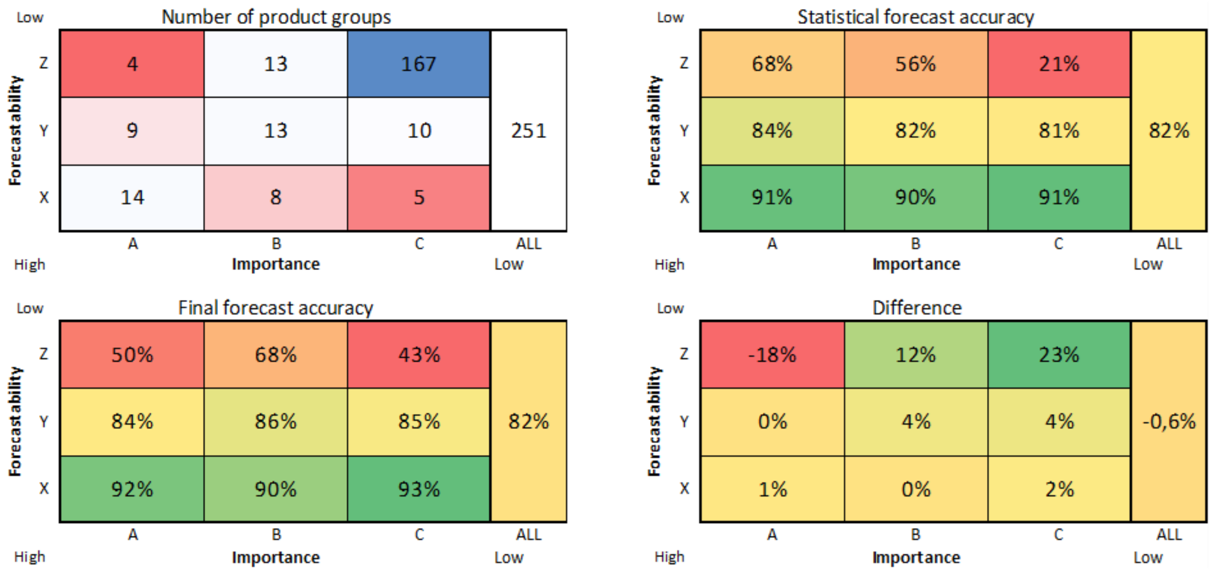


*Figure 20: Classification with aggregated accuracy Country A*

### 4.5.2 Aggregated classification of Country B

What strikes directly by comparing the forecast accuracies on SKU level (Figure 18), with the forecast accuracies on aggregated level (Figure 21), is the 26% increase on aggregated level, compared to the 1% on SKU level for class AZ. This makes clear that qualitative adding for this group is done well, since this is done on aggregated level. On aggregated level the overall increase is 6,6%, while on SKU level this was a decrease of 0,3%. This implies that adding qualitative adding for Country B is not harmful on an overall level. However, there are still some classes (AY and CZ), where adding qualitative information is decreasing the forecast accuracy. Moreover, class BZ had higher improvements on SKU level than on aggregated level, but the accuracies are much higher on aggregated level.

Number of product groups

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 15 | 20 | 76 | |
| Y | 6 | 0 | 3 | 139 |
| X | 4 | 2 | 1 | |

Importance (High → Low)

Statistical forecast accuracy

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 40% | 38% | 31% | |
| Y | 80% | N/A | 80% | 65% |
| X | 77% | 83% | 86% | |

Importance (High → Low)

Final forecast accuracy

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 66% | 41% | 20% | |
| Y | 79% | N/A | 78% | 72% |
| X | 77% | 84% | 89% | |

Importance (High → Low)

Difference

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 26% | 3% | -10% | |
| Y | -1% | N/A | -1% | 6,6% |
| X | 0% | 1% | 3% | |

Importance (High → Low)

*Figure 21: Classification with aggregated accuracy Country B*

### 4.5.3 Aggregated classification of Country C

Figure 22 shows the values for Country C on aggregated level. Adding qualitative information to the statistical forecast seems to have a higher impact on aggregated level than on SKU level for class AZ, BZ, CZ and CY. However, especially for class BX and also for class BY it has an even more negative impact on the accuracy. This is of course not desirable. Using different approaches per class, where class BX for example should (almost) only be focused on the statistics this can be avoided, increasing the accuracy.

Number of product groups

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 4 | 16 | 84 | |
| Y | 6 | 5 | 5 | 139 |
| X | 5 | 2 | 0 | |

Importance (High → Low)

Statistical forecast accuracy

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 61% | 51% | 36% | |
| Y | 79% | 80% | 78% | 78% |
| X | 88% | 86% | N/A | |

Importance (High → Low)

Final forecast accuracy

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 76% | 68% | 42% | |
| Y | 77% | 78% | 74% | 81% |
| X | 89% | 76% | N/A | |

Importance (High)

Difference

| Forecastability (Low → High) | A | B | C | ALL |
|---|---|---|---|---|
| Z | 14% | 17% | 6% | |
| Y | -2% | -2% | -5% | 3,4% |
| X | 0% | -10% | N/A | |

Importance (High → Low)

*Figure 22: Classification with aggregated accuracy Country C*

### 4.5.4 Conclusion

The stated hypothesis that the forecast accuracy increases (final compared to statistical) are higher on aggregated level than on SKU level is only the case for class Z. For all the other classes the increases are somewhat the same or even lower. We can conclude that even on aggregated level adding the qualitative information is not always done correctly, which applies for all the three countries. This confirms the essentiality that different approaches are necessary to get a better focus and to be able to become more consistent in increasing (instead of decreasing) the forecast accuracy with the input from Sales & Marketing.

## 4.6 Setting forecast accuracy targets

Differentiated target setting is a valuable outcome of the classification. The evaluation of the performance can be done more precisely. Easier to forecast and more important products should have a higher forecast accuracy target than products which are more difficult to forecast and less important. Therefore, we give for all classes and for all countries a different target. We set the targets in such a way that the more important products have a higher target and also easier to forecast products a higher target. We took into account both the thresholds as well as the performance (statistics and final) of 2018 for setting the targets. We discuss first the forecast accuracy targets for Country A after which we give the targets for Country B and finally for Country C.

### 4.6.1 Targets Country A

For Country A we set the threshold for class X to 80%. However, Figure 17 makes clear that the weighted average of the classes is much higher than this threshold. This implies that the forecast accuracies of some products are significantly higher. For these groups the focus should be on the statistics and spend as little time as possible on adding the market intelligence, especially for class B and even less for class C. Therefore, we want to have the highest target for class AX. Based on the data of 2018, class AX reached a statistical and final forecast accuracy of 87%. To create some space for improvements, we set the target for this group to 90%. Although the final forecast accuracies of class BX and CX are similar to class AX we set the target for these classes to 85%. The statistical forecast accuracies are close to this 85% and with the classification we almost only want to focus on the statistics for these classes. The focus of qualitative information adding should be shifted to other more important and more difficult to forecast classes.

Class Y, where the products have a statistical forecast accuracy of around 65%-80% for at least 9 months, may have a lower target. Currently, for class AY the average statistical and final forecast accuracy was around 82% in 2018. We expect that in particular the qualitative information adding can be done better as a result of using the classification. Therefore, we set the target to 85% for class AY. Class BY has a statistical and final forecast accuracy of respectively 78% and 83%. Because with this class the focus should be on the statistics, with some qualitative input we set the target to 85%. Class CY is less important and we set the target to 80%.

The products in class Z have the lowest targets as a result that they are more difficult to forecast. With adding the qualitative information the forecast accuracy can be increased, but in general not to the same extent as more predictable products would have. For class CZ, we set the target to 45%. These products are difficult to predict and have low importance. Putting much attention to get a high forecast accuracy for these products will be time-consuming. Besides, as we mention in Section 4.2, a low forecast accuracy for these products is not a bad case. Holding a higher safety stock, which is a consequence of the low forecast accuracy will not be a high investment since in this class only the products with relatively low revenues are placed. It is more important to focus on the AZ products, which we set the target to 60%. This may seem to be a high target, given the fact that currently the final forecast accuracy is only 27%. However, the statistical forecast accuracy was almost 50% for this class and since with the increased attention we expect that this can and also should be increased. We set the target for class BZ to 50%. All the targets together with the necessary improvements are shown in Figure 23. The overall target for all products together is 80%.

Target final forecast accuracy

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z (Low) | 60% | 50% | 45% | 80% |
| Y | 85% | 85% | 80% | |
| X (High) | 90% | 85% | 85% | |

Importance: High (A) – Low (ALL)

Final accuracy - Target

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z (Low) | -33% | -20% | -19% | -5% |
| Y | -3% | -2% | 1% | |
| X (High) | -3% | 1% | 2% | |

Importance: High (A) – Low (ALL)

*Figure 23: Final forecast accuracy targets Country A*

Some improvements may be made by the focus on statistics, but the most will be by correctly adding the market intelligence. As Figure 27 in Chapter 5 makes clear, the most effort and improvements should be made for important and difficult to forecast products. The targets of class BX, CX and CY are about the same as the current statistical forecast accuracy. For these products no real improvements are needed. The focus should be especially on class AZ, where the highest improvement is needed and also possible.

### 4.6.2 Targets Country B

With similar reasoning the targets are set for Country B. The threshold for class X is for Country B set to 65% for at least 9 months. Based on the data from 2018, we expect that it should be possible to get at least 80% of forecast accuracy for class AX and 70% for class BX and CY. The target for AX is the highest improvement of the current situation. We expect that because of the focus on statistics and some manual adjustments the forecast accuracy can be increased for this group. For the two other groups, the targets may be similar to the current statistical accuracies, since with the classification no high improvements will be expected or necessary.

For class Y, adding the qualitative information resulted in a lower forecast accuracy in 2018. This needs to be avoided and with more attention, it should result in an increase. Especially for class AY, this should be done better. We, therefore, set the target for class AY to 75% and for class CY to 60%.

The forecast accuracies for class Z were very low for Country B in 2018. This is the class where statistics normally fall short. With more focus and more carefulness adding the qualitative information should result in higher increases of the statistical forecast accuracies. For class AX, we set the target to 50%, which means that an improvement of 29% is needed. Class BZ and CZ had only a final forecast accuracy of respectively 7% and 5%. This would suggest that forecasting is not possible or helpful for these products. However, with some more attention it can be increased and we set the targets to respectively 45% and 35%. The different targets for the different classes are shown in Figure 24. The overall target is 60%, which is an increase of 10% of the forecast accuracy of 2018.

Target final forecast accuracy

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z (Low) | 50% | 45% | 35% | 60% |
| Y | 75% | N/A | 60% | |
| X (High) | 80% | 70% | 70% | |

Importance: A = High, ALL = Low

Target final forecast accuracy

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z (Low) | -29% | -38% | -32% | -10% |
| Y | -8% | N/A | -4% | |
| X (High) | -6% | 1% | 1% | |

Importance: A = High, ALL = Low

*Figure 24: Final forecast accuracy targets Country B*

### 4.6.3 Targets Country C

Country C has slightly higher targets than Country B, although the thresholds for the classification are the same. Class AX has a target of 80%, which means a small increase of 2018 is necessary. Class BX has a target of 75%, which is the same as the statistical forecast accuracy of 2018. This class should also (almost) only be made using statistics.

Class Y products have a bit lower target. For class AY we set the target to 70%, which leaves some space for improvements for especially the qualitative adjustments. Class BY can be improved a bit by means of better qualitative adding and gets a target of also 70%. Class CY can be dependent on the statistics alone and, we therefore set the target to 65%, which is the statistical forecast accuracy of this class for 2018.

Country C has the highest forecast accuracies for class Z compared to Country B and Country A. This means that there are not many sales that are really hard to forecast. Only class CZ has a very low forecast accuracy, and we set therefore the target for this class to 40%. The other classes had better results in 2018 and we set therefore for both of these classes the forecast accuracy to 65%. The different targets together with the necessary improvements of the current situation are shown in Figure 25.

Target final forecast accuracy

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z (Low) | 65% | 65% | 40% | 75% |
| Y | 70% | 70% | 65% | |
| X (High) | 80% | 75% | N/A | |

Importance: A = High, ALL = Low

Final forecast accuracy - Target

| Forecastability | A | B | C | ALL |
|---|---|---|---|---|
| Z (Low) | -6% | -8% | -19% | -6% |
| Y | -7% | -4% | -6% | |
| X (High) | -2% | -7% | N/A | |

Importance: A = High, ALL = Low

*Figure 25: Final forecast accuracy targets Country C*

## 4.7 Explanation of Excel tool for making the classification

The classifications of this chapter are the outcomes of the Excel tool we made for this research. Although the analysis of this research is based on level 7/8/9, it is possible to make the classification on different types of hierarchies. Moreover, as this chapter made clear, different countries require different thresholds for the classification. This research is only focused on three countries, although Wavin is active in more than twenty countries in Europe. Even other thresholds than in this research is stated may be necessary. Therefore the Excel tool makes it possible to alter the thresholds and make the classification based on this. After importing the data (sales, forecasts and revenues) and choosing the hierarchies and setting the thresholds the classification can be made easily by one push of a button.

This is done by using macros created in VBA. A screenshot of the dashboard of the tool with the example of Country B is shown in Figure 26.

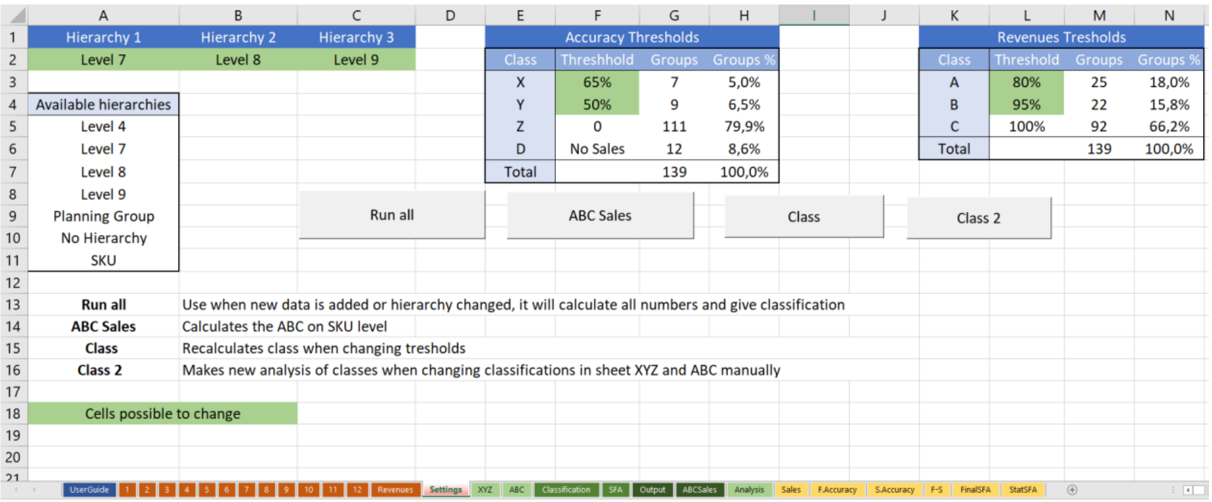



*Figure 26: Screenshot of the dashboard of the Excel tool to make the classifications*

The outcome of the tool are the diagrams (Figures 16-25) shown in this chapter. The classification is made using the statistical forecast accuracy thresholds on SKU level, but gives also the analysis on aggregated level. We refer to Appendix H, for a more detailed explanation of the Excel tool, where the user guide of the Excel tool is given.

Besides the tool for making the classification and the corresponding targets, we also made a tool for evaluating the monthly forecasts. This tool gives the values of the last month, per class and compared to the set targets. This tool can be used during the forecast meetings, where the Demand Manager discusses with the sales representatives to adjust the statistical forecast with the qualitative input. The tool includes a drop-down box where the sales representative can be selected and upon next, only the results of the products corresponding the sales representative are shown. This tool also includes a diagram with bars for the forecast accuracies of this year, lines for the forecast accuracies of the previous year and a green filled area with the target. In this way the performance can be easily compared with the values of the previous months and previous year. In Figure 48, Figure 49 and Figure 50 in Appendix I, screenshots of this tool are shown.

## 4.8 Conclusion

For making the ABC classification we use revenues as the parameter with the highest 80% of the revenues classified to class A, the next 15% to class B and the final 5% to class C. For making the XYZ classification we use the statistical forecast accuracy determined by the wMAPE as the parameter. This in contrast what often in literature is argued where they use CoV. However, this CoV does not take into account seasonality. Moreover, the wMAPE is more directly related to the forecastability, especially because of the fact that Wavin already uses this measurement for evaluating their forecasts. The thresholds for Country A are 80% for class X and 65% for class Y. The remaining products are classified in class Z. For Country B and Country C the threshold for class X is set to 65%, because otherwise (almost) no products will be in class X due to the different market characteristics. The threshold for class Y is set to 50% for these two countries.

We also tested what would happen when the CoV would be used when the demand would be deseasonalized. We saw that when the CoV thresholds are calculated to optimal, especially for Country B and also for Country C, most of the times the same classification would be assigned. We therefore can say that there is a clear relation between the deseasonalized CoV and the wMAPE. Since the wMAPE is easier to use and to understand, we argue that the wMAPE is the most appropriate parameter for Wavin.

Moreover, this chapter contains the classifications of the three countries based on the given thresholds. The analysis of the classifications made clear that for some classes there are significant differences between the statistical and final forecast accuracies. This is not always positive; often the judgmental adjustments have a damaging effect on the forecasts, decreasing the forecast accuracy. This should be avoided as much as possible, of which the classification will help.

Because adding qualitative information is normally done on aggregated level, we also analyzed the forecast accuracies on aggregated level. However, there seems to be not much better improvements. This confirms that adding qualitative information can and should be done better in future. The classification will contribute to this.

We also give forecast accuracy targets to the different classes for Country A, Country B and Country C. We set the highest targets for easy to forecast products and also for more important products. Using these targets, one can evaluate more specifically whether the forecasts are done well, instead of the current single target for class A products.

For making the classifications and evaluating the monthly performance we created two Excel tools for Wavin, which they can use easily for all countries. After importing the necessary data, with one push of the button, the classification and performance are the outcomes.

# 5. Implications of the classification

In the previous chapter the parameters and corresponding thresholds are set, and the classifications for the three countries are shown. The next step is to describe what the implications of the classification are in order to know how to deal with the different classes to reach better focus and a higher forecast accuracy. It is important to know what actions should be taken as Paul Saffo, a famous technology forecaster, quoted:

*"The goal of forecasting is not to predict the future but to tell you what you need to know to take meaningful action in the present."*

In Section 5.1 we discuss the overall strategies per class. In Section 5.2, we list more specific how to put the focus on statistical forecasts. Included in this section is the analysis of which statistical methods is most suitable for which class and method combination. Section 5.3 is about the qualitative part of forecasting and we explain in this section which qualitative factors to consider for which class and what the impact of these factors are. Section 5.4 contains the conclusion of this chapter.

## 5.1 Strategy per class

Each class requires another strategy for making the forecasts. Some classes are more important and some are more difficult to forecast. In Figure 27 we summarize the strategies per class which is an adaption of Figure 8 in Section 3.3.

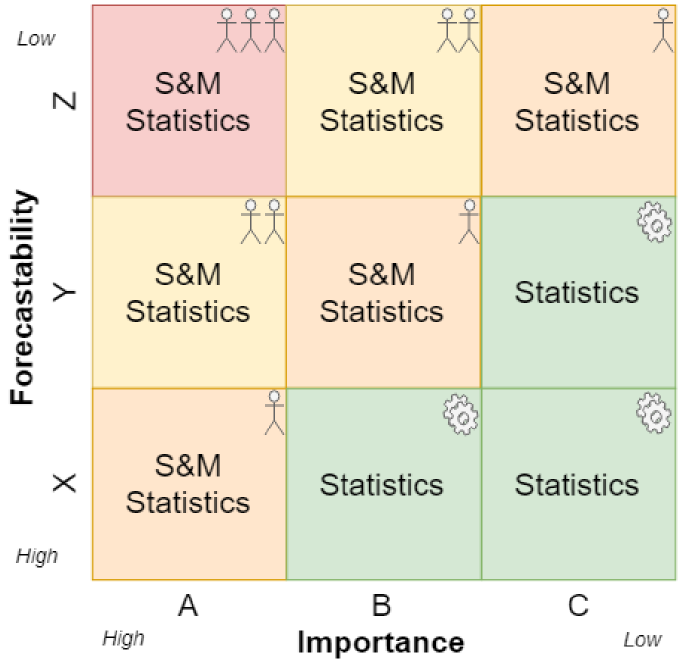



*Figure 27: Strategies per class (adaption from Kepczynski et al. (2018))*

Statistical forecasts should be considered for all classes. For class BX, CX and CY the statistical forecast should be sufficient for the final forecast and can therefore be automatically made. No Sales & Marketing (S&M) information is needed as illustrated in Figure 27. This may only be the case when there is a major factor that the Sales & Marketing come across, which has with a high certainty an impact on the forecast. With any other occasion, the Sales & Marketing should leave the products in these classes alone. They have only limited time and it is better to focus on more important and more difficult to forecast products. For class AX, they should add also very limited qualitative data to the statistical forecasts. Chapter 4 shows that in the past there is barely an improvement of the final forecast over the statistical forecast and there are even some negative effects for this class. More effort should be on the AY and even more on the AZ products. These products are difficult to forecast using statistics and more improvements are possible for these products. Some attention should be on the BY and BZ classes as well. Forecasting class CZ with statistics will not help very much. In this group there are many products which are difficult to forecast and have sporadic demand. It is better to ask Sales & Marketing to keep track of these products when they expect demand of these products. Moreover, for these products a higher safety inventory can be held, since the revenues are low and it is difficult to forecast these products. Therefore, there should not put too much attention for forecasting these products.

Table 16 summarizes what the basic actions per class should be, together with the priorities for making the statistical forecasts and adding the qualitative data. In this table is also indicated how much time should be relatively spend per class. For example, for every hour the Demand Manager has for making the statistical forecast, he/she should spend 15 minutes on class AX and the Sales & Marketing should spend 5 minutes on this class of every hour they spend on making their forecasts. Note that this is rather a suggested recommendation to give the Demand Managers guidelines, than it is the 'golden rule'.

*Table 16: Actions and priorities per class*

| Class | Actions | Priority statistics | Priority S&M | Minutes to spend per hour (statistics) | Minutes to spend per hour (S&M) |
|---|---|---|---|---|---|
| AX | Focus on Statistics, but also ask S&M input | 1 | 6 | 15 | 5 |
| AY | Statistics on a bit lesser extent and more input from S&M | 2 | 2 | 12 | 12 |
| AZ | Little focus on statistics, but emphasize to S&M for their input | 7 | 1 | 3 | 15 |
| BX | Focus on Statistics, only change when big event of S&M | 3 | 8 | 10 | 1 |
| BY | Statistics on a bit lesser extent. And more input from S&M | 4 | 4 | 6 | 8 |
| BZ | Little focus on statistics, but ask to S&M for their input | 8 | 3 | 2 | 10 |
| CX | Focus on Statistics, only change when big event of S&M | 5 | 9 | 6 | 1 |
| CY | Focus on Statistics, only change when big event of S&M | 6 | 7 | 5 | 1 |
| CZ | Very little focus on Statistics, but ask S&M when they expect sales | 9 | 5 | 1 | 7 |
| Total | | | | 60 | 60 |

This table gives both the Demand Managers and Sales & Marketing guidelines how to approach each class. With this table they know where to focus and which groups should be prioritized. How the focus can be done on statistics and qualitative information, we explain respectively in Section 5.2 and Section 5.3.

