MASTER

The impact of different levels of data aggregation on demand forecasting accuracy

Verschoor, Julian

Award date:
2023

Eindhoven University of Technology

Department of Industrial Engineering & Innovation Sciences

Julian Verschoor (1512978)

The impact of different levels of data aggregation on demand forecasting accuracy

---

Supervisors:
M (Martijn) Sprangers - Ricoh
D. (Dennis) Beerendonk - Ricoh
Dr. L. (Laura) Genga - TU/e
Dr. R. (Rob) Broekmeulen - TU/e

In partial fulfillment of the requirements for the degree of Master of Science in Operations Management and Logistics

Bergen op Zoom, March 27, 2023

Abstract

In this thesis the impact of different levels of data aggregation on forecast performance are studied. This concept is also referred to as hierarchical forecasting and aims at improving forecast performance by generating forecast at a different level of aggregation. Most companies generate forecasts on the "natural" level of aggregation on which they require to make all of their business decisions. However, several studies in literature proved that in some cases forecast accuracy can be improved by obtaining forecast values on a level of aggregation other than this one. In this study, this assumption is tested based on extensive amounts of demand data from a multinational specialized in (digital) office automation. Within the context of this study a total of three different dimensions are considered (i.e., product, customer and region dimension) from which results show that a middle-out strategy performs best. A middle-out strategy represents time series data which is not aggregated on the highest level nor on the lowest level of granularity. Furthermore, this study shows that there is no universally best level of aggregation, since it depends on the type of product and its underlying demand distribution. It is argued that many findings within this study are relevant for both practitioners and researchers.

Preface

This thesis concludes my Master degree in 'Operations Management and Logistics' and therefore marks the end of my time as a student at Eindhoven University of Technology. I would like to take this opportunity to thank several people who have guided me during the last couple of months.

First of all, I would like to express my gratitude towards Martijn Sprangers for guiding me through this process over the past months. During our meetings, he shared a lot of his feedback which provided me with valuable insights that most definitely has improved my work. His expertise and extensive knowledge on the demand planning department was exactly what I needed. Furthermore, I would like to thank Dennis Beerendonk for giving me the opportunity to write this thesis at Ricoh. His expertise in project management helped me a lot in understanding how you should set-up such a project. At last, I want to thank everyone within Ricoh who supported me throughout this project. This includes all participants who I was able to interview, people who helped me in gathering the data and in general all of the people who were always open to answer all of my questions.

Next up, I would like to thank both of my University supervisors. I want to thank Laura Genga for all of her feedback that she shared with me. Our regular meetings helped me a lot to come up with new ideas whenever I got stuck or had any doubts. Secondly, I want to thank Rob Broekmeulen for all of his feedback. All of the insights which he shared were especially of value in helping me to analyze all of the results.

Finally, I would like to thank all of my friends and family for their endless support. Each and everyone of them made my time as a student an unforgettable one.

Julian Verschoor, March 27, 2023

Executive Summary

Problem context
This thesis project takes place at Ricoh and addresses approaches to improve their demand forecasting accuracy. Ricoh is a leading multinational specialized in (digital) office automation for which this project is conducted specifically for the demand planning department located in Bergen op Zoom. Ricoh's location in Bergen op Zoom is responsible for all supply chain related tasks within the EMEA region. This includes the creation of a forecast on a regular basis for which management have expressed increasing doubts and concerns. The current forecast creation process includes forecast being derived based on statistical methods from which all of its predictions are reviewed by a team of demand planners. At this moment in time, many of these predictions are perceived as inaccurate by the demand planners and are therefore adjusted accordingly. Ricoh firmly believes that the amount of adjustments can be reduced by improving the reliability of the statistically generated predictions. However, contextual information that is only known by the demand planner results in the fact that some of the adjustments are just inevitable. That is why this project is focused on the forecast creation process as a whole from which improvements should contribute to an enhanced trust and alignment between a planner and a system.

Demand forecasting as a general concept is considered important for every supply chain and is therefore a topic that has received a lot of attention within literature. While most of the attention is dedicated towards the development of different quantitative forecasting techniques, other approaches also have proven to benefit the demand forecast process to some extent. The two approaches that are studied within this specific project are the hierarchical and judgmental forecasting approaches. Hierarchical forecasting aims at improving statistically generated predictions by generating those on a different level of aggregation, while judgmental forecasting is more focused on studying a planner's 'judgment' during the reviewing process. Researching both of these topics will enable this thesis to provide recommendations relevant for the forecast creation process from start to finish.

Research approach
The main aim of this research is to study whether a different level of aggregation can lead to any substantial forecast improvements. This resulted in the following main research objective:

Identifying and testing the impact of different levels of aggregation on demand forecasting performance.

This research starts of by describing the current forecasting procedure in more detail. Thereafter, its corresponding hierarchical structure is defined in which a total of three dimensions (i.e., product, customer, region) are used to break-down Ricoh's business. Each dimensions consist out of multiple levels of aggregation and forecast within this research are generated based on historical data that is aggregated across one of these levels within each dimension. Within the context of this research, the different aggregation combinations are referred to as a '*Hierarchical Scenario*'. A total of 36 different hierarchical scenarios are considered and tested with the aid of a forecast model that is created within Pyton.

Results
Based on a minimum of two years of historical data this Python model generates a 12-month forecast specifically for two types of products (i.e., machines and supplies). Forecast output is generated for both product types on all 36 different hierarchical scenarios from which all output is evaluated in terms of forecast accuracy. With the aid of different statistical approaches it is analyzed whether specific hierarchical scenarios significantly lead to more accurate forecast predictions. Based on these analysis, both product types clearly show poor forecast performance whenever forecast are generated on historical data that is either very much aggregated or mostly disaggregated. Furthermore, it is found that the impact of different levels of aggregation is much larger for the supply product type in comparison to the machine product type. This is mainly due to the fact that the supply product type only performs well on one specific level of aggregation within the product dimension. For the machine product type, a total of 11 out of 36 scenarios are never statistically outperformed by any of the other scenarios. For supplies there are only 5 scenarios out of a total of 36 scenarios which are not outperformed. Conducting two more detailed statistical analyses, comparing forecast performance of the first six months and on a predefined selection including only important SKU's, provided the same findings and therefore validate all of the earlier drawn conclusions.

Furthermore, in order to realize a better alignment between a planner and a system, a good understanding of the reviewing process is also considered crucial. To do so, a total of seven team members of the demand planning department are interviewed about their adjustment motives, whether they are missing important insights and if they feel like that the system should incorporate other additional (data) insights. It is found that most planners use year-on-year values and recent monthly demand values to determine whether an adjustments is needed. Whenever a planner recognizes an explicit demand pattern than most emphasis is put on year-on-year values, while monthly demand values are more considered whenever historical demand is perceived as erratic and volatile. Other motives to adjust initially generated forecast values are too better comply with standard seasonal behaviour, to take into account the impact of supply chain challenges or based on feedback from marketing (e.g., promotional activities, product introductions, sales projections). Moreover, these interviews showed that the current forecast approach of machines is a very manual one and therefore also sensitive to a lot of adjustments. The forecast approach for the supply product type, however, is much more system driven. It is argued by the interviewee's that demand data insights into different levels of aggregation could support them during the reviewing process. Finally, all interviewee's emphasised the need for a more structured data cleaning approach.

Conclusions and recommendations

This study emphasises the need for such research since its results showed that different levels of aggregation lead to substantial differences in terms of forecast performance. This study therefore contributes in drawing the attention of both researchers and practitioners as it demonstrates the potential for any company to improve their forecast performance through the concept of hierarchical forecasting. Moreover, results showed that there is really no universally best level of aggregation and that this depends on type of product that is considered. However, in general it is concluded that a middle-out strategy is favoured based on the statistical analysis conducted across all the different hierarchical scenarios.