### 5.1.1 Combining statistical and qualitative forecasts

Ideally, adjusting the statistical forecasts with judgmental input results in an improvement of the forecast accuracy. However, as Chapter 4 made clear, this is not always the case. Although the classification and the described implications will likely decrease the occurrence of harmful judgmental adjusting, we would also argue for using weights for combining the statistical and final forecasts. During the forecast meetings where the adjustments are done, the sales representatives bring their own forecasts which are based on the statistical forecast. However, they are likely to make too large adjustments. This can be best illustrated by examining the error biases of the three countries, which is shown in Table 17 (values are in millions). The total bias is calculated by the sum of the forecasts (statistical or final) minus the sum of the sales of the country. For Country A both biases (statistics and final) are positive, which means that the forecasts are in general too high. The final bias is even almost 9 million higher than the statistical bias. For Country B and Country C the statistical forecasts are in general too low, while the judgmental adjustments result in over-forecasting. This implies that especially for Country B and Country C a forecast that lies between the statistical and final forecast would seem to be the best.

Table 17: Total bias (in millions) of the three countries in 2018

| | Statistical bias | Final bias | Difference | Total Sales |
|---|---|---|---|---|
| Country A | 3,38 | 12,3 | 8,92 | Confidential |
| Country B | -3,79 | 0,87 | 4,66 | |
| Country C | -0,27 | 0,21 | 0,48 | |

This over-forecasting behavior may be a result of the sales managers trying to increase the forecasts in order to ensure product availability as Shapiro (1977) explains. Therefore, combining the forecasts may be desirable. A way to unify the forecasts is to average the different forecasts (Lawrence et al., 1986). We examined what the forecast accuracies would have been in 2018, when equal or optimal weights where used. The optimal weights per class are determined such that the highest forecast accuracy is achieved. We calculated the optimal weights by using the built-in solver of Excel. We calculated the forecast accuracy when the forecast was calculated by the formula: $weight_1$*statistical forecast + $weight_2$*final forecast (with $w_1 + w_2 = 1$). In Table 18**,** Table 19 and Table 20 we list what the forecast accuracies of 2018 per class would have been when the forecasts where combined using the equal and optimal weights.

Table 18: Weights and corresponding forecast accuracies for Country A

| Class | Stat Accucary | Final Accuracy | Optimal Weight Stat | Optimal weight Final | Accuracy optimal weights | Accuracy equal weights | Increase when optimal weights | Increase when equal weights |
|---|---|---|---|---|---|---|---|---|
| AX | 86,8% | 87,4% | 0,44 | 0,56 | 88,2% | 88,1% | 0,7% | 0,7% |
| AY | 81,1% | 82,4% | 0,49 | 0,51 | 83,5% | 83,4% | 1,1% | 1,1% |
| AZ | 49,0% | 26,9% | 1,00 | 0,00 | 49,0% | 38,8% | 22,1% | 11,9% |
| BX | 86,8% | 86,4% | 0,59 | 0,41 | 87,9% | 87,9% | 1,5% | 1,5% |
| BY | 77,8% | 82,5% | 0,14 | 0,86 | 82,5% | 81,2% | 0,0% | -1,3% |
| BZ | 14,7% | 30,1% | 0,05 | 0,95 | 30,2% | 25,5% | 0,1% | -4,6% |
| CX | 83,8% | 86,9% | 0,02 | 0,98 | 86,9% | 86,2% | 0,0% | -0,7% |
| CY | 75,9% | 80,9% | 0,00 | 1,00 | 80,9% | 79,3% | 0,0% | -1,7% |
| CZ | 8,6% | 26,1% | 0,16 | 0,84 | 27,5% | 24,5% | 1,4% | -1,6% |
| Total | 76,3% | 75,4% | | | 78,9% | 77,4% | 3,5% | 2,1% |

When the optimal weights were used for the forecasts of Country A, the overall final forecast accuracy could have been increased with 3,5% compared to an increase of 2,1% when the forecasts would have been averaged. This is mainly due to the 22,1% increase of AZ, where the statistical forecast accuracy was much higher than the final forecast accuracy. In general, the highest weights are given for the final forecasts. This makes sense, since in Chapter 4, the analysis made clear that the judgmental adjustments do not have much damaging results, except for class AZ.

Table 19: Weights and corresponding forecast accuracies for Country B

| Class | Stat Accucary | Final Accuracy | Optimal Weight Stat | Optimal weight Final | Accuracy optimal weights | Accuracy equal weights | Increase when optimal weights | Increase when equal weights |
|---|---|---|---|---|---|---|---|---|
| AX | 74,6% | 74,2% | 0,58 | 0,42 | 75,8% | 75,7% | 1,5% | 1,5% |
| AY | 68,2% | 67,0% | 0,59 | 0,41 | 69,4% | 69,4% | 2,4% | 2,4% |
| AZ | 19,9% | 21,2% | 0,52 | 0,48 | 32,5% | 32,3% | 11,3% | 11,1% |
| BX | 70,2% | 70,6% | 0,45 | 0,55 | 71,8% | 71,8% | 1,1% | 1,1% |
| BY | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| BZ | 0,0% | 6,9% | 0,00 | 1,00 | 6,9% | 4,8% | 0,0% | -2,2% |
| CX | 70,3% | 71,3% | 0,34 | 0,66 | 71,6% | 71,4% | 0,3% | 0,2% |
| CY | 56,9% | 56,1% | 0,87 | 0,13 | 56,9% | 56,8% | 0,9% | 0,8% |
| CZ | 13,1% | 2,8% | 0,92 | 0,08 | 14,2% | 11,7% | 11,5% | 8,9% |
| Total | 49,9% | 49,5% | | | 54,7% | 54,5% | 5,2% | 4,9% |

The optimal weights for Country B resulted in an overall forecast accuracy increase of 5,2%, while it was 4,9% when equal weights would have been used. These equal weights resulted only for class BZ in a decrease, where the final forecast accuracy was significantly higher than the statistical forecast accuracy.

*Table 20: Weights and corresponding forecast accuracies for Country C*

| Class | Stat Accucary | Final Accuracy | Optimal Weight Stat | Optimal weight Final | Accuracy optimal weights | Accuracy equal weights | Increase when optimal weights | Increase when equal weights |
|---|---|---|---|---|---|---|---|---|
| AX | 77,5% | 77,8% | 0,44 | 0,56 | 78,8% | 78,8% | 1,0% | 1,0% |
| AY | 64,9% | 62,6% | 0,83 | 0,17 | 65,5% | 65,1% | 2,9% | 2,5% |
| AZ | 47,6% | 59,3% | 0,00 | 1,00 | 59,3% | 55,4% | 0,0% | -3,9% |
| BX | 75,0% | 68,4% | 1,00 | 0,00 | 75,0% | 72,5% | 6,6% | 4,1% |
| BY | 66,0% | 65,5% | 0,56 | 0,44 | 66,5% | 66,5% | 1,0% | 1,0% |
| BZ | 42,2% | 57,1% | 0,00 | 1,00 | 57,1% | 52,2% | 0,0% | -4,9% |
| CX | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| CY | 64,8% | 59,0% | 0,65 | 0,35 | 65,3% | 64,7% | 6,3% | 5,8% |
| CZ | 13,6% | 21,0% | 0,18 | 0,82 | 21,7% | 20,8% | 0,7% | -0,2% |
| Total | 65,9% | 68,7% | | | 69,9% | 68,8% | 1,2% | 0,1% |

The optimal and equal weights for Country C resulted in increases of respectively 1,2% and 0,1%. Three times a weight of 1 is given; for class AZ and BZ for the final forecast and for BX for the statistical forecast. These are also the classes with the biggest differences between the final and statistical forecast accuracies.

While the optimal weights give the highest forecast accuracies, it does not mean the same weights would be optimal in future. It is based on the data of 2018. However, we can see some patterns in the weights of the three countries. In Table 21 we listed the average, minimum and maximum values of the optimal weights of the statistical and final forecasts of the three countries.

*Table 21: Average, minimum and maximum optimal weights of the three countries*

| | Optimal Weights Statistics | | | Optimal Weights Final | | |
|---|---|---|---|---|---|---|
| | Average | Min | Max | Average | Min | Max |
| AX | 0,49 | 0,44 | 0,58 | 0,51 | 0,42 | 0,56 |
| AY | 0,64 | 0,49 | 0,83 | 0,36 | 0,17 | 0,51 |
| AZ | 0,51 | 0,00 | 1,00 | 0,49 | 0,00 | 1,00 |
| BX | 0,68 | 0,45 | 1,00 | 0,32 | 0,00 | 0,55 |
| BY | 0,35 | 0,14 | 0,56 | 0,65 | 0,44 | 0,86 |
| BZ | 0,02 | 0,00 | 0,05 | 0,98 | 0,95 | 1,00 |
| CX | 0,18 | 0,02 | 0,34 | 0,82 | 0,66 | 0,98 |
| CY | 0,51 | 0,00 | 0,87 | 0,49 | 0,13 | 1,00 |
| CZ | 0,42 | 0,16 | 0,92 | 0,58 | 0,08 | 0,84 |

For some classes, we can clearly give some recommendations for the weights to use. For class BZ, the final forecast should be leading, since the weights for the final forecast for the three countries are close to 1. This class is also one of the classes that is most difficult to forecast using statistics. The average of the statistical and final forecast gives a good forecast for the classes AX, AY and BX, since for this class all countries have an improvement using the equal weights. For the other classes, setting the same weight for all countries is less suitable. The optimal weights differ more and setting an overall weight may mean an increase for one country while a decrease for another. For these classes the optimal weights per country can be used. Close monitoring for the new year is however desirable since the quality of the judgmental adjustments may change resulting in that the optimal weights may change.

## 5.2 The focus on statistics

Currently, the forecasters at Wavin spend about equal time with a bit more focus on class A products for making the statistical forecasts. They correct the outliers and use planning groups to fit the statistical forecast on. With SAP APO the statistical forecasts are made and with this software a variety of statistical methods are available. With the ‘best pick’ feature of SAP, all available statistical methods are fitted on the data to see which statistical method is best for which planning group. With the new classification this approach can still be used. However, some adaptions depending on the class are necessary. In this section we discuss outlier correction, and how the statistics can be increased by using also other methods and method combination. Moreover, we give guidelines which methods are most suitable for each class.

### 5.2.1 Outlier correction

Outlier correction is an important activity for cleaning the history. In past there may have been some market evolvements which resulted for peaks (or lows) in the demand that are only happening once. When these outliers are not corrected, it influences the statistical method chosen. It may happen that the statistical method predicts a similar peak in future, while it probably will not happen again. Figure 28 shows a good example of outliers occurring in the data. The figure shows the monthly Sales (2016-2018) in Country A of the SKU *X*. This group is classified according the thresholds defined in Chapter 4 as class AX.

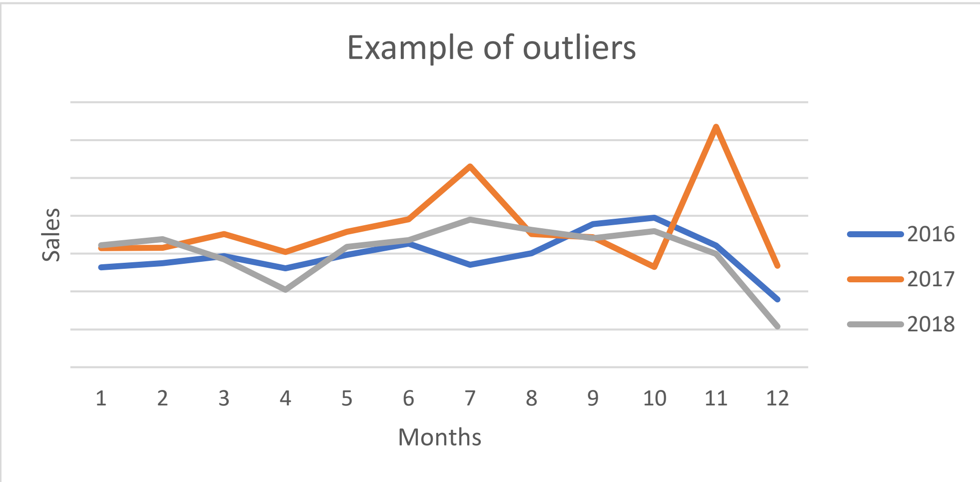



*Figure 28: Example of outliers in data*

There are two clear outliers in the data. One in July of 2017 and an even clearer outlier in November in the same year. These outliers may occurred due to a market event (e.g. promotion). These datapoints are just happening once; in 2016 and 2018 these high peaks did not occur. To avoid that these peaks have an impact on the fitted statistical method it needs to be corrected, before fitting the statistical methods on the data.

We argue that this outlier correction is most important for the easier to forecast and more important products. Therefore, most attention for outlier correction should be on class AX products. These important products have generally predictable demand patterns. However, it may happen that due to market events an outlier could occur. Correcting these outliers avoid that it will have an effect on future forecasts. Outlier correction should be on a lesser extent for BX, CX and AY products. Normally, correcting the outliers for easy to forecast products are also relatively less time consuming than for more difficult to forecast products. Since these products are more forecastable, less outliers occur and thus less time to determine why these outliers occurred and correcting it is needed.

The effort on outlier correction for the other product groups should be minimized. Outlier correction for more difficult to forecast and more random patterns of data will not have a big effect. The first reason is that statistics can simply not predict well the future sales for these products because the demand is more random. Besides, when there are many outliers in past, the chance is high that there will be many outliers in future. Therefore, correcting all outliers may consequently even have a negative impact on the forecasts. Because when including the outliers, a statistical forecast will also predict at least some of the outliers in future.

We recommend that for class AX products the outlier correction is done on SKU level. These important products have generally not many outliers, which makes it possible to do it on SKU level. More important, for these products the statistics is leading and therefore it is key to have a good cleaned history for a better fit of the statistical method. For class BX, CX and AY the outlier correction may be done on a bit lower level of detail. For the other classes, outlier correction will have not a major impact and outlier correction is not really needed. Occasionally, it may be done on high aggregated levels.

### 5.2.2 Statistical methods

For this research we took a deep dive into how to improve the statistical forecasts. We tested also other methods then are used currently by Wavin and we examined whether method combination results in a higher forecast accuracy. We tested the forecasting methods in RStudio. R is a free programming language and there are thousands of add-on packages to do many things of which forecasting is one of them. For this research we used the *forecast* package, containing the most popular forecasting methods. We first explain which methods we used for making the forecasts, after which we explain the dataset we used and the way of working. Then, we give the results of the forecasts, the most used methods per class and the results of method combination.

#### *5.2.2.1 Methods*

In this section we shortly explain the methods that we used for testing. We refer to Section 3.4 for a more elaborate (technical) explanation of the methods. Respectively we discuss the average, naïve, exponential smoothing, ARIMA, Croston and regression methods.

##### Average methods

There are several average methods available. However, for our research we left out the weighted methods and the moving average. These methods require a setup of the parameters manually. The weights and the period over which the moving average is used needs to be set, which makes the forecasts subjective. Since the exponential smoothing methods have similar characteristics but calculates the parameters to optimum, the exponential smoothing methods will most likely outperform these methods.

We, therefore, consider only the *simple average* or *mean* method, which averages the historical data and projects the average of the history into future.

##### Naïve methods

The naïve methods are very simple forecasting techniques. It copies the last observed values into the future. There are several adaptions of this method.

The *naïve* method only considers the last observed value. This last observed value will then be extrapolated into future and will give the same forecast for all other months.

The *seasonal naïve* is similar to the naïve method but differs that it also takes into account the seasonality. This method is therefore suitable for constant data patterns with seasonality.

The *naïve with drift* method is also similar to the naïve method, but differs that it is allowed to change over time. The trend is set equal to the average change of the historical data.

### Exponential smoothing

While the parameters of the weighted and moving averages methods needs to be set in front, the parameters of exponential smoothing can be calculated to optimum. Therefore, exponential smoothing normally gives more accurate forecasts.

The S*imple Exponential Smoothing (SES)* method is suitable for data where there is no clear trend or seasonality. It extrapolates the same value for all periods based on the weights of the past data.

*Double Exponential Smoothing (DES)*, also known as Holt's linear trend method, includes a trend.

The T*riple Exponential Smoothing (TES)*, also known as Holt-Winters' seasonal method includes both a trend and a season.

The automated **E**xponen**T**ial **S**moothing, or ***E****rror,* ***T****rend,* ***S****easonality (ETS)* method fits automatically all the possible exponential smoothing methods and chooses the best method according to the AIC and BIC. It compares the methods using additive or multiplicative errors, without a trend, with a trend or with a damped trend and without seasonality, with additive seasonality or with multiplicative seasonality.

The *Theta* method has shown to be a good forecasting method (Makridakis & Hibon, 2000). It is a special case of the SES where the drift parameter is set to be half the slope of the linear trend fitted to the data.

We left out Brown's linear exponential smoothing and adaptive-response-rate SES in this research, although they are available in SAP IBP. Both are not available in the forecast package for R (Hyndman & Athanasopoulos, 2018). Moreover, research has not proven that the adaptive-response-rate SES is an improvement of the normal SES (Chatfield, 1978; Flowers, 1980; Ekern, 1981). Besides, Brown's exponential smoothing method is very similar to the double exponential smoothing method.

### ARIMA

While exponential smoothing models are based on a description of the trend and seasonality in the data, *ARIMA* models aim to describe the autocorrelations in the data. With R, the package automatically chooses the best ARIMA model based on AIC and BIC. This package also includes the seasonal part of the ARIMA models.

### Croston

The *Croston* method is useful for forecasting intermittent demand. It is the most well-known and popular method for forecasting these types of demand patterns.

### Regression

With regression there can be included different factors that influence the forecast. We consider linear and seasonal regression.

*Linear regression* fits a line through the data and extrapolates the line in future. This method therefore includes some possible trends.

*Seasonal regression* adds another factor involving the seasonality. Therefore, the data would be extrapolated using both the trend and seasonality.

The regression models can also include more explanatory variables that may contribute for better forecasting. Think for example about the GDP that expresses the purchasing power, which may influence the sales patterns. However, we left this out of this research. Investigating which factors have influence and to what extent can be a whole study on its own and including it in this research would make the scope too broad. Besides, this would make the monthly forecasts more time-consuming and complicated, which is not desirable.

### Summary

In Table 22 all the methods, together with whether they include trend and/or seasonality, are shown. In this table is also indicated whether the method is available in SAP APO or SAP IBP. Recall that SAP APO is the current system Wavin uses and SAP IBP is the system Wavin is intent to use in future.

*Table 22: Tested statistical methods*

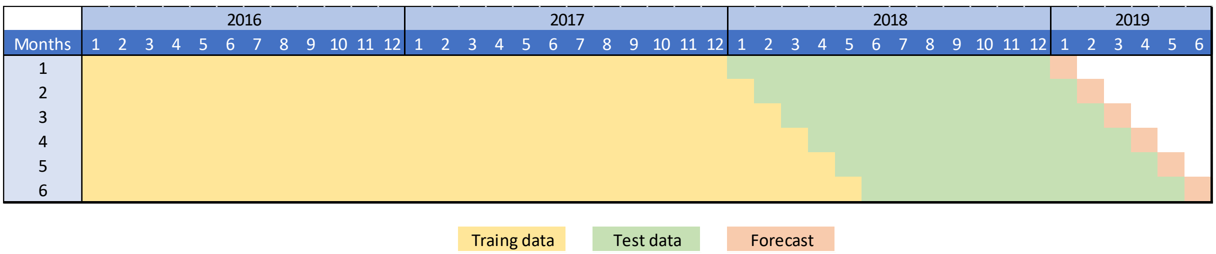
| Number | Method | Trend | Season | SAP APO | SAP IBP |
|---|---|---|---|---|---|
| 1 | Mean | | | Yes | Yes |
| 2 | Naïve | | | | Yes |
| 3 | Seasonal naïve | | Yes | | Yes |
| 4 | Naïve with drift | Yes | | | |
| 5 | SES | | | Yes | Yes |
| 6 | DES | Yes | | Yes | Yes |
| 7 | TES | Yes | Yes | Yes | Yes |
| 8 | ETS | Yes | Yes | Yes | Yes |
| 9 | Theta | Yes | | | |
| 10 | (S)ARIMA | Yes | Yes | | Yes |
| 11 | Croston | | | Yes | Yes |
| 12 | Linear regression | Yes | | Yes | Yes |
| 13 | Seasonal regression | Yes | Yes | Yes | Yes |

#### *5.2.2.2 Dataset and evaluation*

For evaluating the statistical methods we had the monthly data from 2016-2018 and the first 6 months of 2019 of the three countries to our disposal. We used the data from 2016-2018 for choosing and fitting the methods and the six months of 2019 for comparing our results with the results of the forecasts Wavin made. We divided the data of 2016-2018 into two parts; training and test data. The training data is used for fitting the methods, where after the test data can be used to evaluate the forecasting method. This test data is necessary since a good fit of the data does not necessarily mean that it will give a good forecast for the future because it may result in overfitting (Rasmussen, 2004).

While Wavin makes their forecast 18 months ahead, they only evaluate the forecast accuracy of one month ahead. Therefore, we also only evaluate the forecasts of one month ahead. This resulted that we had more data to use for the 2$^{nd}$, 3$^{rd}$ months and so one. In order to keep the same amount of testing data, we expanded the training data for even better estimation of the parameters. In Table 23 the monthly data we used for training, test and forecast is illustrated.

*Table 23: Illustration of the training, test and forecast data*



Outlier correction is normally an important activity for cleaning the history as we explain in Section 5.2.1. However, when there is no information why the outlier happened and there is not much historical data available to see if it is seasonality rather than being a outllier, correcting the outlier can have a negative effect on fitting the model. We did not have the information about why the outlier occurred and we only had three seasons of historical data. Therefore, we fitted the methods on the 'uncleaned' data. Important to note is that the historical data contains the sales and not the demand of the last years. It may have happened in the past that the demand for a certain month was very high, but that it was not forecasted and the production could not supply all the products. This ensured that the highest peaks are already somewhat corrected. Although these facts, we still found it interesting to see the results when the data would have been corrected with the outliers. Therefore, we also give the forecast accuracies of the data when outlier correction is applied. We did this outlier correction using the automated function in RStudio.

We used the grouping of level 7/8/9 to fit the data on. Throughout this report we use this grouping for evaluating all results. This is also the grouping Wavin mostly uses to fit the statistical methods on. To recall, for Country A there are 251 groups and for Country B and Country C there are both 139 different groups.