In addition to more general conclusions, this research also provides more business specific conclusions. It is argued that Ricoh's forecasting process has much potential to be improved through a variety of aspects. First up, a different level of aggregation for especially the supply product type seems to improve the system generated forecast. It is therefore recommended to set-up a pilot phase in which during a longer time period forecast are generated for the most promising hierarchical scenarios (for both product types). Next up, it is concluded that there is still much potential to be gained by improving the data quality of the historical demand data that is used by both the system and a planner. At this moment in time, data is primarily cleaned through inconsistent manual initiatives and it is therefore advised that Ricoh defines a more structured and standardized data pre processing approach. Demand data that better reflects normal demand behaviour will provide more accurate system derived demand predictions, while it will also improve human interpretation. Finally, it is highly recommended that Ricoh starts to monitor forecast accuracy of both their system generated forecast and the final submitted forecast. Doing this will provide Ricoh with a better understanding of what kind of products can be forecasted solely based on statistical methods. Moreover, it will also provide insight into when and why an adjustment improved a forecast and even more importantly when it did not.

Contents

List of Figures

List of Tables

Acronyms

AP Asia-Pacific. 45

CV Coefficient of Variation. 38, 45–48, 50–53, 67

DES Double Exponential Smoothing. 17, 38, 42, 45

EDC European Distribution Center. 1

EMEA Europe, Middle-East and Africa. iii, 1, 2, 4, 16, 25, 34, 65

HWA Holts-Winter Additive. 17, 42

HWM Holts-Winter Multiplicative. 17, 42

MAPE Mean Absolute Percentage Error. 13, 15, 17, 40, 44–48, 50–55, 67

OpCo Operational company. 25, 34, 35, 46, 47, 60, 61, 63, 66

RMSE Root Mean Squared Error. 44, 45

SES Simple Exponential Smoothing. 17, 38, 41–43

SKU Stock-keeping unit. iv, 4, 12, 16, 17, 19, 20, 22–24, 26, 28, 30–33, 37–39, 44, 46, 48–51, 53, 56, 58–63, 66, 67

Chapter 1

Introduction

1.1 Company description

This Master thesis is conducted at Ricoh, a leading multinational founded back in 1936 and specialized in (digital) office automation. Their core business is and has been traditionally focused on the production and sales of many types of copiers and printers (see Figure 1.1). Since financial margins within these traditional markets are slowly shrinking, Ricoh is shifting their business towards being more of a digital service provider. Over time, Ricoh grew into a huge global organization with over 110,000 employees, generating approximately €15 billion in annual revenues, and selling their products in over 200 countries.

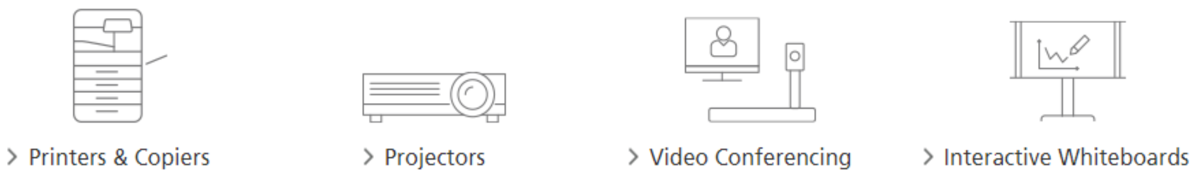


Figure 1.1: Products sold within Ricoh's product portfolio

To be able to efficiently manage such a huge international organization, Ricoh splits its operation into multiple operational regions (e.g. Europe, Asia, America). Each of these regions are operated out of multiple offices, and are all responsible for a different business area. For instance, one location is mainly responsible for all marketing activities, while another location is responsible for all supply chain related tasks. This thesis project is focused on Ricoh Europe and more specifically on their European Distribution Center (EDC) located in Bergen op Zoom. This location carries responsibility for a lot of supply chain related activities within the Europe, Middle-East and Africa (EMEA) region such as warehousing, configuration, distribution, and transportation processes. Furthermore, Ricoh Bergen op Zoom has five main departments, Finance, Demand Planning, Order Fulfillment, Operations and Strategy & Business Excellence. This thesis project is part of the Demand Planning operation, as its topic revolves around demand forecasting.

1.2 Demand forecasting

In today's business world, it is essential to accurately predict future demand. Generating accurate forecast figures supports a supply chain in their strategic and operational decision making (Lawrence et al., 2006). A good quality and precise forecast process results in improved service-levels and more optimal inventory levels (Hyndman & Athanasopoulos, 2018). However, according to Syntetos et al. (2016), consistently providing reliable forecasts is often proven to be a challenging task for many companies. More specifically, Rostami-Tabar et al. (2013) stated that future demand is impacted through many different variables that all contribute to an increased demand uncertainty. The paper of Chen & Blue (2010) even stated that demand uncertainty is currently one of the biggest management challenges. Companies are therefore often supported by some kind of forecasting system, enabling them to better deal with these uncertainties. Most of these systems apply statistical models on historical demand information to systematically derive forecast figures (Petropoulos et al., 2022).

While uncertainties are in some extent inevitable, in order to increase the reliability of forecast calculations, one must reduce and control it as much as possible. In addition to adopting quantitative statistical methods, other approaches also have proven to improve the demand forecasting process to some extent. According to Fildes et al. (2022), one method to provide more accurate demand forecasts is through the concept of hierarchical forecasting. Hierarchical forecasting aims at improving forecast reliability by generating demand forecast at different levels of aggregation. Babai et al. (2022) claims that in some cases forecast accuracy may be improved since the 'natural' level on which forecasts are obtained not always proves to be most accurate. Another approach which is increasingly acknowledged in literature, is the implementation of human judgment within the field of demand forecasting (Perera et al., 2019). This concept, also referred to as judgmental forecasting focuses on introducing a planner's *"judgment"* in creating or adjusting a forecast.

1.3 Problem context

Ricoh Europe has over 4,000 products in their product portfolio, actively sold in over 25 different countries within the EMEA region. Each of these products all have their own future demand patterns which have to be predicted as accurately as possible. Overestimating customer demand could lead to high (potentially obsolete) inventories, while underestimating demand could lead to dissatisfied customers due to bad product availability. These consequences, alongside rapidly changing customer demands, emphasis the need for an accurate forecast.

At this moment in time, Ricoh's management has increasing doubts and concerns about their current forecast creation approach. More specifically, Ricoh firmly believes that the forecasting process includes potential to be improved. That is why in this particular study, Ricoh wants to challenge their as-is forecasting processes. The current forecast process of Ricoh includes forecast being statistically derived based on historical data from which its output is reviewed by a team of demand planners. During the reviewing process a planner needs to decide if he/she agrees with the system generated forecast and if not, a decision needs to be made on the kind of adjustment that is required. Ricoh's current system-driven forecast is often perceived as not accurate enough, resulting in many essential but inefficient manual adjustments.

The currently used level of aggregation has never been challenged and therefore made Ricoh wonder whether other levels of aggregation prove to result in more accurate forecast predictions. However, unusual demand, promotional activities, product introductions and several supply chain disruptions (e.g., Covid-19, material shortages) are perfect examples of factors that are hard to be captured by any statistical model. Such factors make the manual reviewing process in some extent inevitable and it is therefore unrealistic to assume that Ricoh can rely on a forecast solely generated by system. That is why, in addition to the system generated forecast, Ricoh believes that there is also a lot of potential for improvement within the reviewing process. Improving the forecast creation process as a whole should contribute to an enhanced trust and alignment between the system's derived predictions and a planner's judgment. Realizing a better alignment might contribute to a more accurate forecast in which less manual (inefficient) adjustments are required. In order to realize a better alignment between a planner and a system, a good understanding of the reviewing process is considered crucial.