As we explain in Section 3.6, there are several error measures to use for evaluating the forecasts. For choosing the best statistical method, we only need to evaluate the time series seperately. A scale independent measurement is therefore not needed. We choose for the MAE as the measure to evaluate the methods. This measure is intuitive and less sensitive to outliers than the MSE and RMSE. Moreover, this is also the measure that is used in SAP to choose the best method.

With RStudio it is possible to plot the data and the fitted methods. Using these plots, trend and seasonality can be easily determined. In Figure 29 the plot of all the monthly sales for Country C is given, together with the six months ahead forecast for 2019.

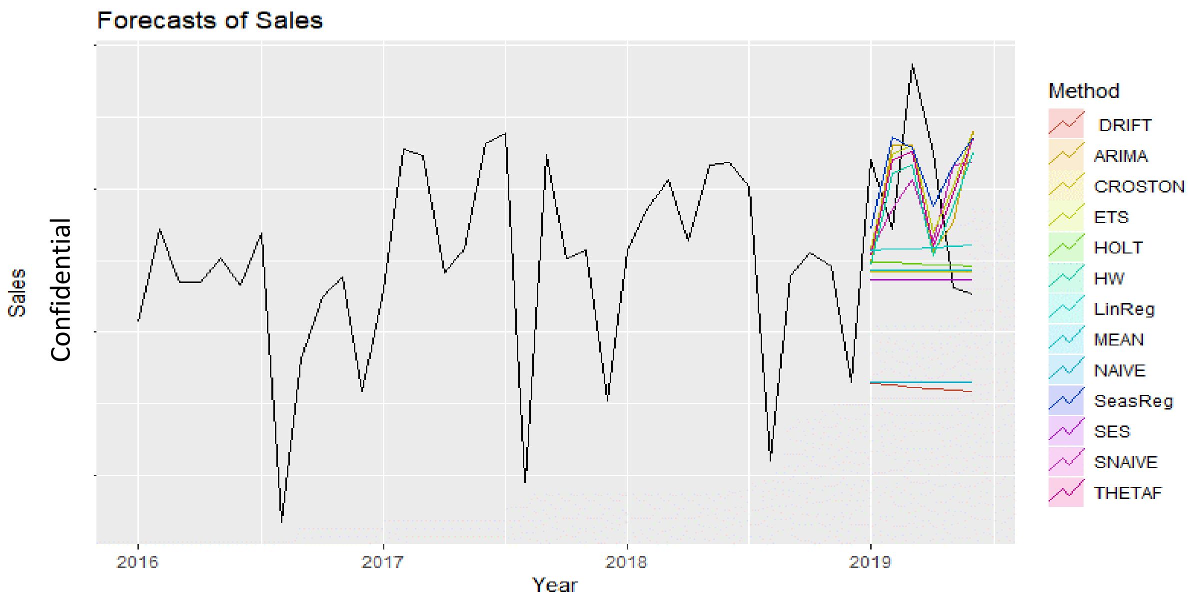



*Figure 29: Time series plot of Total Sales in 2016-2018 and the forecasts for 2019 for Country C*

The three clear troughs visable in the figure, are all for the month August. This is because in this month the factory closes for a couple of weeks for holidays. Besides, December is also a low sales month, because then companies normally want to reduce stocks for the end of the year. These lows in December are also the case for Country B and Country A as can be seen in Figure 51 and Figure 52 in Appendix J.

### 5.2.2.3 Outcomes of forecast accuracies

To evaluate the forecasts we calculated the forecast accuracies (1 – wMAPE) of the forecasts by selecting the method with the lowest MAE on test data. We compared these results with the forecast accuracy of how Wavin has forecasted the products.

The forecast accuracies of the forecasts Wavin made are somewhat different than the forecast accuracies we calculated. This may have several reasons. First of all, we also used other methods that are not available in SAP APO which is used for making the forecasts by Wavin, like the (seasonal) naïve, Arima and Theta methods. Besides, SAP may use different ways or algorithms to calculate the best parameters for the methods than the methods in RStudio do. Another possible reason is that in SAP more historical data was available than we had access to. In Table 24, the forecast accuracies per class of the total six months per class are shown for Country A.

*Table 24: Forecast accuracies of month 1-6 of 2019 for Country A*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Current (SAP APO) | 89% | 86% | 79% | 89% | 83% | 54% | 87% | 82% | 47% | 84,8% |
| New (without outlier correction) | 89% | 86% | 70% | 91% | 78% | 45% | 87% | 80% | 54% | 83,8% |
| New (with outlier correction) | 90% | 86% | 76% | 92% | 78% | 45% | 88% | 82% | 55% | 84,7% |

Using RStudio and the thirteen forecasting methods we forecasted with an accuracy of 83,8%. Surprisingly this is a decrease of about 1,0% of the forecast accuracy when Wavin made the forecast. This can be the cause of the extra methods we used, which seemed to be a good fit for the data of 2018, but not for the first six month of 2019. When we used the outlier correction function, we had almost the same result as the forecasts made by Wavin. The forecasts of RStudio performed the worst for class AZ, BY and BZ compared to the current forecasts. However, for class BX and CZ our forecasts gave better results.

For Country B the forecasts in RStudio give a better result than the currently generated forecasts, as can be seen in Table 25.

*Table 25: Forecast accuracies of month 1-6 of 2019 for Country B*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Current (SAP APO) | 86% | 83% | 53% | 74% | N/A | 61% | 88% | 77% | 20% | 74,8% |
| New (without outlier correction) | 82% | 84% | 66% | 72% | N/A | 62% | 89% | 84% | 42% | 77,7% |
| New (with outlier correction) | 84% | 84% | 61% | 71% | N/A | 63% | 90% | 84% | 33% | 76,8% |

There is an increase of almost 3,0%, when the forecasts are made with all the methods in RStudio. Outlier correction did not have increased results. Especially for class AZ, CY and CZ, the forecasts with RStudio gave better results, while class AX and BX ware less accurate than the forecasts of Wavin.

Also for Country C there are improvements using the extra methods and RStudio as is illustrated in Table 26.

*Table 26: Forecast accuracies of month 1-6 of 2019 for Country C*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Current (SAP APO) | 85% | 56% | 49% | 83% | 74% | 39% | N/A | 52% | 0% | 66,3% |
| New (without outlier correction) | 86% | 62% | 39% | 80% | 76% | 44% | N/A | 55% | 2% | 67,9% |
| New (with outlier correction) | 85% | 61% | 45% | 80% | 76% | 31% | N/A | 51% | 2% | 66,2% |

The forecasts generated by RStudio mean an increase of 1,6%. However, when outlier correction is used it is almost the same as for the forecasts of Wavin. The forecasts for class CZ were close to zero, even with the extra methods we used. The biggest increase is for class AY products with an increase of about 6%.

#### *5.2.2.4 Classification with methods*

The results of the forecasts generated with all the thirteen methods with Rstudio and the forecasts of Wavin showed to have some major differences for at least some classes. This is most likely the case because other forecasting methods are used, but some differences may also be explained that another historical data horizon is used or because of the fact that Rstudio uses other algorithms to determine the optimal parameters. We calculated the percentages how often a method had the lowest MAE on the test data per class. Therefore, the numbers in each column adds up to 100%. In Table 27 these percentages are shown for Country A. The top three highest percentages (when at least 10%) per class are highlighted with green.

*Table 27: Statistical methods per class for Country A*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 0% | 13% | 13% | 2% | 12% | 21% | 3% | 32% | 14% |
| Naïve | 0% | 2% | 8% | 2% | 4% | 12% | 20% | 7% | 22% |
| Seasonal naïve | 68% | 17% | 38% | 67% | 14% | 9% | 57% | 17% | 7% |
| Naïve with drift | 0% | 4% | 4% | 0% | 3% | 5% | 0% | 3% | 4% |
| Croston | 0% | 6% | 8% | 6% | 5% | 18% | 7% | 33% | 14% |
| ETS | 6% | 7% | 4% | 2% | 9% | 13% | 0% | 2% | 9% |
| Arima | 1% | 4% | 4% | 6% | 10% | 4% | 0% | 2% | 5% |
| Theta | 2% | 4% | 0% | 0% | 0% | 3% | 0% | 2% | 5% |
| SES | 1% | 2% | 8% | 4% | 6% | 5% | 3% | 3% | 7% |
| DES | 0% | 0% | 0% | 0% | 0% | 1% | 0% | 0% | 4% |
| TES | 4% | 6% | 8% | 6% | 18% | 6% | 0% | 0% | 1% |
| Linear regression | 1% | 7% | 0% | 2% | 0% | 1% | 0% | 0% | 4% |
| Seas regression | 17% | 30% | 4% | 2% | 19% | 3% | 10% | 0% | 4% |

The seasonal naïve is in the top three of most used methods for all classes except for BZ and CZ. Especially for the more stable products which are in the classes of X it is a popular method (AX: 68%, BX: 67%, CX: 57%). This may explain the difference between the forecasts Wavin made and the forecasts in RStudio. When we leave out the seasonal naïve method for the forecast, more often ARIMA is used and we get results of 84,6% without outlier correction, which is very close to the 84,8% of Wavin. With outlier correction we get even an accuracy of 85,4%. These are very interesting outcomes, since this method seemed to be a good fit according to the test data, but gives less accurate forecasts for 2019.

The mean method has often the lowest MAE on the test data for the class Y and Z products. At one hand this may be surprising, since these classes include the more difficult to forecast products, with more fluctuations. However, at the other hand, the data patterns are so random in these classes, that no method can fit really well and therefore just taking the average is the best.

Seasonal regression is used relatively often for class AX, AY, BY and CX, which are in general the more easy to forecast products. Croston works well for the less important and more difficult to forecast products of which the method is also designed for.

For Country B we see different results than Country A as the percentages in Table 28 illustrates.

*Table 28: Statistical methods per class for Country B*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 17% | 8% | 6% | 25% | N/A | 11% | 0% | 22% | 13% |
| Naïve | 8% | 3% | 7% | 8% | N/A | 5% | 0% | 22% | 31% |
| Seasonal naïve | 8% | 0% | 6% | 0% | N/A | 18% | 0% | 6% | 6% |
| Naïve with drift | 4% | 11% | 3% | 8% | N/A | 8% | 0% | 6% | 5% |
| Croston | 0% | 0% | 3% | 0% | N/A | 10% | 0% | 0% | 6% |
| ETS | 0% | 3% | 4% | 8% | N/A | 4% | 0% | 0% | 3% |
| Arima | 0% | 6% | 6% | 0% | N/A | 3% | 0% | 0% | 3% |
| Theta | 8% | 0% | 7% | 8% | N/A | 8% | 0% | 0% | 7% |
| SES | 17% | 3% | 4% | 8% | N/A | 5% | 0% | 0% | 3% |
| DES | 0% | 17% | 6% | 0% | N/A | 2% | 0% | 0% | 5% |
| TES | 13% | 14% | 13% | 0% | N/A | 6% | 50% | 0% | 4% |
| Linear regression | 13% | 8% | 10% | 17% | N/A | 5% | 0% | 11% | 7% |
| Seas regression | 13% | 28% | 26% | 17% | N/A | 15% | 50% | 33% | 7% |

The seasonal regression had for all classes except for class CZ at least 10% of the times the lowest MAE on the test data. For class CX this method is used half of the times, although we need to mention that only one group is classified in this class. Also the mean method is popular and is used often for both easy to forecast products (class AX and BX and also CY) as well as more difficult to forecast products (BZ and CZ).

In contrast with Country A, the seasonal naïve has not the lowest MAE for many classes. Only for class BZ it is in the top three methods for Country B. The triple exponential smoothing (TES) is, however, more popular especially for class A and CX products.

The results of the forecasting methods used for Country C are shown in Table 29.

*Table 29: Statistical methods per class for Country C*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 0% | 0% | 13% | 25% | 20% | 17% | N/A | 13% | 13% |
| Naïve | 0% | 3% | 13% | 8% | 7% | 11% | N/A | 3% | 36% |
| Seasonal naïve | 37% | 47% | 4% | 0% | 23% | 23% | N/A | 3% | 6% |
| Naïve with drift | 0% | 3% | 0% | 0% | 3% | 4% | N/A | 0% | 4% |
| Croston | 0% | 3% | 8% | 17% | 0% | 3% | N/A | 7% | 9% |
| ETS | 0% | 0% | 17% | 8% | 3% | 2% | N/A | 3% | 4% |
| Arima | 10% | 0% | 0% | 0% | 0% | 3% | N/A | 0% | 3% |
| Theta | 0% | 3% | 8% | 0% | 0% | 7% | N/A | 7% | 6% |
| SES | 0% | 3% | 21% | 17% | 0% | 2% | N/A | 3% | 4% |
| DES | 0% | 0% | 4% | 0% | 3% | 7% | N/A | 13% | 5% |
| TES | 3% | 14% | 13% | 0% | 20% | 2% | N/A | 17% | 3% |
| Linear regression | 0% | 3% | 0% | 0% | 0% | 1% | N/A | 7% | 4% |
| Seas regression | 50% | 22% | 0% | 25% | 20% | 17% | N/A | 23% | 3% |

Seasonal regression is also for Country C a method that has often the lowest MAE on the test data and is therefore used often for making the forecasts. This was only not the case for class AZ and CZ. The latter class had a very low forecast accuracy and no method seemed to give an accurate forecast for this class. For class AZ the ETS and SES are used most often. These methods may have similar characteristics since with the ETS it can have any form of exponential smoothing and thus also of single exponential smoothing.

The seasonal naïve is used almost half of the times for class AY, while more than a third of the times for class AX and a quarter of the times for class BY and BZ. Besides, the mean method was used often for all class B products.

### *5.2.2.5 Outcomes of forecast combination*

Literature is consistent that combining forecasts improves the forecast accuracy of which taking the average of the methods is most common (e.g. Makridakis et al., 1982). However, the number of statistical methods to combine for the best forecast is not clear. While Silver et al. (2017) argues for combining only two methods, Armstrong (2001) found that combining at least five methods gives the best forecast. We also examined for the case of Wavin whether forecast combination contributes to better forecasting and how many forecasting methods to use.

We combined the best fitted two methods, the best fitted three methods up till using a combination of all methods. Besides, we also calculated the forecast accuracy when combining all the methods that include seasonality (see Table 22). We used the simple but, as literature showed, effective approach of averaging the forecasts. In Table 30, the outcomes of forecast combination are shown for Country A. Per combination and per class the increase compared to when only a single method is used is shown. The combination which resulted in the highest increase are highlighted with green. When combining methods did not result in an increase of the forecast accuracy for a class, we highlighted the lowest decrease with red.

*Table 30: Forecast Combination for Country A*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| One method | 89% | 86% | 70% | 91% | 78% | 45% | 87% | 80% | 54% | 84% |
| Combi(Seas) | -0,3% | 2,4% | 0,0% | -1,7% | 3,7% | -0,6% | -1,1% | 0,1% | 10,8% | 1,5% |
| Combi(2) | 0,0% | 0,8% | -1,1% | -1,5% | 3,6% | 2,3% | -1,9% | 0,6% | 6,7% | 0,6% |
| Combi(3) | 0,2% | 0,9% | -0,8% | -1,3% | 3,6% | 1,0% | -1,6% | 1,2% | 5,7% | 0,7% |
| Combi(4) | 0,6% | 1,0% | 1,3% | -1,2% | 4,3% | 0,0% | -2,1% | 1,5% | 5,6% | 1,2% |
| Combi(5) | 0,9% | 1,4% | 1,4% | -1,2% | 5,1% | 0,5% | -2,1% | 1,4% | 6,9% | 1,6% |
| Combi(6) | 1,1% | 1,6% | 4,3% | -1,5% | 5,5% | 0,7% | -2,0% | 1,4% | 6,5% | 2,0% |
| Combi(7) | 1,2% | 1,8% | 7,9% | -1,9% | 5,7% | -0,7% | -2,4% | 1,5% | 5,6% | 2,4% |
| Combi(8) | 1,2% | 2,3% | 9,7% | -1,9% | 4,9% | -2,2% | -2,9% | 1,8% | 5,3% | 2,7% |
| Combi(9) | 1,2% | 2,5% | 10,2% | -2,0% | 4,8% | -1,6% | -3,4% | 1,7% | 5,7% | 2,9% |
| Combi(10) | 1,4% | 3,0% | 10,3% | -2,2% | 5,3% | 0,2% | -2,6% | 1,6% | 7,0% | 3,2% |
| Combi(11) | 1,2% | 3,1% | 10,2% | -2,4% | 5,1% | -0,1% | -2,7% | 2,0% | 7,3% | 3,2% |
| Combi(12) | 0,8% | 3,4% | 11,2% | -2,6% | 5,0% | -0,3% | -2,8% | 2,1% | 7,3% | 3,3% |
| Combi(all) | 0,4% | 3,6% | 12,9% | -3,0% | 4,8% | -0,4% | -2,5% | 2,1% | 7,8% | 3,4% |

Except for class BX and CX, for all other classes, forecast combination resulted in an increase of the forecast accuracy. How many methods to combine differs per class. While for class AY, AZ and CY it has the best results when all methods are combined, for class BZ, combining the only two best methods give the best results. For AZ the highest increase is possible with 12,9%, which would result in a forecast accuracy for this class of about 83%. Then the highest increase is for class CZ with 10,8% when only the seasonal methods are combined. When only one type of forecast combination is used for all products, the best results are get when all methods are combined and it can be increased by 3,4%. When for each class the type of combination is chosen which gets the highest increase, the total accuracy would even be 88%, which is an increase of 4,2%.

For Country B, the results are less significant of which the results are shown in Table 31.

*Table 31: Forecast Combination for Country B*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| One method | 82% | 84% | 66% | 72% | N/A | 62% | 89% | 84% | 42% | 78% |
| Combi(Seas) | 3,7% | 2,9% | -10,3% | 2,8% | N/A | -1,2% | -1,6% | 0,1% | -22,2% | -0,4% |
| Combi(2) | 1,8% | 0,1% | -17,7% | 0,9% | N/A | 2,4% | -0,5% | -3,8% | -3,4% | -3,2% |
| Combi(3) | 2,2% | -0,2% | -5,4% | 1,6% | N/A | 1,6% | -2,5% | -4,1% | -4,9% | -1,0% |
| Combi(4) | 2,1% | 0,7% | -8,5% | 2,3% | N/A | 1,5% | -3,6% | -4,5% | -4,6% | -1,1% |
| Combi(5) | 3,2% | 0,5% | -8,3% | 2,8% | N/A | -1,9% | -3,7% | -4,6% | -3,9% | -1,3% |
| Combi(6) | 3,3% | 0,8% | -7,1% | 3,3% | N/A | -2,2% | -3,9% | -5,3% | -3,4% | -0,9% |
| Combi(7) | 3,1% | 0,4% | -4,9% | 3,8% | N/A | -1,0% | -4,1% | -4,7% | -4,2% | -0,6% |
| Combi(8) | 3,8% | 0,9% | -3,4% | 3,6% | N/A | -1,5% | -5,3% | -3,9% | -8,4% | 0,0% |
| Combi(9) | 3,2% | 1,1% | -3,8% | 4,2% | N/A | -1,0% | -5,6% | -3,8% | -8,4% | 0,0% |
| Combi(10) | 3,2% | 1,2% | -1,4% | 4,1% | N/A | -0,7% | -6,3% | -3,9% | -8,3% | 0,6% |
| Combi(11) | 2,5% | 1,3% | -0,4% | 3,5% | N/A | 0,1% | -6,7% | -4,0% | -7,5% | 0,8% |
| Combi(12) | 1,3% | 1,1% | 1,5% | 3,6% | N/A | 0,6% | -6,8% | -4,4% | -8,8% | 0,9% |
| Combi(all) | 0,9% | 1,1% | 2,6% | 4,1% | N/A | 1,1% | -7,2% | -5,4% | -7,2% | 1,2% |

Forecast combination is only an (almost) consistent improvement for class AX, AY and BX. For the other classes, forecast combination only decreases the forecast accuracy for at least some combinations. The highest increase is for class BX and is when the best nine methods are combined. For class CZ it means that forecast combination decreases the accuracy, and with even over 22% when only the seasonal methods are combined. There is an increase of 1,2% of the overall accuracy when for all classes all methods are combined and an increase of 2,8% when the type of combination is used per class that has the highest increase.

Country C has the highest overall increase in forecast accuracy when forecast combination is used as Table 32 shows.

*Table 32: Forecast Combination for Country C*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| One method | 86% | 62% | 39% | 80% | 76% | 44% | N/A | 55% | 2% | 68% |
| Combi(Seas) | 2,3% | 3,2% | 7,6% | 2,9% | 1,7% | -5,2% | N/A | 0,2% | -2,4% | 1,4% |
| Combi(2) | 2,1% | -1,5% | 5,6% | -2,2% | 1,5% | 4,7% | N/A | -1,2% | 0,2% | 1,9% |
| Combi(3) | 2,5% | 1,2% | 7,3% | -0,4% | 2,9% | 2,8% | N/A | 0,9% | -1,1% | 2,6% |
| Combi(4) | 2,3% | 2,5% | 6,2% | 0,9% | 3,6% | 2,6% | N/A | 2,1% | -0,5% | 2,7% |
| Combi(5) | 2,8% | 5,2% | 7,0% | 1,8% | 2,4% | 2,5% | N/A | 2,7% | -2,4% | 3,4% |
| Combi(6) | 3,1% | 6,4% | 7,4% | 3,2% | 1,4% | 3,5% | N/A | 2,9% | -1,9% | 3,9% |
| Combi(7) | 2,8% | 6,8% | 7,8% | 4,3% | 0,9% | 2,5% | N/A | 2,7% | -0,5% | 3,8% |
| Combi(8) | 3,7% | 6,5% | 6,9% | 4,0% | 0,2% | 2,8% | N/A | 3,0% | -0,8% | 4,1% |
| Combi(9) | 3,7% | 6,5% | 7,5% | 4,5% | -0,2% | 2,6% | N/A | 3,5% | -2,4% | 4,0% |
| Combi(10) | 3,4% | 6,6% | 9,3% | 4,8% | -0,6% | 3,2% | N/A | 3,4% | -2,3% | 4,1% |
| Combi(11) | 3,6% | 6,1% | 9,2% | 4,9% | -0,1% | 2,9% | N/A | 2,6% | -1,2% | 4,2% |
| Combi(12) | 2,6% | 6,3% | 10,2% | 5,4% | -1,6% | 1,4% | N/A | 2,5% | -2,4% | 3,4% |
| Combi(all) | 1,9% | 5,8% | 9,9% | 5,4% | -2,7% | 1,5% | N/A | 2,8% | -2,4% | 2,8% |

There is an increase in forecast accuracy for all classes when forecast combination is used. However, for most classes it means a different number of methods to combine. Especially for class AZ, forecast combination improves the forecast accuracy and is 10,2%. Combining many methods gives also for Country C a good result on an overall level. The increase can be 4,2% when eleven methods are combined. This increase can even be 4,9% when this is differentiated per class.