Furthermore, at this moment in time Ricoh initiated a huge global ongoing project in which the entire organisation is shifting to the same supply chain planning system. This new system also comprises new specialized forecasting software and can therefore also be interpreted as one of the triggers for this study. To conclude, the main research interest for Ricoh are the following:

- Providing insight into different levels of data aggregation that might be of interest for Ricoh's forecasting process.
- Testing these different levels of aggregation in order to conclude whether a different level of data aggregation improves the system's obtained forecast accuracy.
- Achieving a good understanding of the current reviewing process. This includes providing insight into into the demand planner's motives to adjust the system generated forecast, whether they are missing important insights and if they feel that the system should incorporate other additional (data) insights.

1.4 Research objective

All information from these previous sections is used to formulate a concise and comprehensive objective. The formulated research objective is shown below. Achieving this research objective will create a better insight and understanding of future demand patterns due to an improved forecast accuracy and efficiency.

Research objective:
Identifying and testing the impact of different levels of aggregation on demand forecasting performance.

In order to achieve this research objective, several sub-research questions are formulated. The overall (common) goal of answering these questions is to challenge the as-is forecast situation. Each of the research questions below are formulated such that the objective is achieved in a structured manner in which all important research phases are considered and included.

Phase 1: Hierarchical forecasting

1. *What is the as-is forecasting procedure?*
2. *Which alternative levels of aggregation are of interest and need to be tested?*
3. *What kind of model is used to test these different levels of aggregation?*
4. *How should the impact of all these tested levels of aggregation be measured and evaluated?*
5. *Which of the levels of aggregation result in the best forecast performance?*
6. *Which validation steps can be taken in order to assure that all gathered results are considered valid?*

Phase 2: Judgmental forecasting

1. *What is the current judgmental forecasting procedure?*
2. *What are the variables/factors that human decision makers consider to access whether a forecast adjustment is necessary?*
3. *What are additional insights that might support a human decision maker during their reviewing process?*
4. *What are additional variables that might be of interest to be incorporated into the forecasting system?*

1.5 Research scope

First of all, as already briefly mentioned in section 1.1, this thesis is focused on Ricoh Europe whom are responsible for all Ricoh related activities within the EMEA region. More specifically, this thesis is conducted for the demand planning department who are responsible for all forecasting related task and are located in Bergen op Zoom. This scope is necessary as conducting this research on Ricoh's worldwide organisation is far too complex and unrealistic looking at the given time-span.

Moreover, Ricoh's product assortment in terms of demand volumes for the larger part consist out of three types of products (i.e., Machines, Options and Supplies). Ricoh's machine product range includes a wide assortment of printers, copiers, production printers and whiteboards. Often sold in combination with one of these machines are one the options Ricoh has to offer. These are products which if requested by a customer can be installed on one of these machines. Finally, supplies can be best explained as products that a machine requires to have to be able to properly function (e.g., toners, cartridges). For specific reasons, discussed in section 3.5, it is decided to proceed this project only on the machine and supply product types. The data as gathered in chapter 5 includes a total of 1036 unique SKU's (i.e., 330 machine SKU's and 706 supply SKU's). Due to various reasons, which are discussed throughout this report, this total number of SKU's is eventually scoped down to a total of 57 machine SKU's and 232 supply SKU's.

In order to conduct this research a total of four years of historical (demand) data is taken. The first three years of this data-set are used by the system to determine the best fitting algorithm and the final year is included to evaluate forecast accuracy. As Covid-19 heavily impacted demand patterns, the data will originate from the period before the pandemic

(2016-2019). By doing this, research is assumed to provide a more fair comparison. More details on this specific time-period are also shared in section 3.5. Finally, the decision is made to focus on the new forecasting system which is not currently in use just yet. This is done to prevent recommendations that might turn irrelevant from the moment the new system is in use. All technical details of this new system are shared in chapter 6.

1.6 Research methodology

This research follows the 'Model-based and integrated process improvement' (MIPI) methodology developed by Adesola & Baines (2005). This seven-step procedure will function as the framework of this thesis and is shown in Figure 1.2 below. This model is customized to this research specifically, meaning that it is mostly used as a general framework.