#### *5.2.2.6 Conclusion*

With the deep dive into the statistical forecast in this section we can conclude that especially for Country B and also for Country C the forecast accuracy can be increased by using more methods and using RStudio. However, for Country A, RStudio only gave lower results. This is the case because for

Country A the seasonal naïve seemed to work well on the test data, but not on the data for 2019. Besides the seasonal naïve, the seasonal regression and also the mean method showed to have often a low MAE on the test data for the three countries. However, clear patterns for specific methods per class are not present. Therefore, giving recommendations for using methods for a certain class is not really possible.

The analysis about forecast combination showed that even higher improvements are possible. For only two classes for two countries there was not an increase in forecast accuracy when combining the best methods. With forecast combination the forecast accuracies can be improved from 1,2% up till 4,9%. These increases are very high, thinking of the research of Syntetos et al. (2010) who showed that a reduction of 1% of the MAPE can be translated into an inventory reduction of about 15-20% and an increase of about 1% for the cycle service levels and fill rates.

## 5.3 The focus on judgmental adjustments

As the figures (Figure 17, Figure 18, Figure 19) about the classifications in Chapter 4 make clear, the qualitative adding is not always done correctly. It happens regularly that these qualitative information decreases the forecast accuracy. This is of course not desirable, especially given the fact that adding the qualitative information is time consuming. With the classifications we show that especially for the relatively easy to forecast products no high improvements and even decreases are the result of adding the qualitative information. These are also the classes where not much improvement is needed. However, decreases in forecast accuracy as a result of adding qualitative information should be avoided as much as possible.

For class BX, CX and CY (almost) no time should be spend on trying to improve the statistical forecast by adding qualitative information. These non-important products are relatively easy to forecast and the focus should be on more important and difficult to forecast products. Only when there is a clear market event that influences the sales, the forecast may be adjusted. For all other classes, qualitative information should be considered at least to some extent as Figure 27 in Section 5.1 illustrates. In this section we discuss which judgmental factors should be considered for which class, together with the type of impact is has on the forecast.

### 5.3.1 Judgmental factors and their impact

In Section 2.4 we list five types of qualitative information Wavin currently considers and in Section 3.5 we explain what is described in literature what influences the forecast. In this section we combine this and explain how these types of information should be used depending on the class. Besides, we also explain what type of impact the factors have according to the four possible effects that the qualitative information can have on the sales, as we describe in Section 3.5.3. These four types are: transient, transferred impact, quantum jump or trend change. Transient is when the factor only influences the data during the period an event is happening. Transferred impact is when the impact is transferred from one period to another, without changing the forecasts of the following periods. Quantum jump is a single event resulting in a permanent down or up change. Trend factors are when it results in a trend change because of an event. We refer to the figures in Section 3.5.3 for an illustration of these impact factors. We do not explain the size of the impact of each factor, since it is very case specific what the extent of the impact is and it is out of scope of this research to explain all the different cases.

#### *5.3.1.1 Open orders*

Orders that are still open and not yet fulfilled should be considered for all classes. When these orders are relatively large it will have a high impact on the forecast. The open orders should be evaluated whether it will be fulfilled before the end of the month or remain open for the coming months. This information should bring Sales to the forecast meeting. The open orders should be analyzed for all product classes, because it can have a major impact on the forecasts of the next months. The impact of the open orders has a transient effect, since it will happen only once, effecting it for the period the open orders will be fulfilled.

#### *5.3.1.2 Open projects*

Some products of Wavin are mostly sold for projects. Therefore, the sales of these products highly depend on whether these projects are happening and whether Wavin is winning the tender. These products are normally very difficult to forecast. There are no clear patterns when these projects are commissioned and even less when Wavin is chosen to be the supplier of the products for the projects. Therefore, products that are mainly project-based are categorized as class Z. Especially for A products, but also for B products, these projects should be evaluated carefully. A good relationship with the customer is key for this. When data is shared by the customer, the probability rate of whether the project is given to Wavin can be estimated better. Wavin only incorporates the projects when the probability rate is high. The impact of the open projects is also transient, since the sales increase has only an effect during the project.

#### *5.3.1.3 Wholesaler trend*

The more stable products are sold to the wholesalers. Mostly by statistics the forecasts can be made accurately for these products. However, for the more important products it is till valuable to analyze the wholesaler trend. This would apply mostly for AY, AX and BY products. By analyzing the trend per product group of the main wholesalers, information can be gathered about whether the wholesaler is planning to buy additional stocking. This could influence the forecasts that cannot be predicted by using merely the statistics. Depending on the situation the impact can be quantum jump or trend change. Quantum jump happens when the wholesaler decides to order more but with the same trend. Trend change is the case when the wholesaler decides to order more on the longer period, increasing the steepness of the trend.

#### *5.3.1.4 Market events*

There are varying market events that influences the sales. Wavin already designed a checklist which market events to consider, which is shown in Figure 31 in Appendix C. The first market event is *sales price management* and should be considered by Sales. Price changes are normally known in advance and are also announced to the customers. If an increase in price is announced to happen in a couple of months, the customers probably want to buy more products in the months before to still be able to get the lower price. The forecasts for these months should be increased. However, on the long run, when the higher price is introduced, the sales may be lower. The customer may shift to other suppliers when possible or will look for cheaper substitute products. A price decrease will have a reversed effect. The customers will wait to buy new products until the price is decreased. This will mean lower sales in the upcoming months, but an increase in the first months when the price is decreased and probably also a long-term increase. The impact can be best understood as trend change or transferred impact. There is a trend change when the price change results that the customers buy with an increasing or decreasing demand. The customer may want to use more or less substitutes or buy some of the products more or less at the competitor. Transferred impact is when the customer still needs to have the same product and where no substitutes are really possible. The sales will be compensated in the next few months. Price changes are known and only made occasionally and should therefore be considered for all classes and specifically for class A.

*Promotions* are the short-term price decreases owned by the Sales & Marketing. These promotions will increase the sales in the months the price is decreased and may decrease the sales in the couple of months afterwards. The impact is therefore most likely to be transferred impact, since the sales change is compensated by a sales change in the consecutive months. Promotions are known like the price changes and should be added to all class products when these promotions are happening.

*Marketing campaigns* done by Marketing have a more long-term impact. The goal of these marketing campaigns is to make the products more known to new and existing customers. Often Marketing already has their expectations and targets how many sales to expect as a result of the marketing campaign, which can be used for the forecasts. The impact will most likely be a quantum jump, since because of the marketing campaigns the products will become more known and more popular to the customers.

*Conditional bonus performance* is when there is a specific short-term incentive for the customer. The customer may be offered discount when they order more products for example. Sales should then use this information for making the qualitative forecast. This is also most likely to be a quantum jump. Discounts are normally given when a certain number of products is ordered. These known conditional bonus performances should be considered for all products.

Wavin continuously innovates and new products are designed. When there is a *product launch*, forecasts need to be made for these products. However, no historical data is available and no statistical forecast can be made. Therefore, it is important to compare the new product launch with previous product launches or comparable products in order to estimate what the forecasts will be. Besides, if the new product substitutes existing products, the forecasts of these products should also be re-evaluated. The product manager is responsible for giving all this information. These new products are most likely to be in class BZ and CZ. Because these products are new, the revenues will be not that high, and there is no historical data which makes it difficult to make a statistical forecast. However, it is important to closely keep track of these products, since a good starting period for the product is key. Therefore, it should be assigned at least for the first months to class AZ, since this class contains the products that are the most closely followed manually. The impact of these new products will be both quantum jump and trend change, since before no sales exists.

The last market event that Wavin considers is c*ompetitive information.* A close examination of the competitors about price or organizational developments is needed to know what effect it has on the forecasts of Wavin. For example, when there are stockouts at the competitor, it means they cannot supply, resulting that the customers will look for other suppliers. Sales should provide this information for the forecast meetings. For example, if the competitors increase their prices, Wavin may expect more sales, since the customer will try to find alternatives. Depending on the type of change of the competitor, it can have any of the four types of impact.

In the literature study in Section 3.5 we found some other market events that may also influence the sales in future and where the forecasts need to be adjusted accordingly. The first is *government policy*, which includes the subsidies, pollution restrictions, import duties and quotas, regulations and safety standards. When this changes, some products may not be allowed to sell anymore or need some adaptions. Besides, prices may increase or decrease, for example when import duties and quotas change for the export products. Since the changes in government policy can be of different nature, the impact can also be different, dependent on the type of change.

The final market event that is important to consider for Wavin is possible *(inter)national crises*. When a crisis happens, demand may be decreased because the customers do not have much money or are careful to spend money. This may affect some, or even all products of Wavin. Therefore, the forecasts need to be adjusted accordingly. The impact may be a trend change or even a quantum jump, when the customers immediately buy less.

#### *5.3.1.5 Sourcing events*

Besides the market events, the sourcing events are also important to considering when making the forecasts. Where the market events have a direct effect on the sales, the sourcing events have a more indirect effect on the sales. When the r*aw material availability* is low for example, some products may not be produced in the same numbers as before for a certain time. This may result that the demand cannot be reached and the sales forecast should have a lower value. When the raw material is available again later, there may be more sales. This results in a transferred impact. However, when there will be no more raw materials in future, it may be a trend change or quantum jump.

The overall *raw material price development* influences the sales, since the price of the products needs to be adapted accordingly. When the prices of the raw materials consistently increase, the sales price also inherently needs to be increased to get the same profit. Then, the raw material price development has the same effect as the price change, which results in a trend change or transferred impact.

For almost all products there is a *phase-in and phase out*. When the product is new, there is a phase-in period. In this period, the sales of the corresponding product increase rapidly. During the phase-out period the product sales are decreasing. This phase-out may happen because other substitute products are on the market or there is simply no demand for these types of products anymore. It is important to keep track of these phase-in and phase-out periods in order to adjust the statistical forecasts. A trend change is the most common impact.

When Wavin receives a *procurement bonus* of their suppliers to buy the products for a lower price, Wavin can use this to create a market opportunity. The products can be sold to a lower price, creating that the products will be sold more. This means a price decrease on short or long term. Consequently, this has the same impact as the price change, which results in a trend change or transferred impact.

A good relationship with the *third-party supplier* is crucial in order to anticipate to possible stock outs of these suppliers. When these stock outs happen, the materials may not be available anymore for a certain time. This means the demand cannot always be satisfied and the sales decrease accordingly. Since stockouts are most often temporary, it will have a transferred impact, since after the stocks are filled again the customers may want to buy extra products to compensate the previous months.

The last sourcing event Wavin considers is *intercompany supplier*. Wavin has almost thirty factories throughout Europe, each having their own product portfolio. The products sold in each country need to be supplied by multiple factories. Therefore, it is important to know the status of each factory, alerting when stock outs are happening or when there are problems with production. A strong alignment is required for this. Depending on the issue, and whether it is long or short term it can have any of the four impacts.

#### *5.3.1.6 Summary of qualitative factors*

In Table 33 we summarize all the qualitative factors, their impact and for which classes the focus should be on. This summary can be used by Sales & Marketing as a checklist to know which qualitative factors to consider and where to put the focus on for which classes. Many factors, especially the market and sourcing events, are marked that the focus should be on A products and then on B&C products. There is not a separated focus advised based on forecastability. This has a couple of reasons.

The first reason is that we expect that for these factors the information is already available. For example, a price increase of a certain product is known by Sales. There does not need to be collected any more information. Therefore, there will not need to be put much more effort to adjust the statistical forecast with the qualitative information. Besides, we expect that these factors are most often the case for more difficult to forecast products and therefore, consequently more attention will be to these products. However, it may happen that for a stable product a rare market event may be the case. Then, we still think the forecast should be adjusted accordingly. We need to remark however, that adjusting the statistical forecasts needs to be done carefully. In particular for the X products. These products have in general already a high statistical forecast accuracy and increasing it can be difficult. The adjustment should only be made when the Sales & Marketing is really sure about making the adjustment.

*Table 33: Summary of qualitative factors and their impact*

| Factor | Description | Impact | 1st focus of class | 2nd focus of class |
|---|---|---|---|---|
| **Open orders** | Will the open orders be fulfilled in next months? | Transient | A | B&C |
| **Open projects** | Are projects happening in next months with probability of at least 90%? | Transient | Z | Y |
| **Wholesaler Trend** | Are there trends at the main wholesalers and are they planning to buy more products in future? | Quantum jump/Trend change | AX/AY/BX | BY/CX |
| **Market events** | | | | |
| *Sales price management* | Are there price changes resulting in short- and longterm sales changes? | Trend change/Transferred impact | A | B&C |
| *Promotions* | Are there promotions resulting in shortterm sales change? | Transferred impact | A | B&C |
| *Marketing campaigns* | Are there planned marketing campaigns increasing the sales short- and longterm? | Quantum jump | A | B&C |
| *Conditional bonus performance* | Are there conditional bonus performances affecting the sales? | Quantum jump | A | B&C |
| *Product launch* | Are there new products introduced into the market? | Quantum jump & Trend change | Classify as AZ | Compare with substitute products |
| *Competitive information* | Are there price or organizational changes at competitors influencing Wavin's sales? | Depending on competitor change | A | B&C |
| *Government policy* | Are there any changes in government policy affecting the sales? | Depending on government policy change | A | B&C |
| *(Inter)national crises* | Are there any (inter)national crisies affecting the sales? | Trend Change/Quantum Jump | A | B&C |
| **Sourcing events** | | | | |
| *Raw material availability* | Are there expected to be shortages on raw material affecting the sales? | Transferred impact/Trend change/Quanmtum jump | A | B&C |
| *Raw material price development* | Are there raw material price changes resulting in a sales price change? | Trend change/Transferred impact | A | B&C |
| *Phase-in & phase-out* | Are there any products in the phase-in or phase-out? | Trend change | A | B&C |
| *Procurement bonus* | Are there any procurement bonuses Wavin receives which may make price changes possible? | Trend change/Transferred impact | A | B&C |
| *Third party supplier* | Are there any stockout at the third party suppliers? | Transferred impact | A | B&C |
| *Intercompany supplier* | Are there any intercompany supplier issues resulting that the demand cannot be satisfied? | Depending on issue of intercompany supplier | A | B&C |

A way to make sure the qualitative information will result in an improvement of the forecast accuracy is by requiring the managers to justify their judgements in writing. Goodwin (2000) found that by documenting, the damaging qualitative adjustments can be decreased from 85% to 35%. Because the classification gives better focus, we strongly recommend spending the extra time available by documenting the adjustments. It is better to make less, more underlined adjustments, than many adjustments resulting that some have a negative effect.

### 5.3.2 Overall remarks

As the Section 5.3.1 illustrates, Wavin already uses a comprehensive list of adding the qualitative information. The question consequently arises: Why is adding the qualitative information not always done correctly? The first answer to this question is that this qualitative information is not yet used in all countries. These factors are recently determined by using Poland as the good example. However, this qualitative information is not yet considered and implemented in all countries. Moreover, although the factors may be known, the type of impact and the degree of the impact can be difficult to predict. Marketing may expect an increase of sales of at least 10%, while this in practice may only be 2% for example. Since there are so many different other (unknown) factors that influences the sales, it is very difficult to give accurate predictions. Therefore, it is always key to be careful when adjusting the forecasts with the qualitative information. We argue that for class X products and especially BX and CX and also CY almost no qualitative adding should be done. Only when there is a major event impacting these forecasts, like a significant price change, or large open orders that with a high certainty will be fulfilled in the next month. This is also the case for AX products. However, for AX products the wholesaler trend may be analyzed as well. Besides, on a bit lower level of detail the qualitative adjustments may be done. However, for these classes, statistics should always have a high weight and qualitative information should be questioned whether it has indeed the expected results.

For AY and BZ, but more specifically for AZ products the qualitative information can have a major impact. For these products it is relatively difficult to make good statistical forecasts and in order to reach a decent final forecast accuracy, qualitative information is needed. Much attention should be spend on these classes, analyzing the market and sourcing events together with open projects and open orders. In 2018, most often for these classes the qualitative information has shown to be of added value. With even better focus, we expect that the qualitative information can be done even better, increasing the final forecast accuracy of these classes. For these forecasts, the statistical forecasts should only be a base forecast. The most emphasize should be on the qualitative information that should be added.

Another issue is that different managers may want to influence the forecasts in varying way. For example, the sales managers often try to increase the forecasts in order to ensure product availability (Shapiro, 1977). This results in lower forecast accuracies, especially because of the fact that Wavin uses the wMAPE to evaluate the forecast accuracies. As Table 9 in Section 4.2 illustrates, an over-forecast has a bigger (negative) influence on the forecast accuracy than an under-forecast. This underlines the recommendation that adjusting the forecasts with qualitative information needs to be done carefully. The classification helps to give the focus needed to do this qualitative information adding correctly.

## 5.4 Conclusion

In this chapter we give the implications of the classification. This chapter helps to understand how to make the approach for each class and where to put the focus. More difficult to forecast and important products need to have more manual input than easier and less important products. Class BX, CX and CY should be automated as much as possible, while class AZ needs the most qualitative input from Marketing and Sales.

Currently, the statistical forecasts are adjusted with the qualitative information from Sales & Marketing. However, they tend to make too big adjustments, which results in over-forecasting. A way to avoid this behavior is to combine the statistical and final forecasts by using weights. When equal weights are used for all classes, the forecast accuracy for Country A, Country B and Country C are increased by respectively 2,1%, 4,9% and 0,1%. When optimal weights calculated per class are used, i.e. the weights that result in the highest increase, the forecast accuracies can be increased with respectively 3,5%, 5,2% and 1,2%.

When the forecasts depend most on the statistics, good outlier correction is crucial, otherwise it may give inaccurate forecasts. Besides, using more methods and with RStudio, for Country C and Country

B the forecasts can be improved with up to 3%. Moreover, combining the forecasts of multiple methods can increase the accuracy even more for all countries with another 1-5%. Our forecasts are mainly made by seasonal naïve, seasonal regression and the mean method, as these methods had often the lowest error on the test data.

It is crucial that correcting the statistical forecast with the qualitative information is done correctly. For knowing which factors to consider for making the judgmental forecasts we made a checklist with all the factors. It is important that Sales & Marketing considers these factors and document the influences in order to reduce wrongly made forecasts. Besides, the Demand Manager should be careful of the over-forecasting behavior of the Sales & Marketing.

# 6. Results of implementation and testing the classification

The classification of Country B as we present in Section 4.4 and Section 4.5 and the suggested implications as we explain in Chapter 5 we tested for two months (July and August of 2019). We give these results in this chapter. Although testing it for only two month is a bit short, we still can give an indication what the use and added value of the classification is. The mathematician Henri Poincaréin said about this:

*"It is far better to foresee even without certainty than not to foresee at all."*

In Section 6.1 we explain how we tested the classification. Then, in Section 6.2, we compare the results of July and August of 2019 with the results of all the months of 2018 (Figure 18 and Figure 21) and with July and August of 2018 only. This is done on both SKU and aggregated level. Section 6.3 is about the experiences of the people working with the forecasts, since the classification is not only for increasing the forecast accuracy but also to give focus and making it easier to carry out the forecasts. We also include guidelines and recommendations on how to implement the classification company-wide in this chapter. We describe this in Section 6.4. The conclusion of this chapter is given in Section 6.5.

## 6.1 The way of testing the classification

Of the three countries in the scope of this research, we chose to test the use of the classification for Country B for several reasons. First of all, Country B is the country with the lowest forecast accuracy as we make clear in Section 2.6. The statistical and final forecast accuracies were both not even 50% over 2018. In Country B there are many project-based sales, which are in general more fluctuating and more difficult to forecast. Another reason that we selected Country B for testing, is that for this country it was possible for us to go to the factory where the forecast meetings take place. During these forecast meetings, the Demand Manager discusses with Sales & Marketing how to adjust the statistical forecast according to the expected market events. Because we were present at two of these meetings, we could explain the Sales & Marketing about the classification and to understand the potential of start using it in practice.

For testing the classification we had close contact with the Demand Manager, explaining the classification and how to use it. Besides, we asked input for possible changes in the Excel tool to make it even more user-friendly. The experiences of the Demand Manager we explain in Section 6.3.

For the test period of July we only showed and explained the classification, using the classification tool. During the meeting, we recognized that another tool for evaluating the monthly forecasts was also desirable. Therefore, we created the second tool explained in Section 4.7 and illustrated in Appendix I. For the forecast meeting in August, we first showed the results of July with this tool, to illustrate the performance. This helped to give more clearance for the Sales & Marketing.

As the results in Section 5.2 made clear, our forecasts generated in RStudio, using more statistical methods and combining (averaging) them, showed to increase the statistical forecast accuracy (for Country B up to 5,7%). Therefore, for testing the classification we also used RStudio, fitting the thirteen statistical methods on the groups (with the same approach as we explain in Section 5.2). Besides, we combined per class specific the optimal combination defined in Section 5.2 (Table 31). The methods with the lowest MAE are used for combining the forecasts. The number of combinations per class that showed to have the best results are repeated in Table 34. In the same table we also list the optimal weights calculated in Section 5.1.1, for combining the statistical forecast with the final forecast. In this section we showed that Sales & Marketing tend to over-forecast, which is the result that they normally make too big adjustments of the statistical forecasts. Combining the forecasts by using weights diminishes this over-adjusting behavior.

*Table 34: Setup for making the forecasts for July and August 2019*

| | AX | AY | AZ | BX | BY | BZ | CX | CY | CZ |
|---|---|---|---|---|---|---|---|---|---|
| Number of methods to combine | 8 | Seasonal | 13 | 9 | N/A | 2 | 1 | Seasonal | 1 |
| Weight statistical forecast | 0,58 | 0,59 | 0,52 | 0,45 | N/A | 0 | 0,34 | 0,87 | 0,92 |
| Weight final forecast | 0,42 | 0,41 | 0,48 | 0,55 | N/A | 1 | 0,66 | 0,13 | 0,08 |

In Chapter 5.3, we also give a list with qualitative information to consider for Sales & Marketing for adjusting the statistical forecast with the qualitative information. Although the judgmental adjustments can be improved by considering all these factors, we could not test it for July and August 2019 for several reasons. First of all, because these months are the holiday period. Therefore, some sales representatives were on holidays for some weeks, resulting that they did not have time to try something new. Secondly, for this research we mostly approached the forecasts in the point of view of the Demand Managers, who make the statistical forecasts and lead the forecast meetings with the Sales & Marketing. We did not have much contact with the Sales & Marketing and requiring them to consider all factors would have been too much. Moreover, more research needs to be carried out how to exactly incorporate these qualitive factors and their impact. Third, there should be carried out some sort of training for the Sales & Marketing to explain the qualitative factors and the necessity of it. We expect that it may take some time before Sales & Marketing fully understands all the factors listed in Table 33.

Although, we could not test the qualitative factors, we still found it necessary to incorporate it in this research. The research is about both quantitative and qualitative forecasting and the classification contributes to know where the focus should be on for which classes. We are convinced that the created checklist is comprehensive and when considered all, it can contribute to better judgmental adjustments. However, as we already stated, it takes time to understand and consider all the factors for the Sales & Marketing and because we only had a short time for testing, we expect that it would not have resulted in major increases in this short term test period.