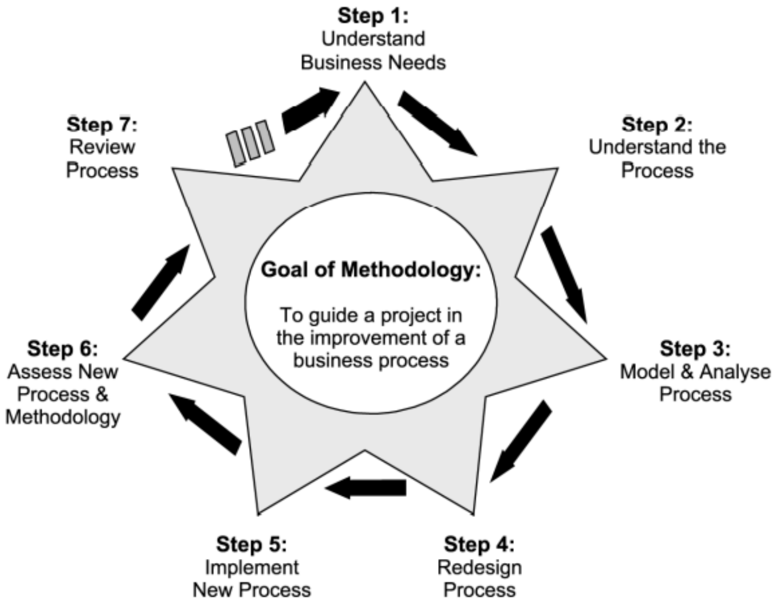


Figure 1.2: Seven-step MIPI methodology

This methodology starts of with the aim to understand the business needs in terms of the project. This is an important first step in any research and focuses on emphasising project necessity by defining current challenges, scoping the research area, aligning research goals, finding relevant literature and more. This is done in the first chapter, which introduces the project, and the second chapter which includes a literature review that discusses literature relevant to this topic. The main topics discussed in this review revolve around hierarchical and judgmental forecasting approaches.

The remainder of this thesis consist out of seven more chapters and are structured according to this methodology. The second step is focused on understanding all processes executed within the pre-defined scope of the project. In other words, describing the as-is situation of all of the relevant processes. That is why in the third chapter, Ricoh's current forecasting procedure is discussed. This includes providing a high-level overview on both the systematical and judgmental forecasting approaches that Ricoh currently uses. Thereafter, in the fourth chapter, the hierarchical structure is defined within the context of Ricoh's business. This structure is discussed in order to conclude on which of these levels of aggregation are of most interest for this research specifically.

The third step of this methodology is focused on modelling and analysing these processes. All of the data that is collected is discussed in the fifth chapter in which data insights are provided for each of the levels of aggregation that were included throughout the fourth chapter. The sixth chapter is subsequently devoted to sharing details on the forecast model that is used to test the impact of all of these different levels of aggregation. The model is created within a Python script and represent a replication of Ricoh's new forecasting system. In order to provide an insight from start to finish into Ricoh's general forecast approach, this step also includes analyzing the judgmental forecasting approaches in more detail. That is why in the eighth chapter interview input from several demand planners is used to get insight into adjustment motives, missing insights and promising variables that might be interesting to be incorporated into a system.

The fourth and fifth steps of this research methodology includes redesigning a process and implementing it. These steps are conducted by generating output as discussed in the sixth chapter based on the different levels of aggregation as defined in the fourth chapter.

The sixth step is a very important one and is used to assess all of the results. This step is conducted throughout the seventh chapter of this thesis in which all of the generated output is analyzed. Multiple statistical approaches are used to conduct (detailed) analysis in order to observe which levels of aggregation perform well and which do not.

Finally, the most important findings are discussed in the last chapter (i.e., chapter 9) and corresponds with the final step of this research methodology (i.e., Review process). Multiple research and business findings are shared, while this chapter also elaborates on research limitations and shares ideas for future research.

Chapter 2

Literature review

This chapter is focused on reviewing and describing literature that is relevant to this thesis project. More specifically, forecasting literature is addressed that discusses hierarchical and judgmental forecasting approaches. The aim of this chapter is to provide a good understanding on both of these topics that are relevant to this thesis project.