## 6.2 Forecast accuracies of the tested months of Country B

In this section we evaluate the forecast accuracies of the tested months July and August of 2019. We use two iterations of the outcomes for evaluating the results. The first iteration is the results when only the approaches as illustrated in Figure 27 are used. The classification of the products is where the research was initiated from. The expectation is that the classification helps both the Demand Managers and Sales & Marketing to know where to put the focus on, where also the research question is about. However, a drawback of testing merely the classification is that normally it takes time to introduce a different way of working. Moreover, it was just a pilot, where it was mainly about explaining the use of the classification to the people working with it and seeing where further improvements had to be made. Therefore, the specific approaches where not yet fully followed. However, as Section 6.3 illustrates, the experience is that the classification contributes greatly for making the forecasts, although it may not be expressed in real numbers yet.

The second iteration are the outcomes when the forecasts are made with the fully new suggested approach. We explain the set-up in the previous section and in Table 34. The results of this iteration are more likely to show more quantitative improvements.

Besides the outcomes of these two iterations, we also give the performance of all the months of 2018 as already given in Section 4.4 and Section 4.5 (Figure 18 and Figure 21) and the performance of only July and August of 2018. The values of 2018 are useful because it is based on a longer time period.

However, we need to remark that in the beginning of 2018 the forecasts where not done very well as Figure 34 in Appendix D illustrates. Because this may give some skewed results we also find it necessary to give the values of July and August of 2018 in able to make a comparison of the same months a year earlier.

We first give the outcomes on SKU level, after which we give the outcomes on aggregated level. As we explain in Section 2.3 and Section 5.2, the forecasting methods are fitted on this aggregated level, where after the proportional factors based on the last six months are used for giving the forecasts on SKU level. The results on SKU level have lower accuracies, since on aggregated level the over and under forecasts on SKU level are balancing somewhat. The outcomes on SKU level are important for production to know how much to produce of each material, while on aggregated level makes more sense for Sales & Marketing, since they give their qualitative information on this level.

### 6.2.1 Results on SKU level

The performance of the test period of the weighted average of July and August of 2019, when only considering the different approaches per class (new 1) and when also included (combining) other statistical methods (new 2), compared to the performance of last year is illustrated in Table 35.

*Table 35: Test results on SKU level compared to performance of year before*

| | Statistical | | | | Final | | | |
|---|---|---|---|---|---|---|---|---|
| | July/August 2019 (new 1) | July/August 2019 (new 2) | Whole 2018 (old) | July/August 2018 (old) | July/August 2019 (new 1) | July/August 2019 (new 2) | Whole 2018 (old) | July/August 2018 (old) |
| AX | 89% | 86% | 75% | 83% | 86% | 88% | 74% | 85% |
| AY | 67% | 67% | 68% | 73% | 62% | 65% | 67% | 76% |
| AZ | 31% | 40% | 20% | 47% | 51% | 47% | 21% | 32% |
| BX | 59% | 61% | 70% | 66% | 50% | 55% | 71% | 62% |
| BY | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| BZ | 35% | 36% | 0% | 5% | 37% | 37% | 7% | 1% |
| CX | 55% | 57% | 70% | 68% | 45% | 49% | 71% | 74% |
| CY | 49% | 50% | 57% | 62% | 44% | 49% | 56% | 59% |
| CZ | 24% | 24% | 14% | 14% | 16% | 24% | 5% | 20% |
| All | 59,7% | 60,6% | 49,9% | 62,7% | 59,4% | 61,1% | 49,6% | 60,9% |

Comparing the outcomes of the test period of July and August 2019 (new 1) with the performance of whole 2018 shows some major differences for both the statistical and final forecast accuracy. This is especially the case for the classes AX, AZ, BX, BZ and CX. With the classification the percentages of these classes are all much higher except for BX and CX. These latter two classes are in general easier to forecast and the suggested approach is to focus (almost) only on statistics. We can conclude that this suggested approach was not completely followed with the test period. The guidelines for these classes is to only adjust the forecasts of these products when there is a clear market event. However, the final forecast accuracy is 9% and 10% lower for respectively class BX and CX of the test period (new 1).

However, these two classes are also less important. Class AX has a higher statistical and final forecast accuracy than whole 2018 and July/August of 2018. Moreover, class AZ has also a higher statistical and final forecast accuracy than whole 2018 and a higher statistical forecast accuracy than July/August of 2018. Class AY of the new method (new 1) has similar statistical forecast accuracy to last year, but the final forecast accuracy of this class of July/August of 2018 has a significant higher forecast accuracy than the new proposed method.

On an overall level, we can conclude that the new approach (new 1) only is an improvement when comparing it to the whole year of 2018 (about 10%), but is a decrease when comparing only to July/August of 2018 (about 1,5-3%). However, we need to remark that there are many factors (of which some may be unknown) influencing the actual sales, resulting it is different than expected, as we explain in Section 6.2.3.

The complete new approach which is iteration 2 (new 2) has some improvements compared to the results of iteration 1 (new 1). On an overall level it resulted in an increase of almost 1% of the statistical forecast accuracy and about 1,5% increase of the final forecast accuracy. Only for class AX the statistical forecast accuracy was lower with the second iteration. However, the final forecast accuracy of the second iteration for this class was higher. While for class AZ the improvement of the final forecast accuracy compared to statistics was 20% for iteration 1, it was only 7% for iteration 2. This is not only because the statistical accuracy of iteration 2 was already higher, but also because we weighted the statistical forecast with the final forecast with iteration 2. Normally only the final forecast is used (new 1), but with the explained weights, less risks and less probability of over-forecasting is happening. However, the downside is that when the qualitative information is actually very good it is not completely used, since it is also weighted with the statistical forecast. This also emphasizes that the weights for combining the statistical and final forecast should be recalculated regularly.

Summarized we can conclude that the numbers of the first iteration does not seem to improve the forecast accuracy, but incorporating extra methods, combining them and using weights for combining the statistical and final forecast increases the overall statistical and final forecast accuracy with respectively 0,9% and 1,7% (compared to new 1).

### 6.2.2 Results on aggregated level

We give a similar table as in the previous section, but now on aggregated level. This is illustrated in Table 36.

*Table 36: Test results on aggregated level compared to performance of year before*

| | Statistical | | | | Final | | | |
|---|---|---|---|---|---|---|---|---|
| | July/August 2019 (new 1) | July/August 2019 (new 2) | Whole 2018 (old) | July/August 2018 (old) | July/August 2019 (new 1) | July/August 2019 (new 2) | Whole 2018 (old) | July/August 2018 (old) |
| AX | 96% | 90% | 76% | 86% | 90% | 91% | 77% | 90% |
| AY | 85% | 84% | 81% | 87% | 76% | 81% | 77% | 90% |
| AZ | 58% | 65% | 28% | 63% | 84% | 75% | 63% | 53% |
| BX | 83% | 84% | 83% | 82% | 62% | 75% | 85% | 76% |
| BY | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| BZ | 84% | 74% | 35% | 49% | 81% | 81% | 41% | 44% |
| CX | 67% | 70% | 86% | 87% | 52% | 58% | 89% | 92% |
| CY | 70% | 66% | 79% | 86% | 61% | 66% | 79% | 82% |
| CZ | 45% | 43% | 33% | 32% | 37% | 44% | 24% | 40% |
| All | 80,5% | 79,9% | 69,7% | 77,9% | 78,4% | 80,2% | 74,1% | 78,0% |

The aggregated statistical forecast accuracy of the first iteration (new 1) for class AX is close to 100%, which shows that the forecast was very close to the actual sales. Another remarkably well performing class is class BZ with a percentage of 84%. Especially this class forecasted much better than whole 2018 and July/August of 2018 for both statistical and final forecast accuracy. However, on SKU level the percentages were only about 35%. This implies that the SKUs in the groups of this class are fluctuating much within the aggregated group.

Class CX and also class CY are performing worse with the new method than last year. However, this may also have some other reasons. For example in class CX there is just one group. The sales of this group were relatively stable in 2018, making it easier to forecast. However, in 2019 the sales were more fluctuating, resulting that it is more difficult to forecast. Figure 30 shows that for June 2019 there was a high peak in the sales of the group in class CX, while the sales in July and August of 2019 where low. This is likely to be the cause of the lower forecast accuracies of this group.

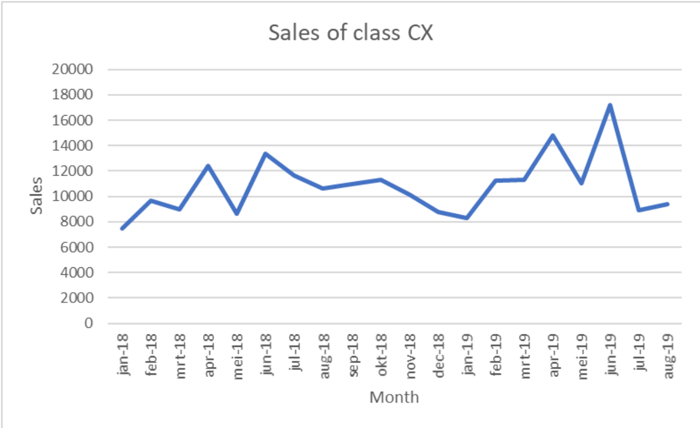



*Figure 30: Sales of class CX of January 2018 till August 2019*

When the fluctuating sales pattern of this group in class CX is going to continue in future, which makes it more difficult to forecast, it may need to be assigned to another class. This emphasizes that it is key to reconsider the classification once or twice a year, to avoid that a product is in the wrong class as a result of that the demand patterns have changed. The created tool makes it possible to make the classification easily and fast.

Moreover, remarkable is that the overall statistical forecast of the first iteration (80,5%) is higher than the statistical forecast accuracy of the second iteration (79,9%) and of July/August of 2018 (77,9%), while on SKU level it had a lower forecast accuracy than the other two. The forecasts on SKU level are based on the proportional factors of the last six months. The difference of 2018 can be explained by the fact that the SKUs were more in line with the proportional sales of the six months before. The proportional factors and sales where, however, the same for the results of iteration 1 and 2. The difference can be explained that for some groups it may be over-forecasted with one approach, while under-forecasted for another approach, resulting in the slight differences.

Comparing the overall percentages on aggregated level of the two new methods with the percentages of last year, we can conclude that the new methods are indeed an improvement for both statistics and final forecasts. While the final forecast accuracy of the first iteration is an improvement of 4,3% and 0,4% of the forecast of respectively whole 2018 and July/August of 2018, the final forecast accuracy of the second iteration was an improvement of 6,1% and 2,2% of the forecasts of respectively whole 2018 and July/August 2018. The improvements in forecast accuracies where less significant on SKU level. However, this is more likely to be the result of the more fluctuating behavior of the SKUs, than that the forecasts where not done well.

### 6.2.3 Overall remarks

While the results of iteration 2 may seem to improve the forecasts, iteration 1 does not seem specifically to show great proven based improvements. Although the overall forecast accuracies of July and August of 2019 are significant higher than the forecast accuracies of all the months of 2018, it is similar to the same months a year earlier. This can have several reasons, like problems with the production and the fact that we could only test in the holiday period. These are just some of the many factors that may have resulted that the actual sales are different from the forecasts. Besides, as the *basic rule* of forecasting says: 'forecasts are always wrong', it is simply not possible to accurately forecast (most of the times). However, key is evaluating the performance to determine why it was inaccurate and how it can be made accurate. The classification contributes to this. With the classification tool, with one glance the forecaster can see easily where the forecasts where good and where not. Upon next, the forecaster can try to find out what resulted in the lower forecast accuracies. Some unforeseen matters may have happened which resulted in the lower forecast accuracy and the current forecast accuracy may have been the best possible. Documenting these issues will contribute for understanding and for making future forecasts. Moreover, as Ross (2004) explains, when replenishment planning and safety stock calculations take account of true measured forecasts inaccuracies, even high errors can be managed to provide acceptable customer service levels.

## 6.3 Experiences of Demand Manager and Sales & Marketing

As we explain, we are convinced of the great use of the classification (iteration 1) although the results may not directly show big improvements. This is because the results are more likely to be visible over a longer period, where it is in the new nature of working, where the evaluation of the forecasts will also help for understanding. This is also in line with what the Demand Managers and Sales & Marketing expect. We elaborate on this in this section with first describing the experiences of the Demand Manager after which we summarize the experiences of the Sales & Marketing.

### 6.3.1 Demand Managers

We had several meetings with the Demand Manager of Country B. During these meetings we explained the classification, the importance, the added value and the use of it. The first reaction was very positive and the Demand Manager was curious about the outcomes. The Demand Manager explained that she experienced that making the forecasts are among others workload intensive, the Sales & Marketing are not always able to give good forecasts, sometimes statistical forecasts do not give an accurate forecast at all, there is a large portfolio of the products, and it is not performing always as expected. Besides, the Demand Manager faced the challenge to get the right balance between the statistical and final forecast. All these factors are also given by Kepczynski et al. (2018) to be common challenges for many forecasters.

The Demand Manager experiences to have a very full working schedule, where other projects are also in the job description. This makes that the Demand Manager does not have much time for making the right forecasts. There is barely time for good outlier correction, error handling and evaluating the statistical forecasts. With the classification, the Demand Manager can now put the first focus on the products where making a statistical forecast makes the most sense, i.e. gives the most accurate forecasts.

The Demand Manager is the end responsible of the forecasts. Their task is not only to make good statistical forecasts, but also to lead the forecast meetings where the Sales & Marketing give their qualitative input. The sales representative normally either comes with information to raise (or lower) the statistical forecast with a number or they come with their own forecasts. In both cases it is important that the Demand Manager not unquestioningly accept these forecasts. They need with a good discussion reach a consensus of the statistical and final forecast. Ofcourse, the Sales & marketing are also hold responsible for the qualitative information, but the Demand Manager should ask for reasoning of the changes.

The classification helps to get a better consensus between the statistical and final forecast as is also experienced by the Demand Manager. Showing the results and explaining the classification helps to persuade the sales representatives for making the right forecast. As the Demand Manager explained, it helps to better lead the forecast meetings.

### 6.3.2 Sales & Marketing

The sales representatives were also in general positive about the classification. They saw the usefulness of the classification, but they first had to see the actual outcomes to be completely persuaded. As with the classification, it means that for some products they may shift the focus to more important and more difficult to forecast products. Especially, classes like BX, CX and CY they may leave alone. The sales representative saw the added value of this. However, the differentiated target setting may push them to come well prepared to the forecast meeting. Besides, because the performance of the qualitative information is also presented to them, it encounters them how well they did their forecasts of the last month. As we illustrate in Chapter 4, and also in this chapter, it is often the case that their input decreased the forecast accuracy. This may be a bit confronting, and may explain the bit reserved reactions of the sales representatives.

Besides, with the differentiated target setting, it really pushes both Demand Managers and Sales & marketing to give accurate forecasts. Both need to spend time and effort for making the forecasts. This may result for a bit more work for some classes, but also less work for other classes.

## 6.4 How to implement and matters to consider in future

While we give the classifications in Chapter 4, the implications in Chapter 5 and the results of the test months in this chapter, we need to give some guidelines how to implement the classification in order that it can be used throughout the whole company. We explain this in this section. We also explain which matters to consider in future.

### 6.4.1 How to implement companywide

As we explain in Section 4.7, we made an Excel tool using VBA for making the classification. With one push of a button the classification can be made on all types of groupings. It is important to make the classification regularly, since the demand of the products over the past cycles is crucial for the classification. Gudehus (2006) recommends making the classification monthly, quarterly or at least yearly, depending on the industry the company is in. For Wavin, we recommend performing the classification at least twice a year. Currently, Wavin also updates their ABC classification two times a year. If it is done more often, some products may be classified into different classes too often. This makes that the Demand Managers and Sales & Marketing needs to shift their focus too often and it can become confusing. If it is done less regularly, the products may be classified in the wrong class for too long, resulting in a wrong focus. When it is updated twice a year, the Demand Managers and Sales & Marketing know the focus for the next six months.

The classifications presented in this report is only for three countries. However, there are over twenty countries for which forecasts need to be made. The Excel file should therefore be distributed and explained to all the Demand Managers in Europe. Besides, the implications of the classification should be explained to the Demand Managers but also to the Sales & Marketing department since they will also be affected by it. The presented figure (Figure 27) and tables (Table 16, Table 33) in Chapter 5 are useful for understanding how to deal with the classifications.

For the Demand Managers an extra training should be carried out to help them to use the Excel file and to give them guidelines how to deal with the classification. At the end, they are responsible for making the forecasts. Moreover, markets differ. We recommend using the same thresholds for all countries as for Country B and Country C. However, as we saw with Country A, it may be the case that for some countries the forecast accuracies are significantly higher (or lower). Then, we argue that the thresholds should be adapted accordingly. Therefore, a good analysis for all countries should be made, determining what thresholds should be used. Based on these thresholds and the performance of the last 12 months, the targets can also be set accordingly.

Besides, as the results showed that improvements can be made using more methods and combining them, we recommend to also implement this. However, using RStudio where the extra methods are available is not desirable, since this would require a change in software, where the Demand Managers have not the skills for. In the near future, Wavin is shifting to SAP IBP, where most of the tested methods are also available. Therefore, when Wavin is going to use this new version of SAP, it will most likely also increase the forecast accuracy. Moreover, method combination is also a feature in this new SAP, which then can also be used.

The weights for combining the statistical and final forecast accuracy resulted also in an increase of the forecast accuracy. Therefore, for all countries these weights should be calculated and used for making the forecasts. However, we need to mention that these weights are most likely to be less valuable when the classification is introduced and followed. We expect that when using the classification, it decreases the damaging effect of adding the qualitative information. Besides, for class X products, (almost) no qualitative information should be added. Therefore, for this class the final and statistical forecast will be very similar. Moreover, for class Z, adding qualitative information is likely to improve the statistical forecasts much, which will result that the final forecast is much more accurate.

### 6.4.2 Matters to consider in future

There are several matters to consider in future. This is because new products may be introduced or forecast accuracies may change. When new products are introduced, we recommend assigning the product to class AZ for the first six months. These products may not yet have much revenues, but it is important to spend time for making the forecasts for these products. Besides, making statistical forecasts is not possible, because there is simply no historical data. Class AZ is the class where the highest focus is on manually making the forecasts. The Sales & Marketing, discussing it with the product manager, should provide the forecasts of these new products, until the behavior of the sales patterns becomes steadier.

The current thresholds for the classes are determined based on the current situation. While the thresholds for all countries should be determined for implementing the classification, in future the situation may change. It may happen that in future more products will be easier or harder to forecast using the statistics. Then, the thresholds and also the targets may need to change. Therefore, we recommend that every year an analysis should be made whether the thresholds and targets should be changed to be more updated to the current situation.

## 6.5 Conclusion

In this chapter we evaluate the results of the test period in July and August of 2019. We do this using two iterations. One iteration where we only took into account the classification and the different approaches and foci. The other iteration also takes into account, when the statistical forecasts are made using the proposed outcomes of Section 5.2 and when the final and statistical forecast is weighted (Section 5.1.1). We compare these results with the performance of whole 2018 and with July and August of 2018.

While iteration 1 shows significant higher forecast accuracies than whole 2018, it has some lower accuracies compared to July/August of 2018. This does not directly proves that the classification is an improvement. Iteration 2, however, shows increases compared to iteration 1 and to the results of the year before. It has the highest overall final forecast accuracy on both SKU and aggregated level, implicating that applying the complete new method is indeed an improvement.

However, we need to mention that the added value of the classification (iteration 1) is likely to be more visible on the long term when both the Demand Managers and the sales representative are more used to the new way of working, seeing the full added value of the classification.

The experiences of the Demand Manager and the sales representatives are more promising. Especially in the point of view of the Demand Manager, the feedback is that it gives more focus for the limited time they have to spend for making the statistical forecast. Moreover, it gives more guidance for the forecast meeting to know where to depend more on statistics and where on market events. Besides, the sales representatives see also the added value, to know for which groups to gather most market intelligence for.

In order to implement the classification throughout the more than twenty countries a training for all the Demand Managers is needed. They are the end responsible for making the forecasts. Moreover, the Sales & Marketing of all the countries should also be updated about the classification, so they know where to put the focus. The thresholds should be determined per country and the targets needs to be agreed on. Preferably, the classification would be updated twice a year, to adapt to the possible changes. Important is to adapt the targets and thresholds when needed as well.

# 7. Conclusions and recommendations

You entered now the end of this report. In this last chapter we give the conclusion, contribution to literature and practice, and recommendations for further research. This chapter summarizes the main findings of the research and explains how to move on. The future is what lies ahead for all of us as the quote of the American inventor Charles F. Kettering illustrates:

*"My interest is in the future because I am going to spend the rest of my life there."*

The chapter starts with the conclusion in Section 7.1, where we give an answer to the main research question posed in the first chapter of this report. Then we explain in Section 7.2, the contributions of this research to both literature as well as practice. The chapter ends with the recommendations for further research in Section 7.3.

## 7.1 Conclusions

All the chapters in this report lead to giving an answer to the main research question as we formulated in Chapter 1:

***Main Research question:*** *How can Wavin improve its forecast accuracy for different types of products for different markets by putting the right focus on quantitative and qualitative methods?*

Like many other companies, Wavin makes their forecasts by first fitting statistical methods on the historical data after which it is adjusted by qualitative factors, like promotions, marketing campaigns or tenders.

However, how the quantitative and qualitative forecast should be combined differs per product. While for the stable products, quantitative methods will work well, for other products it can only be forecasted accurately using qualitative information. Classifying the products according to importance and forecastability, the approaches can be defined for the products. This is called differentiated forecasting (Kepczynski et al., 2018).

This is what we did in this research, with the case of Wavin. Using the scope of Country A, Country B and Country C, we determined the thresholds for the revenues (ABC) and statistical forecast accuracy (XYZ), making the classifications. Each of the nine classes, requires another forecasting approach. While for the more important and more difficult to forecast products, much qualitative information is needed, for the easy to forecast products with low importance, it can be more automated, fitting forecasting methods on the historical data.

By giving the right focus, the forecast accuracy is likely to increase, given the fact that currently for Country B and Country A the statistical forecast accuracy is higher than when the forecasts are adjusted with the qualitative data.

Since Sales & Marketing tend to make too big adjustments of the statistical forecast we also propose to set different weights for balancing the statistical and final forecast. Using equal weights (average of statistical and final forecast) the accuracy can be increased by 0,1-4,9% and using optimal weights it can be increased with 1,2%-5,2%.

Included in this research is a deep dive into the statistical and qualitative forecasts. For improving the statistical forecasts, we tested multiple statistical methods. Adding extra statistical methods showed to improve the forecast accuracies of up to 3%. The most common methods are the seasonal naïve and seasonal regression. Furthermore, averaging multiple forecasting methods is likely to improve the statistical forecasts with an extra 1-5%.

Besides, we listed seventeen qualitative factors that may affect the demand of the products. We made a list with these factors and the corresponding type of impacts and for which classes the focus of gathering this qualitative information should be on. This list can be used by Sales & Marketing for making their forecasts, adjusting the statistical forecasts.

The results of the implementation and test period of merely the classification in July/August of 2019 did not show any proven based improvements. However, the experience is that it helps for making the forecasts giving the right focus. Besides, the differentiated target setting helps to evaluate the forecasts massively. Therefore, the results are more likely to be present on the longer term. The results of also incorporating the extra statistical methods and combining them and weighting the statistical and final forecast seems to be indeed an improvement. The percentages are an improvement of the old approach, which confirms the outcomes of the deep dive in the statistics.

## 7.2 Contribution of the research

In this section we explain how this research contributes to both literature and practice. With the contribution to literature we describe how we filled the gaps in literature. With the contribution to practice we explain why this research is important for Wavin and we give recommendations how they can use this research in their business.

### 7.2.1 Contribution to literature

As we state in Section 3.9, we tried to fill five gaps in literature with this research. We can conclude that we indeed filled these gaps with our research. The first gap (I) we filled with our research is how to make the classification based on which parameters and which thresholds. While in literature, the coefficient of variation (CoV) of the sales is used, we would argue for another parameter; an error measurement. In literature, a high CoV value implicates that the sales are more random, which decreases the forecastability. However, seasonal patterns also increase the CoV, although with seasonal statistical methods it may still be easy to forecast. Although also other error measurements may be useful for other companies, we would argue for using the wMAPE, which is scale independent and easy to interpret. Wavin already uses this measurement, which makes it easy to use. The thresholds of wMAPE for making the classification should be based on the situation. As the three countries showed, even though it is all one company, markets differ, and different thresholds needed to be determined (II). Therefore, one classification with the same thresholds for all markets is not desirable.

Besides, with this research we elaborated on how to deal with the different classes (III). In this report we explain how to do the specific focus on the statistics and the qualitative information. We explain how the classification can be used and what qualitative factors to consider for which classes.

With the deep dive into the statistical forecast we also give guidelines, which statistical methods are most suitable for which classes (IV). Moreover, our research contributed to literature in the way that we showed that method combinations results in an improvement of the forecast (V). Depending per class the amount of methods to combine differs, but on an overall level, combining many methods (11+) results in the highest increase.

### 7.2.2 Contribution to practice

This research also contributed to practice on multiple areas. Wavin, which commissioned the research, is intending to implement the classifications as presented in this research. The plan is to give a workshop to all Demand Managers about differentiated forecasting, explaining the classification tool. This will be done on the short term, in order to test it for multiple countries and setting it up. From January 2020 on, Wavin is going to use the classification company-wide, throughout whole Europe. Wavin is convinced of the added value of using the classification, to give the right focus for both the Demand Managers and the Sales & Marketing. It will help them to be more efficient, focusing on the aspects that matter, instead of losing time on the aspects that they do not have a (positive) impact on. Moreover, the classification enables better target setting which can not only be differentiated per country but also per class. This helps to evaluate and compare the performance of the forecasts massively.

The results of the part of the statistical methods also contributes to practice. It helps Wavin to decide for shifting to SAP IBP from SAP APO, where more statistical methods are available. Our results showed that adding the extra methods increases the forecast accuracy up till 3%. An even higher increase of the forecast accuracy can be achieved when method combination is used meaning an extra increase of up till almost 5%. The results of the test period of Country B show slightly lower improvements, but confirms the added value of it. Therefore, we recommend to Wavin that they consider using more methods, maybe after testing it on a longer term.

Besides, the checklist with the qualitative factors can be used by Sales & Marketing. This checklist helps them to consider all factors that may have influence on the future sales, altering the forecasts accordingly. We recommend that a training or a workshop is carried out, explaining each qualitative factor and how it should be considered and can impact the sales.

The last recommendation for Wavin is to consider using weights, for combining the statistical and final forecasts. In this research we showed that in general Sales & Marketing tend to make too big adjustments, resulting in over-forecasting. Using different weights per class for the statistical and final forecast will most likely increase the forecast accuracy. Another way to decrease the sometimes damaging effect of adding the qualitative information, is to let the Sales & Marketing document their adjustments. This makes sure that they really need to think about the changes they want to make, explaining and giving reasons for it. In this way, only the changes where they are really sure about are taken into consideration.

## 7.3 Recommendations for further research

Although the results presented in this research are promising, more research is needed to confirm it. Besides, during the process of this research we came across other interesting fields that were out of scope of this research but which are interesting to examine. We discuss some of these recommendations in this section.

First of all, because we could only test the classification for two months in the holiday period, the outcomes were not that convincing. Besides, we expect that it takes time to fully incorporate the classification, enabling that it becomes in the new nature of working. Therefore, further research where close monitoring of the implementation of the classification is carried out for multiple countries, holding multiple interviews is needed. In this way, a more proven based result of the classification can be achieved.

Besides, with this research we focused on the case for Wavin. Interesting to know would be to examine if a similar approach of the classification would work for other companies as well. Further research can, therefore, be focused on working out the classification for multiple companies. Point of interest is determining the best parameter to use for the XYZ distinction with corresponding thresholds. This can be done for multiple companies in the same industry, or for multiple companies from different industries.

Moreover, in order to get more fundamental results of the focus on statistics, more data is needed. For this research we only had the first six months of 2019 to evaluate the forecasts using RStudio. To get more complete results, testing it for at least six other months is desirable. Besides, it would be interesting to experiment with the selection of statistical methods that can be used for making the forecasts. For example, what accuracy could be achieved when only the exponential smoothing methods would be used for making the forecasts of a specific class. Doing this, the extra effort for using multiple and more complicated statistical methods can be compared against the increase in the forecast accuracy.

One of the statistical methods that can be extended, is the multiple regression. We currently only used the regression that includes a trend and seasonality. However, it would be interesting to determine if other factors, like GDP, have an impact on the forecasts. Moreover, further research could examine how to incorporate the qualitative factors explained in this research into the regression models. Gathering the necessary data and determining the impact of the factors may show some patterns, which helps to incorporate the qualitative information more data based.

# References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

Antle, R., & Eppen, G. D. (1985). Capital rationing and organizational slack in capital-budgeting. *Management Science, 31*(2), 163-174.

Armstrong, J. S. (2001). *Principles of Forecasting.* Boston: Kluwer Academic Publisher.

Armstrong, J. S., Green, K. C., & Graefe, A. (2015). Golden rule of forecastng: Be conservative. *Journal of Business Research, 68*, 1717-1731.

Axsäter, S. (2006). *Inventory Control* (2nd ed.). Lund: Springer.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Operational Research Quaterly, 20*(4), 451-468.

Box, G. E., & Jenkins, G. M. (1970). *Time series analysis: forecasting and control.* San Francisco: Holden-Day.

Brown, R. G. (1959). *Statistical forecasting for inventory control.* New York: McGraw-Hill.

Burnham, K. P., & Anderson, D. R. (2004). Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261-304.

Cachon, G. P., & Lariviere, M. A. (2001). Contracting to assure supply: how to share demand forecasts in a supply chain. *Management of Science, 47*(5), 629-646.

Chambers, J. C., Mullick, S., & Smith, D. D. (1971). How to choose the right forecasting technique. *Harvard Business Review, 49*, 45-71.

Chatfield, C. (1978). The Holt-Winters forecasting procedure. *Applied Statistics, 27*(3), 264-279.

Chatfield, C. (1988). Apples, oranges and mean square error. *International Journal of Forecasting, 4*, 515-518.

Clemen, R. (1989). Combining forecasts: A review and annotated bibliography with discussion. *International Journal of Forecasting, 5*, 559-608.

Copeland, T., Koller, T., & Murrin, J. (1994). *Valuation: Measuring and managing the value of companies* (2nd ed.). New York: Wiley.

Costantino, F., Di Gravio, G., Patriarca, R., & Petrella, L. (2017). Spare parts management for irregular demand items. *Omega*, 1-10.

Crittenden, V. L., Gardiner, L. R., & Stam, A. (1993). Reducing conflict between marketing and manufacturing. *Industrial Marketing Management, 22*(4), 299-309.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operations Research Quaterly, 23*(3), 289-303.

DeSmet, B. (2018). *Supply Chain Strategy and Financial Metrics; The Supply Chain Triangle of Service, Cost and Cash.* Kogoan Page Limited.

Duffuaa, S. O., & Raouf, A. (2015). *Planning and Control of Maintenance Systems; Modelling and Analysis* (2nd ed.). Switzerland: Springer.

Ekern, S. (1981). Adaptive exponential smoothing revisited. *Journal of Operational Research Society, 32*(9), 775-782.

Errasti, A., Chackelson, C., & Poler, R. (2010). An expert system for inventory replenishment optimization. *International Federation for Information Processing, 322*, 129-136.

Fildes, R., & Goodwin, P. (2007). Against Better Judgment? How Organizations Can Improve Their Use of Management Judgement in Forecasting. *INFORMS Journal on Applied Analytics, 37*(6), 570-576.

Fildes, R., & Petropoulos, F. (2015). Simple versus complex selection rules for forecasting many time series. *Journal of Business Research, 68*, 1692-1701.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an emperical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting, 25*, 3-23.

Fildes, R., Madden, G., & Tan, J. (2007). Optimal forecasting model selection and data characteristics. *Applied Financial Economics, 17*, 1251-1264.

Fliedner, G. (1999). An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers and Operations Research, 26*, 1133-1149.

Flores, B. E., & Whybark, D. C. (1987). Implementing multiple criteria ABC analysis. *Journal of Operations, 7*(1-2), 79-85.

Flowers, A. D. (1980). A simulation study of smoothing constant limits for an adaptive forecasting system. *Journal of Operations Management, 1*(2), 85-94.

Gardner Jr., E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science, 31*(10), 1237-1246.

Goodwin, P. (2000). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting, 12*, 85-99.

Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research 68*, 1678-1685.

Griffin, A., & Hauser, J. R. (1992). Patterns of communication among marketing, engineering and manufacturing: a comparison between two new product teams. *Management of Scinece, 38*(3), 360-373.

Griffin, A., & Hauser, J. R. (1996). Integrating R&D and Marketing: a review and analysis of the literature. *Journal of Product Innovation Management, 13*(1), 191-215.

Grunfeld, Y., & Griliches, Z. (1960). Is aggregation necessarily bad? *Review of Econimics and Statistics, 42*(1), 1-13.

Gudehus, T. (2006). *Dynamische Disposition – Strategien zur optimalen Auftrags- und Bestandsdisposition.* Berlin: Springer.

Guvenir, H. A., & Erel, E. (1998). Multicriteria inventory classification using a genetic algorithm. *European Journal of Operational Research, 105*, 29-37.

Hamel, G., & Prahalad, C. K. (1989). Strategic intent. *Harvard Business Review, 67*(3), 63-78.

Harvey, N. (2007). Use of heuristics: Insights from forecasting research. *Thinking and Reasoning, 13*(1), 5-24.

Holt, C. C. (1957). Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. *ONR Memorandum, 52*.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting, 20*, 5-10.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). Melbourne, Australia: OTexts.

Hyndman, R. J., & Billah, B. (2003). Unmasking the Theta method. *International Journal of Forecasting, 19*(2), 287-290.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*, 679-688.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing; The State Space Approach.* Berlin: Springer.

Kang, H. (1986). Unstable weights in the combination of forecasts. *Management Science, 32*(6), 683-695.

Keating, E. K., Oliva, R., Repenning, N., Rockart, S. F., & Sterman, J. D. (1999). Overcoming the improvement paradox. *European Management Journal, 17*(2), 120-134.

Kepczynski, R., Jandhyala, R., Sankaran, G., & Dimofte, A. (2018). *Integrated Business Planning: How to integrate planning processes, organizational structures and capabilities, and leverage SAP IBP technology.* Cham: Spinger.

Koch, R. (1997). *The 80/20 principle: The secret of achieving more with less.* New York: Doubleday.

Kremer, M., Moritz, B., & Siemsen, W. (2011). Demand forecasting behavior: system neglect and change detetection. *Management Science, 57*, 1827-1843.

Krupp, J. A. (1994). Are ABC codes an obsolete technology? *APICS - The Performance Advtange*, 34-35.

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research, 33*(2), 18-229.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). esting the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54*(1-3), 159-178.

Lapide, L. (2005). An S&OP maturity model. *Journal of Business Forecasts, 24*(3), 15-20.

Lawrence, M. J., Edmundson, R. H., & O'Connor, M. J. (1986). The accuracy of combining judgmental and statistical forecasts. *Management Science, 32*(12), 1521-1532.

Lawrence, M., Edmundson, R. H., & O'Connor, M. J. (1985). An examination of the accuracy of judgmental extrapolation of time series. *International Journal of Forecasting, 1*(1), 25-35.

Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting, 22*(3), 493-518.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: the bullwhip effect. *Management of Science, 43*, 546-558.

Makridakis, S., & Hibon, M. (2000). The M3-competition: results, conclusions and implication. *International Journal of Forecasting, 9*, 451-476.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., . . . Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting, 1*(2), 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simmons, L. F. (1993). The M-2 competition: a real-time judgmentally based forecasting study. *International Journal of Forecasting, 9*, 5-23.

Markin, S., & Sinha, A. (2018). *SAP Integrated Business Planning.* Boston: Rheinwerk Publishing.

Marmier, F., & Cheikhrouhou, N. (2010). Structuring and integrating human knowledge in demand forecasting: a judgmental adjustment approach. *Production Planning and Control, 21*(4), 399-412.

McCarthy, T. M., Davis, D. F., Golicic, S. L., & Mentzer, J. T. (2006). The Evolution of Sals Forecasting Management: A 20-year Longitudinal study of forecasting practices. *Journal of Forecasting, 25*, 303-324.

Oliva, R. (2001). Tradeoffs in responses to work pressure in the service industry. *California Management Review, 43*(4), 26-43.

Oliva, R., & Sterman, J. D. (2001). Cutting corners and working overtime: quality erosion in the service industry. *Management of Science, 47*(7), 894-914.

Oliva, R., & Watson, N. (2009). Managing Functional Biases in Organizational Forecasts: A Case Study of Consensus Forecasting in Supply Chain Planning. *Production and Operations Management, 18*(2), 138-151.

Onwubolu, G. C., & Dube, B. C. (2006). Implementing an improved inventory control system in a small company: a case study. *Production Planning & Control, 17*(1), 67-76.

Pandaya, B., & Thakkar, H. (2016). A review on inventory management control techniques: ABC-XYZ analysis. *REST Journal on Emerging trends in Modelling and Manufacturing, 2*(3), 82-86.

Pareto, V. (1896). Cours d'economie politique. *University of Lausanne, 1*, 430.

Petropoulos, F., Kourentzes, N., Nikolopoulos, K., & Siemsen, E. (2018). Judgmental Selection of Forecasting Models. *Journal of Operations Management, 60*, 34-46.

Rasmussen, R. (2004). On time series data and optimal parameters. *Omega, 32*(2), 111-120.

Ross, A. (2004). Why bad forecasts can be good for business. *Manufacturing Engineer, 83*(3), 26-29.

Rowe, G. (2007). A guide to Delphi. *Foresight: The International Journal of Applied Forecasting*(8), 11-16.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting, 15*(4), 353-375.

Sankaran, G., Sasso, F., Kepczynski, R., & Chiaraviglio, A. (2019). *Improving forecasts with Integreated Business Planning: From short-term to long-term demand planning enabled by SAP IBP.* Zwitserland: Springer.

Scholz-Reiter, B., Heger, J., Meinecke, C., & Bergmann, J. (2012). Integration of demand forecasting in ABC-XYZ analysis: practical investigation at an industrial company. *International Journal of Productivity and Performance Management, 61*(4), 445-451.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464.

Shapiro, B. P. (1977). Can Marketing and Manufacturing Coexist? *Harvard Business Review, 55*(5), 104-114.

Shapiro, F. R. (2006). *The Yale book of quotations.* New Haven: Yale University Press.

Silver, E. A., Pyke, D. F., & Thomas, D. J. (2017). *Inventory and Production Management in Supply Chains* (4th ed.). Boca Raton: Taylor & Francis Group.

Stein, J. C. (1997). Internal capital markets and the competition for corporate resources. *Journal of Finance, 52*(1), 111-133.

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of Operational Research Society, 56*(5), 495-503.

Syntetos, A. A., Boylan, J. E., & Disney, S. (2009). Forecasting for inventory planning: a 50-year review. *Journal of the Operational Research Society, 60*(1), 149-160.

Syntetos, A. A., Nikolopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting, 26*, 134-143.

Terwiesch, C., Ren, Z. J., Ho, T. H., & Cohen, M. A. (2005). An emprical analysis of forecast sharing in the semiconductor equipment supply chain. *Management of Science, 51*(2), 208-220.

Vanston, J. H. (2003). Better forecasts, better plans, better results. *Research Technology Management, 46*(1), 47-58.

Wallace, T. F. (2004). *Sales & Operations Planning: The how-to handbook* (2nd ed.). Cincinnati: T.F. Wallace & Company.

Wavin. (2019, April). *History*. Retrieved from https://www.wavin.com/en-EN/About/Organisation/History

Winkler, R. L., & Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society, 146*(2), 150-157.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science, 6*, 324-342.

# Appendix A: Countries per Demand Manager

*Table 37: Countries per Demand Manager*

| Demand Manager | Countries |
|---|---|
| 1 | Estonia<br>Latvia<br>Lithuania<br>Poland |
| 2 | Czech Republic<br>Hungary<br>Slovakia |
| 3 | Germany<br>Switzerland<br>Austria<br>Italy |
| 4 | Denmark<br>Finland<br>Norway<br>Sweden |
| 5 | Belgium<br>Netherlands<br>France |
| 6 | Turkey |
| 7 | United Kingdom<br>Ireland |

# Appendix B: Available statistical methods in SAP IBP

*Table 38: Statistical forecasting methods in SAP IBP*

| Method | Forecasting Effort | Stable Demand | Intermittent Demand | Trend | Season | Cycle |
|---|---|---|---|---|---|---|
| Simple Average | Low | Yes | No | No | No | No |
| Simple Moving Average | Low | Yes | No | Yes | No | No |
| Weighted Average | Low | Yes | No | No | No | No |
| Weighted Moving Average | Low | Yes | No | Yes | No | No |
| Naïve | Low | Yes | No | No | No | No |
| Seasonal naïve | Low | No | No | No | Yes | No |
| Single Exponential Smoothing (SES) | Low | Yes | No | No | No | Yes |
| Adaptive-response-rate SES | Low | No | No | Yes | No | Yes |
| Double Exponential Smoothing | Medium | No | No | Yes | No | No |
| Triple Exponential Smoothing | Medium | No | No | No | Yes | No |
| Automated Exponential Smoothing | Medium | No | No | No | No | No |
| Croston Method | Medium | No | Yes | No | No | No |
| Multiple Linear Regression | Medium | No | No | No | No | Yes |
| Brown's Linear Exponential Smoothing | Medium | Yes | No | Yes | No | No |
| ARIMA | High | Yes | No | Yes | No | Yes |
| SARIMA | High | Yes | No | Yes | Yes | Yes |

# Appendix C: Checklist for adjusting statistical forecasts

| Market Event checklist | Owner | Possible impact |
|---|---|---|
| ☐ Sales price management | Sales | Dependent on price elasticity could lead to short term stocking if lower price vs postone if high price |
| ☐ Promotions | Sales & Marketing | Volume increase, possible short term price decrease |
| ☐ Marketing campaign (events) | Marketing | Focused target audience and campaign objectives |
| ☐ Conditional bonus performance | Sales | Specific short term incentives for customer (product mix / growth / volume). Depending on performance, trend and behaviour of the customer could lead to volume increase. |
| ☐ Product launch | Product Management | Speed-up or postpone launch of products, needs continuous alignment with supply chain on availability and timing |
| ☐ Competitive information | Sales | Competitors price development, major projects/ tenders lost/won, organisational developments etc could lead to both short and long term market share development (positive and negative) |

*Figure 31: Checklist for adding market events*

| Sourcing Event checklist | Owner | Possible impact |
|---|---|---|
| ☐ (Raw) material availability | Procurement | Sudden availability loss, without alternative sources, leads to scarcity. |
| ☐ Overall Raw material price development | Procurement | Overall RM price development is input for the sales price development. Significant price development could lead to disrupted sales price management (both positive and negative) |
| ☐ Phase-in / Phase-out | Product management | Replacement period, phase out |
| ☐ Procurement bonus | Procurement | Specific short term incentives for Wavin can lead to periodical higher or lower cost prices which can give a limited opportunity in the market. |
| ☐ Third party supplier | Supply Chain | Periodical out of stock |
| ☐ Intercompany supplier | Supply Chain | Strong intercompany alignement on forecast is crucial, golden rules of intercompany. |

*Figure 32: Checklist for adding sourcing events*

# Appendix D: Forecast accuracies per country over 2018

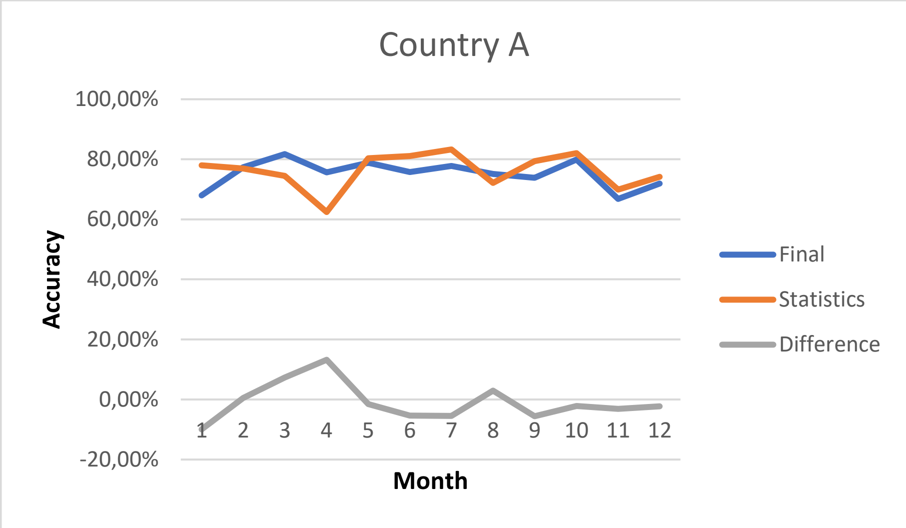



*Figure 33: Forecast accuracies per month over 2018 Country A*

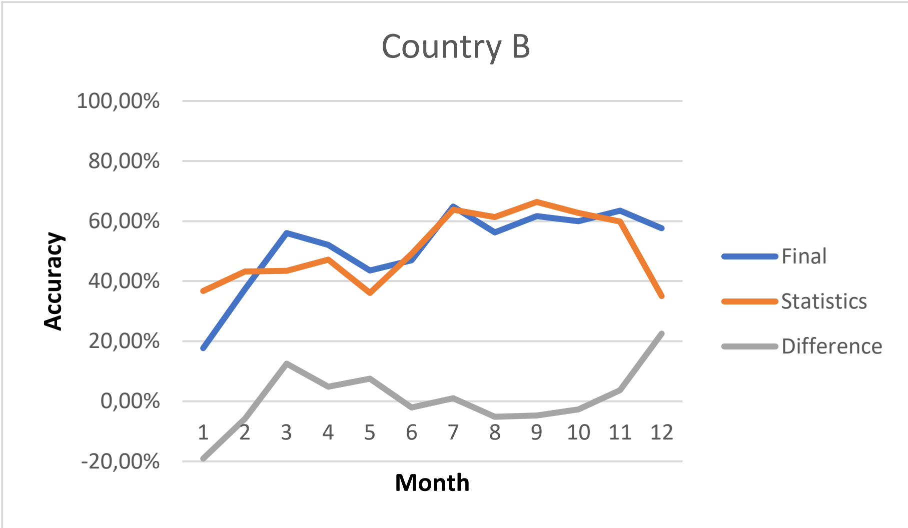



*Figure 34: Forecast accuracies per month over 2018 Country B*

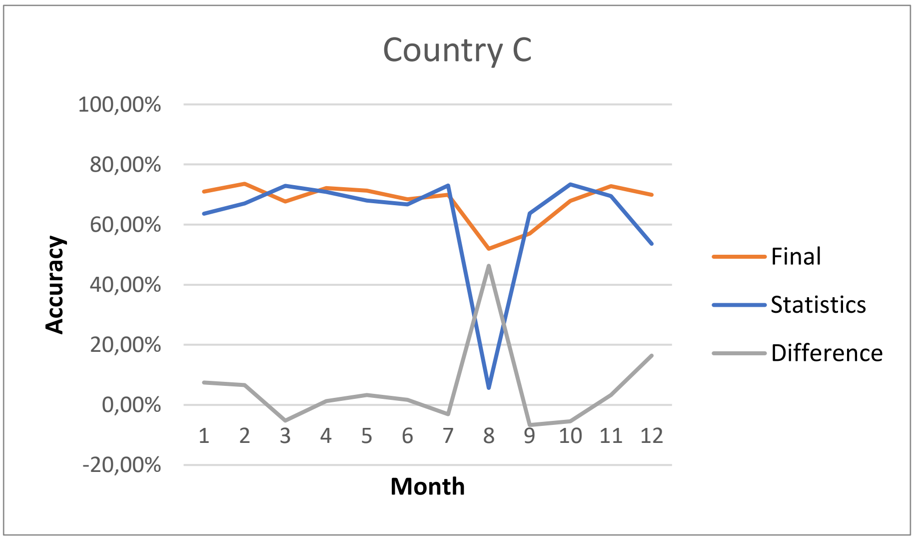



*Figure 35: Forecast accuracies per month over 2018 Country C*

# Appendix E: Golden Rule Checklist

| | | |
|---|---|---|
| **1.** | | **Problem formulation** |
| 1.1 | | Use all important knowledge and information by… |
| 1.1.1 | ☐ | selecting evidence-based methods validated for the situation |
| 1.1.2 | ☐ | decomposing to best use knowledge, information, judgment |
| 1.2 | | Avoid bias by… |
| 1.2.1 | ☐ | concealing the purpose of the forecast |
| 1.2.2 | ☐ | specifying multiple hypotheses and methods |
| 1.2.3 | ☐ | obtaining signed ethics statements before and after forecasting |
| 1.3 | ☐ | Provide full disclosure for independent audits, replications, extensions |
| **2.** | | **Judgmental methods** |
| 2.1 | ☐ | Avoid unaided judgment |
| 2.2 | ☐ | Use alternative wording and pretest questions |
| 2.3 | ☐ | Ask judges to write reasons against the forecasts |
| 2.4 | ☐ | Use judgmental bootstrapping |
| 2.5 | ☐ | Use structured analogies |
| 2.6 | ☐ | Combine independent forecasts from judges |
| **3.** | | **Extrapolation methods** |
| 3.1 | ☐ | Use the longest time-series of valid and relevant data |
| 3.2 | ☐ | Decompose by causal forces |
| 3.3 | | Modify trends to incorporate more knowledge if the… |
| 3.3.1 | ☐ | series is variable or unstable |
| 3.3.2 | ☐ | historical trend conflicts with causal forces |
| 3.3.3 | ☐ | forecast horizon is longer than the historical series |
| 3.3.4 | ☐ | short and long-term trend directions are inconsistent |
| 3.4 | | Modify seasonal factors to reflect uncertainty if… |
| 3.4.1 | ☐ | estimates vary substantially across years |
| 3.4.2 | ☐ | few years of data are available |
| 3.4.3 | ☐ | causal knowledge is weak |
| 3.5 | ☐ | Combine forecasts from alternative extrapolation methods, data |
| **4.** | | **Causal methods** |
| 4.1 | ☐ | Use prior knowledge to specify variables, relationships, and effects |
| 4.2 | ☐ | Modify effect estimates to reflect uncertainty |
| 4.3 | ☐ | Use all important variables |
| 4.4 | ☐ | Combine forecasts from dissimilar models |
| **5.** | ☐ | **Combine forecasts from diverse evidence-based methods** |
| **6.** | ☐ | **Avoid unstructured judgmental adjustments to forecasts** |

*Figure 36: Golden Rule Checklist (Armstrong et al., 2015)*

# Appendix F: Grouped products per minimum number of months

*Table 39: Grouped per class X & Y based on minimum amount of months Country A*

| Min month | X | Y |
|---|---|---|
| 6/12 | 41 | 31 |
| 7/12 | 41 | 28 |
| 8/12 | 34 | 31 |
| 9/12 | 27 | 32 |
| 10/12 | 16 | 37 |
| 11/12 | 5 | 41 |
| 12/12 | 0 | 28 |

*Table 40: Grouped per class X & Y based on minimum amount of months Country B*

| Min month | X | Y |
|---|---|---|
| 6/12 | 13 | 8 |
| 7/12 | 12 | 8 |
| 8/12 | 11 | 7 |
| 9/12 | 7 | 9 |
| 10/12 | 4 | 10 |
| 11/12 | 3 | 10 |
| 12/12 | 1 | 12 |

*Table 41: Grouped per class X & Y based on minimum amount of months Country C*

| Min month | X | Y |
|---|---|---|
| 6/12 | 13 | 19 |
| 7/12 | 12 | 19 |
| 8/12 | 10 | 16 |
| 9/12 | 7 | 16 |
| 10/12 | 3 | 16 |
| 11/12 | 3 | 14 |
| 12/12 | 1 | 14 |

# Appendix G: Different classes deseasonalized CoV with wMAPE

*Table 42: Analysis of different given XYZ class CoV and wMAPE Country A*

| Group | CoV | Class | Accuracy | Class | CoV difference from TH | Accuracy difference from TH | Months needed to change for other class | Variable close to TH |
|---|---|---|---|---|---|---|---|---|
| 2 | 0,145 | X | 79,6% | Y | 0,01 | 0,4% | 2 | Both |
| 4 | 0,094 | X | 85,4% | Y | 0,07 | 0,0% | 2 | Accuracy |
| 6 | 0,142 | X | 78,6% | Y | 0,02 | 1,4% | 2 | Both |
| 7 | 0,124 | X | 82,5% | Y | 0,04 | 0,0% | 2 | Both |
| 36 | 0,144 | X | 85,2% | Y | 0,02 | 0,0% | 1 | Both |
| 39 | 0,118 | X | 75,2% | Y | 0,04 | 4,8% | 4 | CoV |
| 80 | 0,137 | X | 85,2% | Y | 0,02 | 0,0% | 1 | Both |
| 101 | 0,329 | Z | 66,5% | Y | 0,10 | 1,5% | 1 | Accuracy |
| 102 | 0,161 | Y | 0,0% | Z | 0,00 | 65,0% | 4 | - |
| 103 | 0,243 | Z | 79,0% | Y | 0,01 | 14,0% | 2 | Both |
| 124 | 0,181 | Y | 81,8% | X | 0,02 | 1,8% | 1 | Both |
| 144 | 0,373 | Z | 74,9% | Y | 0,14 | 9,9% | 2 | Accuracy |
| 148 | 0,172 | Y | 17,9% | Z | 0,06 | 47,1% | 3 | - |
| 150 | 0,152 | X | 35,1% | Z | 0,08 | 29,9% | 7 | - |
| 176 | 0,153 | X | 80,0% | Y | 0,01 | 0,0% | 3 | CoV |
| 198 | 0,168 | Y | 51,5% | Z | 0,06 | 13,5% | 6 | - |
| 203 | 0,234 | Z | 79,8% | Y | 0,00 | 0,2% | 3 | CoV |
| 204 | 0,161 | Y | 82,4% | X | 0,00 | 2,4% | 1 | Both |
| 207 | 0,273 | Z | 71,2% | Y | 0,04 | 6,2% | 1 | Both |
| 210 | 0,131 | X | 82,3% | Y | 0,03 | 0,0% | 2 | Both |
| 211 | 0,146 | X | 82,9% | Y | 0,01 | 0,0% | 1 | Both |
| 230 | 0,229 | Y | 25,4% | Z | 0,00 | 39,6% | 3 | CoV |
| 233 | 0,147 | X | 87,2% | Y | 0,01 | 0,0% | 1 | Both |
| 244 | 0,288 | Z | 70,4% | Y | 0,06 | 5,4% | 2 | Accuracy |

*Table 43: Analysis of different given XYZ class CoV and wMAPE Country B*

| Group | CoV | Class | Accuracy | Class | CoV difference from TH | Accuracy difference from TH | Months needed to change for other class | Variable close to TH |
|---|---|---|---|---|---|---|---|---|
| 2 | 0,245 | X | 70,2% | Y | 0,07 | 5,2% | 1 | Accuracy |
| 77 | 0,504 | Z | 54,3% | Y | 0,02 | 4,3% | 2 | Both |
| 80 | 0,258 | X | 71,0% | Y | 0,05 | 0,0% | 1 | Both |

*Table 44: Analysis of different given XYZ class CoV and wMAPE Country C*

| Group | CoV | Class | Accuracy | Class | CoV difference from TH | Accuracy difference from TH | Months needed to change for other class | Variable close to TH |
|---|---|---|---|---|---|---|---|---|
| 36 | 0,293 | Y | 78,3% | X | 0,06 | 13,3% | 3 | - |
| 39 | 0,456 | Z | 60,6% | Y | 0,03 | 4,4% | 1 | Both |
| 44 | 0,333 | Y | 68,9% | X | 0,10 | 3,9% | 1 | Accuracy |
| 45 | 0,277 | Y | 69,2% | X | 0,05 | 4,2% | 1 | Both |
| 51 | 0,308 | Y | 56,2% | Z | 0,08 | 6,2% | 2 | Accuracy |
| 60 | 0,450 | Z | 60,9% | Y | 0,02 | 4,1% | 2 | Both |
| 96 | 0,346 | Y | 57,4% | Z | 0,08 | 7,4% | 1 | Accuracy |
| 99 | 0,339 | Y | 60,4% | Z | 0,09 | 4,6% | 2 | Accuracy |
| 131 | 0,295 | Y | 76,3% | X | 0,06 | 11,3% | 1 | Accuracy |

# Appendix H: User guide Excel tool Differentiated Forecasting

## 1. Introduction

This user guide explains how to deal with the developed Excel tool for making the ABC-XYZ classification for differentiated forecasting. We recommend reading this document carefully before making use of the Excel tool. The classification should be made twice a year, since values may change resulting in different classes. This user guide is made by grouping on level 7/8/9 and using Country B as an example. In Figure 37 the "UserGuide" sheet is shown which is the first sheet of the Excel file and which summarizes what each sheet contains.

| Legenda |
|---|
| Input |
| Run model |
| Output |
| Extra analysis |

| Sheet | Including/Action |
|---|---|
| 1-12 | In these sheets the monthly data (forecasts/accuracies) of the last 12 months needs to be copied |
| Revenues | In this sheet the Revenues needs to be copied |
| Settings | In this sheet the macro can be run after importing the data, changing the thresholds and selecting the hierarchies |
| XYZ | In this sheet the XYZ classification is made. It is possible to change classes manually, but then use the "Class 2" button in Settings |
| ABC | In this sheet the ABC classification is made. It is possible to change classes manually, but then use the "Class 2" button in Settings |
| Classification | The classification and results with diagrams are shown |
| SFA | This is the same as the cassification sheet but now on SFA |
| Output | This is the output sheet with the classes that needs to be copied to the monthly evaluation sheet |
| ABCSales | Gives the ABC sales classification on SKU level based on revenues and orderliness |
| Analysis | Some extra analyisis of the overall performance of the 12 months |
| Sales | Gives the sales per group per month |
| F.Accuracy | Gives the Final Forecast accuracy per month |
| S.Accuracy | Gives the Statistical Forecast accuracy per month |
| F-S | It gives the difference between Final and Statistics on monthly basis |
| FinalSFA | Final Forecast accuracy on group level |
| StatSFA | Statistical Forecast accuracy on group level |

*Figure 37: User Guide sheet*

## 2. Interface

In Figure 38 the first sheet of the Excel file is shown. With this sheet the classification can be made automatically with one push of a button after the data is loaded into the file (see Section 3). All the green cells can be changed, but only do when necessary.

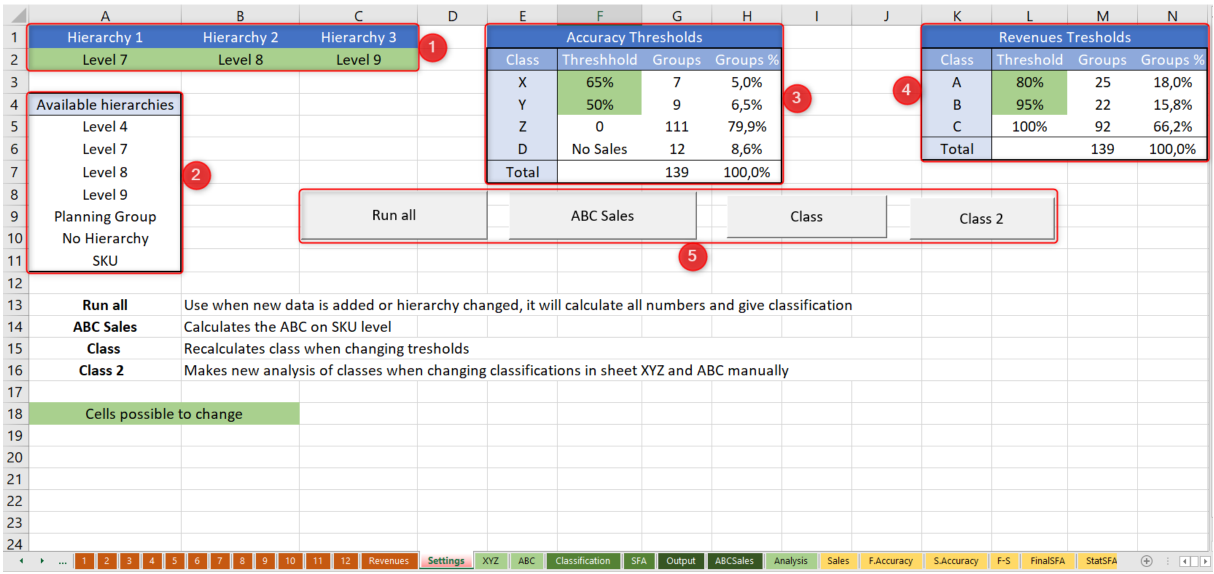



*Figure 38: Sheet Settings*

1. Hierarchies and groupings can be set using a drop-down box.
    a. Up to three hierarchies at the same time can be chosen.
    b. When selecting “Planning Group” or “SKU”, the other hierarchies should be set to “No Hierarchy”.
2. All the available hierarchies that can be chosen for making the classification are shown.
3. This gives the overview of the XYZ classification.
    a. The Thresholds can be changed, but only do when discussed with others.
    b. Class D products are the products with no sales.
4. This gives an overview of the ABC classification
    a. The thresholds can be changed, but only do when discussed with others.
5. These are the buttons to run the macro for making the classification.
    a. First make sure that the data (monthly sales, forecast accuracies, revenues etc.) is loaded before running the macro.
    b. When data is loaded, the “Run all” button can be used. Dependent on the country and the type of hierarchy chosen it may take some seconds up to several minutes to run the macro. Be aware that other Microsoft programs may not work properly during the macro is running. It may happen that it is shown that the excel file is “not responding”. This is not the case; it is just running the macro and you need to wait a while.
    c. Run ABC Sales, which is the ABC classification on SKU level.
    d. When for some reasons the thresholds for the classification are changed (point 3 and 4), after the “Run all” button is already used, the “Class” button can be used to make a new classification.
    e. Sometimes it may be the case that it is desirable to change the class of some products by hand. This can be done in sheet “XYZ” and ABC”. When this is changed, the analysis of the classification does not change automatically. The button “Class 2” can be used to change the classification without overwriting it by using the thresholds.

## 3. Load Data

The first step is to load the data into the tool such that the classification can be made. For making the classification, the information on material level is needed. The actual sales, the final and forecasted sales and corresponding accuracies, and the revenues are needed for making the classification. In Section 3.1 and 3.2, we explain how this data exactly should look like. The data can be exported from SAP.

### 3.1 Forecast accuracies and sales

In Figure 39, one of the sheets of 1-12 is shown. In these sheets the data should be loaded in order to make the classification. The data of the last 12 months should be exported from SAP and copied in these sheets.

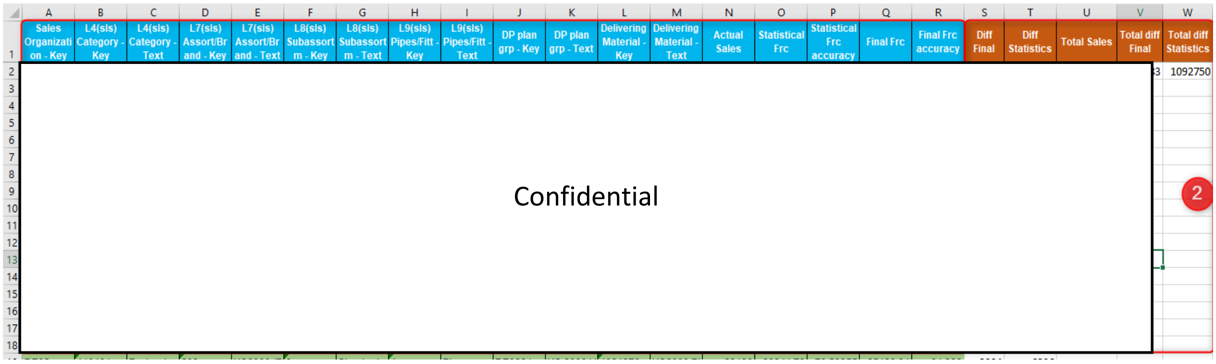



*Figure 39: Sheet 1-12*

1. In these columns the data needs to be copied. Make sure to only copy in the green marked cells. It is very important that the same layout is used when loading the data. If some columns are missing or changed the tool will not work! So compare the output whether the right information is copied into the right column.
2. These columns are just auxiliary columns for running the macro faster. These columns should be left alone and not be changed.

### 3.2 Revenues

In the sheet "Revenues" shown in Figure 40, the data of the Revenues should be loaded.

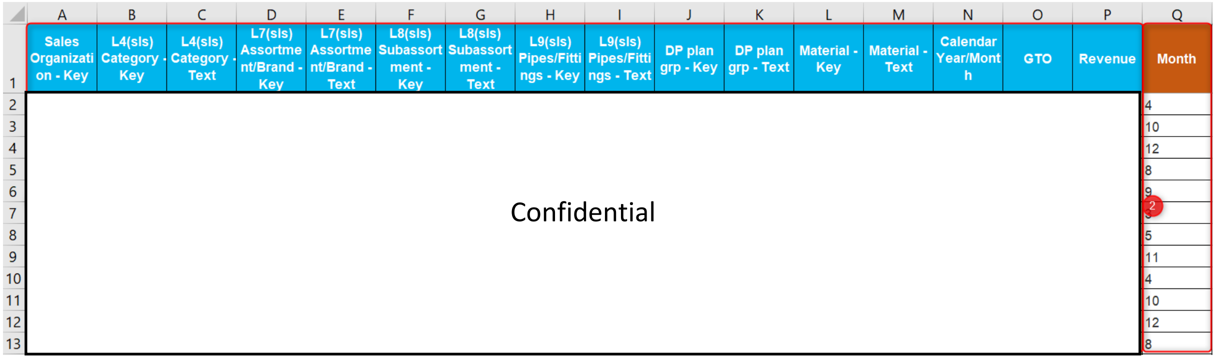



*Figure 40: Sheet Results*

1. The same layout of data as in Figure 40 needs to be imported to the excel file in the green marked cells.
2. It is just an auxiliary column making the macro faster.

## 4. Extra analysis sheets

The sheets “Sales”, “F.Accuracy” (final forecast accuracy, SKU level), “S.Accuracy” (statistical forecast accuracy, SKU level), “F-S” (difference between final and statistical forecast accuracy), “FinalSFA” (Final accuracy on aggregated level), “StatSFA” (statistical accuracy on aggregated level)  look very similar, of which the example of “F.Accuracy” is shown in Figure 41.

| | Hierarchy 1 | Hierarchy 2 | Hierarchy 3 | Month | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No | Level 7 | Level 8 | Level 9 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Average | W.Average | St.dev |
| 1 | 208 | Confidential | | 49,873 | 67,711 | 92,858 | 90,775 | 82,718 | 76,788 | 83,636 | 88,282 | 49,128 | 66,474 | 77,61 | 85,904 | 75,97967 | 74,7942113 | 14,20744 |
| 2 | 208 | | | 43,966 | 53,17 | 78,842 | 82,163 | 71,084 | 53,66 | 88,079 | 85,457 | 60,302 | 56,908 | 76,866 | 83,33 | 69,48572 | 67,4824851 | 14,45729 |
| 3 | 229 | | | 64,594 | 56,419 | 37,218 | 59,563 | 52,426 | 53,099 | 68,378 | 56,743 | 48,487 | 42,291 | 38,182 | 71,325 | 54,06043 | 54,0244832 | 10,66936 |
| 4 | 414 | | | 29,063 | 49,487 | 51,453 | 64,575 | 58,472 | 47,469 | 22,838 | 20,811 | 22,303 | 51,276 | 51,26 | 50,023 | 43,25244 | 44,6200003 | 14,56329 |
| 5 | 41 | | | 51,107 | 50,039 | 49,172 | 65,539 | 54,764 | 61,097 | 62,527 | 57,228 | 59,3 | 75,068 | 61,629 | 50,316 | 58,14899 | 60,0063738 | 7,369778 |
| 6 | 54 | | | 50,784 | 49,776 | 65,969 | 71,921 | 69,925 | 33,93 | 84,33 | 75,446 | 66,204 | 64,563 | 73,403 | 62,73 | 64,08178 | 65,2003215 | 12,98202 |
| 7 | 54 | | | 46,066 | 45,291 | 50,575 | 60,431 | 54,975 | 45,87 | 52,629 | 45,166 | 17,725 | 30,083 | 26,269 | 49,497 | 43,71479 | 44,8528032 | 12,06423 |
| 8 | 6 | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 6 | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 7 | | | 63,687 | 58,981 | 80,695 | 86,403 | 92,996 | 77,895 | 79,077 | 88,619 | 74,523 | 72,207 | 63,627 | 59,996 | 74,89209 | 75,3049684 | 10,96861 |

*Figure 41: Sheet F.Accuracy*

1. For easiness, the exceltool assigns numbers to the groups.
2. The values of the hierarchies are shown.
3. The monthly values per group are shown. For the forecast accuracy, the SKU forecast accuracies are weighted with the sales, because WMAPE is used for calculating the accuracies.
4. The averages and standard deviations are shown. “W.Average” is the average of each month weighted to the sales. These values are used for making and analyzing the classification. When the standard deviation is very high, it means the forecast accuracies (or sales) fluctuates much.

## 5. XYZ classification

The XYZ classification is made in sheet “XYZ” shown in Figure 42. When needed, the classes can be changed manually. Be aware that when this is done, the button “Class 2” needs to be used in the “Settings” sheet for the right output.

| | Hierarchy 1 | Hierarchy 2 | Hierarchy 3 | | Months higher than TH | | |
|---|---|---|---|---|---|---|---|
| No | Level 7 | Level 8 | Level 9 | Accuracy | | Sales | Class |
| 109 | Confidential | | | 79,06143 | 12 | Confidential | X |
| 10 | | | | 76,44 | 8 | | Y |
| 110 | | | | 75,1616 | 10 | | X |
| 1 | | | | 74,20713 | 6 | | Y |
| 117 | | | | 73,16511 | 10 | | X |
| 113 | | | | 72,82701 | 10 | | X |
| 11 | | | | 72,5109 | 7 | | Y |
| 106 | | | | 72,07949 | 7 | | Y |
| 102 | | | | 71,47411 | 7 | | Y |
| 2 | | | | 70,64212 | 7 | | Y |
| 105 | | | | 69,99593 | 12 | | Y |

| Class | #Accuracy |
|---|---|
| X | 4 |
| Y | 16 |
| Z | 107 |
| D* | 12 |
| Total | 139 |

*No Sales

| Treshold | Accuracy |
|---|---|
| X | 70 |
| Y | 50 |
| Z | 0 |

| Accuracy | # | Cum |
|---|---|---|
| 10 | 50 | 50 |
| 20 | 13 | 63 |
| 30 | 10 | 73 |
| 40 | 15 | 88 |
| 50 | 20 | 108 |
| 60 | 12 | 120 |
| 70 | 9 | 129 |
| 80 | 10 | 139 |
| 90 | 0 | 139 |
| 100 | 0 | 139 |

*Figure 42: Sheet "XYZ"*

1. The groups are sorted based on the statistical forecast accuracy on descending order.
2. Not only the statistical forecast accuracy is a criterium for making the classification, but it also needs to have a higher forecast accuracy than the threshold for at least 9 out of 12 months.
3. The classes are assigned based on the statistical forecast accuracy thresholds for 9 out of 12 months.
4. Summarizes the number of groups per class.
5. The Thresholds chosen in the “Settings” sheet.
6. Summarizes the number of groups that have a certain forecast accuracy with steps of 10%.

## 6. ABC classification

The ABC classification is made in sheet "ABC" shown in Figure 43. When needed, the classes can be changed manually. Be aware that when this is done, the button "Class 2" needs to be used in the "Settings" sheet for the right output. The sheet "ABCSales" has a similar analysis but then on SKU level.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | Hierarchy 1 | Hierarchy 2 | Hierarchy 3 | | | | | |
| 2 | Number | Level 7 | Level 8 | Level 9 | Revenues | % of total rev | Cumulative | % part | Class |
| 3 | 1 | Confidential | | | | 8,32% | 8,32% | 1% | A |
| 4 | 113 | | | | | 7,47% | 15,79% | 1% | A |
| 5 | 58 | | | | | 6,93% | 22,71% | 2% | A |
| 6 | 2 | | | | | 5,89% | 28,60% | 3% | A |
| 7 | 91 | | | | | 5,72% | 34,33% | 4% | A |
| 8 | 84 | | | | | 4,23% | 38,55% | 4% | A |
| 9 | 57 | | | | | 4,09% | 42,64% | 5% | A |
| 10 | 10 | | | | | 3,97% | 46,61% | 6% | A |
| 11 | 108 | | | | | [illegible] | 49,83% | 6% | A |
| 12 | 109 | | | | | 3,01% | 52,84% | 7% | A |
| 13 | 30 | | | | | 2,70% | 55,54% | 8% | A |
| 14 | 4 | | | | | 2,67% | 58,22% | 9% | A |
| 15 | 45 | | | | | 2,66% | 60,88% | 9% | A |

| K | L | M | N | O | P |
|---|---|---|---|---|---|
| Class | # | Lower | Upper | Amount | # products |
| A | 25 | 0,0% | 80,0% | 80,0% | 18,0% |
| B | 22 | 80,0% | 95,0% | 15,0% | 15,8% |
| C | 92 | 95,0% | 100,0% | 5,0% | 66,2% |
| Total | 139 | | | | 100,0% |

*Figure 43: Sheet ABC*

1. The revenues per group, the percentage of the total revenues, the cumulative revenues and the cumulative of the number of groups are given. The data is sorted on descending order based on the revenues.
2. The class is given. For the current thresholds, the first products accounting until 80% of the total revenues (column G) are assigned to class A. The next 15% to class B and the remaining 5% to class C.
3. Summary of the thresholds and  the number of groups there are in every group.

## 7. The classifications

The sheet "Classification" contains the analysis of the classification on SKU level with multiple figures. The same analysis is done on aggregated level in sheet "SFA". In Figure 44 and Figure 45 the sheet for "Classification is shown.

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | Hierarchy 1 | Hierarchy 2 | Hierarchy 3 | | | | | | | | | |
| 2 | No | Level 7 | Level 8 | Level 9 | ABC | XYZ | Class | Sales | Revenues | F.Accuracy | Diff Final | S.Accuracy | Diff Stat |
| 3 | 1 | Confidential | | | A | X | AX | Confidential | | 74,70% | 483394 | 74,18% | 493336 |
| 4 | 2 | | | | A | Y | AY | | | 66,96% | 1053430 | 70,18% | 950705 |
| 5 | 3 | | | | B | Z | BZ | | | 20,96% | 28259 | 24,03% | 27161 |
| 6 | 4 | | | | A | Z | AZ | | | 17,60% | 54831 | 15,98% | 55913 |
| 7 | 5 | | | | B | Z | BZ | | | 43,93% | 43318 | 51,18% | 37712 |
| 8 | 6 | | | | B | Z | BZ | | | 60,90% | 15811 | 50,52% | 20010 |
| 9 | 7 | | | | B | Z | BZ | | | 20,39% | 38933 | 23,07% | 37624 |
| 10 | 8 | | | | C | Z | CZ | | | 0,00% | 2533 | 0,00% | 2533 |
| 11 | 9 | | | | A | X | AX | | | 73,58% | 20537 | 74,94% | 19481 |
| 12 | 10 | | | | A | X | AX | | | 66,36% | 19610 | 70,19% | 17380 |
| 13 | 11 | | | | B | Z | BZ | | | 15,81% | 7135 | 21,80% | 6627 |
| 14 | 12 | | | | C | Z | CZ | | | 18,20% | [illegible] | 26,65% | 13423 |
| 15 | 13 | | | | A | Z | AZ | | | 4,05% | [illegible] | 12,12% | 14684 |
| 16 | 14 | | | | A | Z | AZ | | | 0,00% | 5784 | 0,00% | 5005 |
| 17 | 15 | | | | B | Z | BZ | | | 0,00% | 9916 | 0,00% | 7576 |
| 18 | 16 | | | | C | Z | CZ | | | 15,31% | 37816 | 27,80% | 32235 |
| 19 | 17 | | | | B | Z | BZ | | | 0,00% | 45663 | 0,00% | 42598 |

| O | P | Q | R | S | T | U | V | W |
|---|---|---|---|---|---|---|---|---|
| Class | # | Sales | Revenues | F.Accuracy | F.W.Accuracy | S.Accuracy | S.W.Accuracy | F-S |
| AX | 4 | Confidential | | 71,79% | 74,2% | 74,6% | 74,6% | -0,4% |
| AY | 6 | | | 68,59% | 67,0% | 68,1% | 68,2% | -1,2% |
| AZ | 15 | | | 26,79% | 23,5% | 25,5% | 21,9% | 1,6% |
| BX | 2 | | | 72,61% | 70,6% | 71,8% | 70,2% | 0,4% |
| BY | 0 | | | N/A | N/A | N/A | N/A | N/A |
| BZ | 20 | | | 20,26% | 11,5% | 18,5% | 7,2% | 4,2% |
| CX | 1 | | | 71,25% | 71,3% | 70,3% | 70,3% | 1,0% |
| CY | 3 | | | 56,77% | 56,1% | 56,5% | 56,9% | -0,8% |
| CZ | 76 | | | 6,37% | 6,9% | 7,4% | 16,6% | -9,7% |
| D | 12 | | | N/A | N/A | N/A | N/A | N/A |
| Total | 139 | | | 49,30% | 49,5% | 49,06% | 49,9% | -0,3% |

*Figure 44: Sheet Classes (1)*


*Figure 45: Sheet Classes (2)*

1. The classes ABC and XYZ are combined (column G).
2. Just some auxiliary columns for calculating the weighted averages of the forecast accuracy per class.
3. Summary of every class and the corresponding (weighted) forecast accuracies.
4. The amount of product groups per class.
5. The statistical forecast accuracies per class.
6. The final forecast accuracies per class.
7. The difference between the statistical and final forecast accuracies.
8. The targets per class. You can change them when needed.
9. The difference between the set target and the final forecast accuracy.

## 8. Overall analysis

In Figure 46 some overall analysis is shown. It summarizes the monthly forecast accuracies together with the sales and revenues. With the diagrams, the overall performance over the last 12 months can be evaluated.

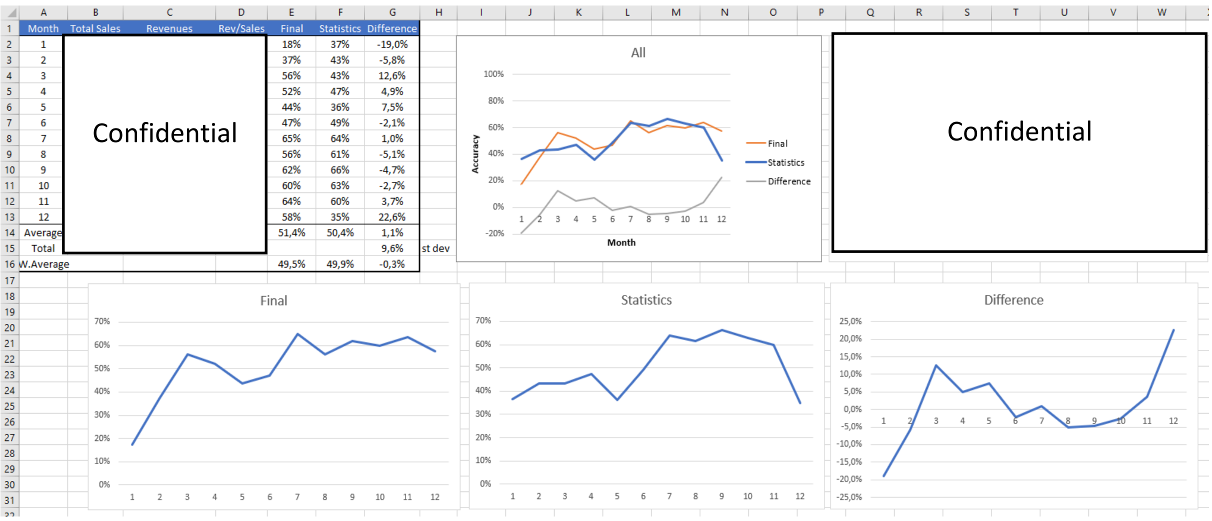
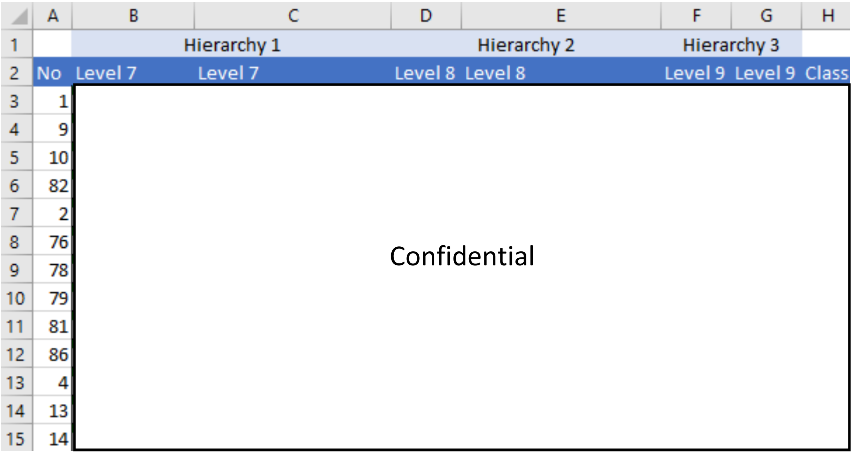



*Figure 46: Sheet Analysis*

## 9. Assigned classes

In Figure 47 the “Output” of the classification is shown. This output with the product groups and the assigned classes can be used for making the forecasts. In column H the classes are shown.



| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | | Hierarchy 1 | | Hierarchy 2 | | Hierarchy 3 | | |
| 2 | No | Level 7 | Level 7 | Level 8 | Level 8 | Level 9 | Level 9 | Class |
| 3 | 1 | | | | | | | |
| 4 | 9 | | | | | | | |
| 5 | 10 | | | | | | | |
| 6 | 82 | | | | | | | |
| 7 | 2 | | | | | | | |
| 8 | 76 | | | | | | | |
| 9 | 78 | | | Confidential | | | | |
| 10 | 79 | | | | | | | |
| 11 | 81 | | | | | | | |
| 12 | 86 | | | | | | | |
| 13 | 4 | | | | | | | |
| 14 | 13 | | | | | | | |
| 15 | 14 | | | | | | | |

*Figure 47: Sheet Products*

# Appendix I: Excel tool for monthly evaluating

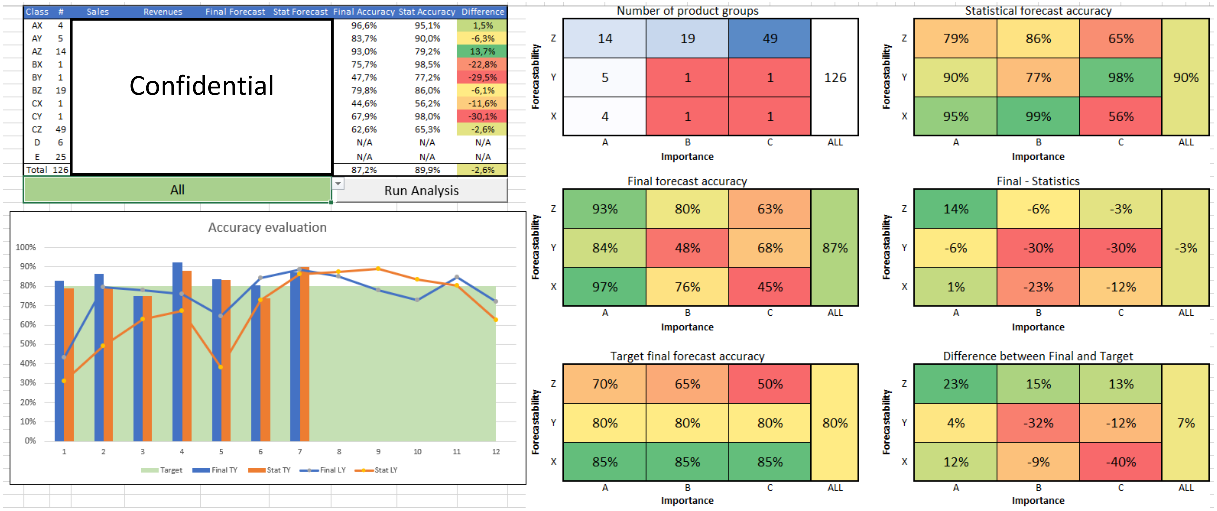



*Figure 48: Screenshot of monthly evaluating tool with main sheet*


*Figure 49: Screenshot of monthly evaluating tool with class specific diagrams*

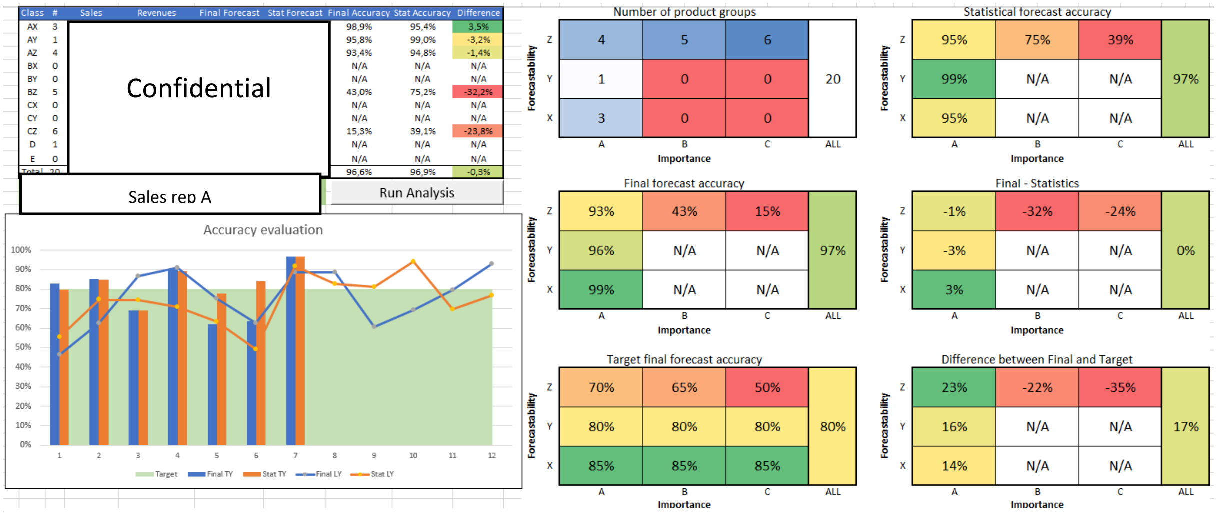



*Figure 50: Screenshot of monthly evaluating tool with a specific sales representative selected*

# Appendix J: Time series plots of total sales

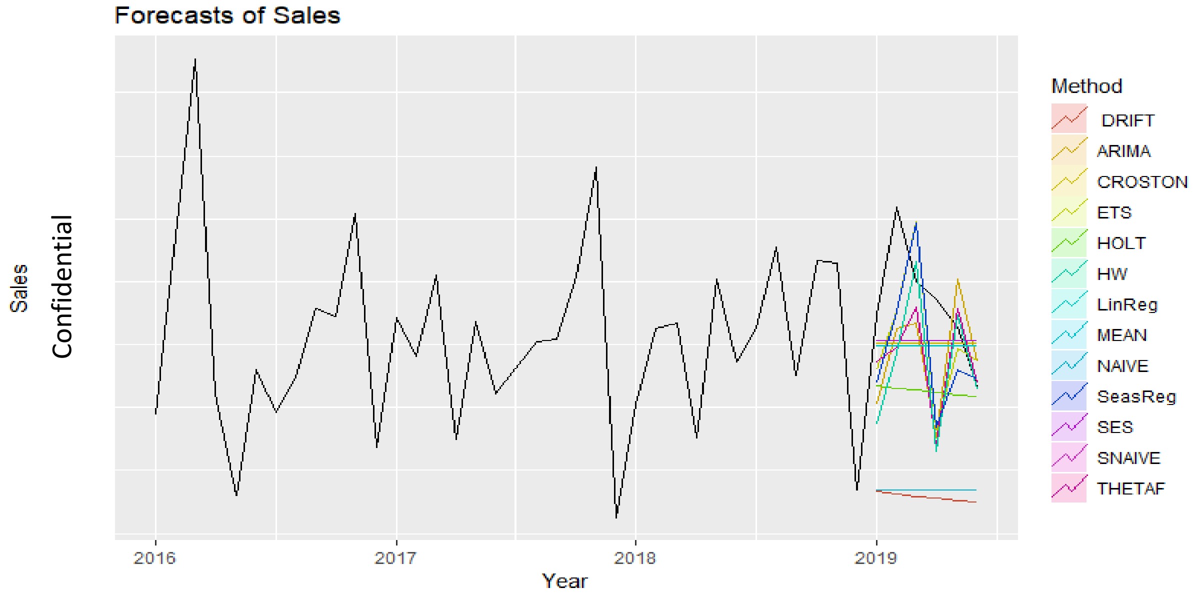



*Figure 51: Time series plot of Total Sales in 2016-2018 and the forecasts for 2019 for Country A*

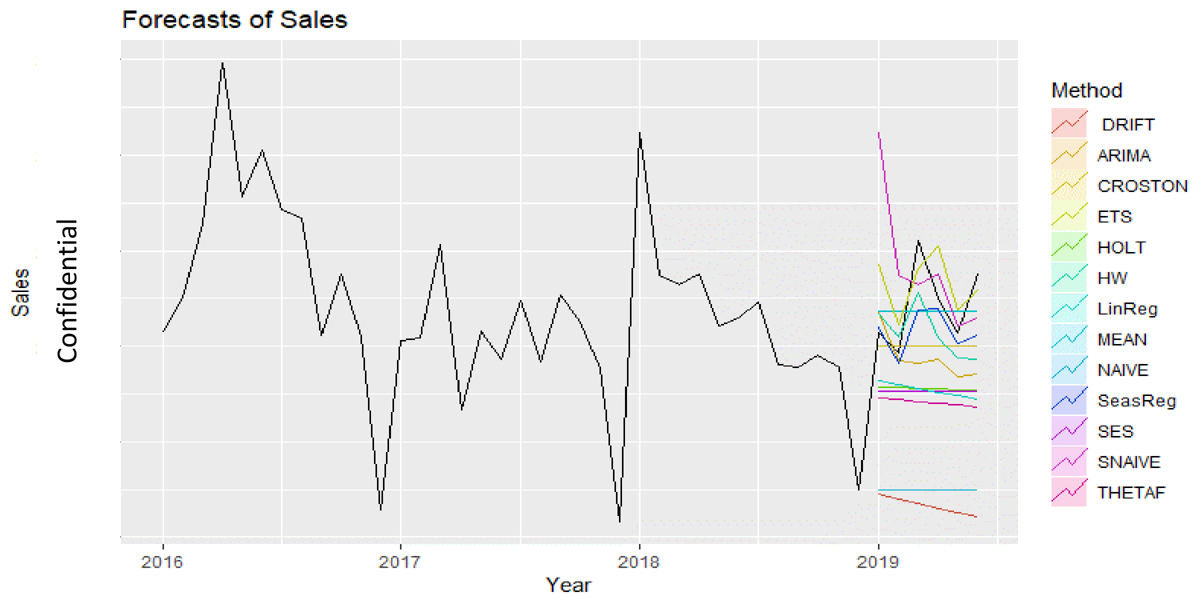



*Figure 52: Time series plot of Total Sales in 2016-2018 and the forecasts for 2019 for Country B*