2.1 Demand forecasting

Forecasting as a general concept is a traditional research area which is well-known and often addressed within scientific literature. It is a broad topic which is applied in many disciplines like finance, statistics, public policy, economics, science, engineering and many others (Montgomery et al., 2015; Mahalakshmi et al., 2016). Within the context of this thesis project, forecasting is studied from a demand forecasting perspective. The paper of Albarune & Habib (2015), describes demand forecasting as a task in which all available knowledge, including historical data is used to come up with a predicted actual value for a future time period. Petropoulos et al. (2022) uses a similar terminology and states that forecasting is based on the premises that future predictions can be made through the usage of current and past knowledge. According to Kalchschmidt (2012), forecasting is considered an important task given its impact on the decision making and strategic planning of many departments. The benefits of good quality forecasts are among others, an efficient use of resources due to operational plannings being made well in advance, operating with lower inventory levels and reducing the risks of stock-outs (Hyndman & Athanasopoulos, 2018).

Given those benefits, both academics and practitioners constantly strive to develop methods aimed at improving this forecasting process. In demand forecasting literature, a lot of attention is therefore devoted to the development and adaptation of quantitative methods. These papers either study the performance of existing techniques, propose an adjustment to one of these existing technique or introduce a completely new developed technique (Zotteri et al., 2005). The common feature of all of these studies is that they aim to improve forecast accuracy to some extent. Over the years, many statistical methods have been developed which range from very simple (intuitive) algorithms to much more sophisticated/complex techniques (De Gooijer & Hyndman, 2006). Although these complex techniques have been developed to improve forecast accuracy, several studies have shown that simpler techniques still prove to have comparable performance (Ouwehand, 2006). This emphasis the fact that being able to accurately predict future demand is proven to be a very difficult task.

According to Rostami-Tabar et al. (2013), being able to accurately predict future demand heavily depends on the level of uncertainty within the underlying time series data. The more continuous and less volatile demand is, the easier it is to accurately predict future demand (Kerkkänen et al., 2010). However, accurate forecasts are unfortunately especially welcomed for demand patterns which prove to be the opposite. In order to tackle this demand uncertainty, researchers have not only focused on developing quantitative statistical methods, but also studied other approaches (Duncan et al., 2001). The approaches that are discussed in this chapter are the hierarchical and judgmental forecasting approaches. Historically, these approaches have received much less attention, but over time, researches have looked more and more into these approaches.

2.2 Hierarchical forecasting

A recurring topic in literature to improve forecast performance is focused on the level of aggregation on which forecasts are calculated (Babai et al., 2022; Fildes et al., 2022). This topic is also referred to as hierarchical forecasting and examines the impact of obtaining forecast through different levels of aggregation. The underlying idea within this approach is that if there are products, regions, customers or other dimensions that have similar demand characteristics, clustering them could potentially improve all individual forecasts (Hyndman et al., 2011; Rostami-Tabar et al., 2015). More specifically, Jin et al. (2015) summarizes the choice of the level of aggregation as a trade-off between variance reduction and capturing individual time series information. The concept of reducing uncertainty through aggregation techniques is also referred to as risk-pooling in the academic community (Chen & Blue, 2010).

The relevance of this topic is summarized by Babai et al. (2022), since they claim that the most accurate forecast are not always obtained from data at the "natural level" of aggregation. That is why, according to Zotteri et al. (2005), an important aspect whilst implementing any forecasting technique is to determine the level of aggregation on which these techniques have to be adopted and evaluated. Traditionally, most companies forecast future demand for a single product over a fixed time bucket and a given region. However, the papers of Ouwehand (2006) and Lapide & LaPiDe (2009) state that forecasting individual items have become increasingly more difficult over the past decades as most businesses have changed. Assortments have grown exponentially, from which many products have a significantly shorter life cycle which is further complicated by the fact that most products are sold in more countries than ever (Lapide & LaPiDe, 2009). As a result it may be no longer optimal to calculate forecast for each individual stock-keeping unit separately (Ouwehand, 2006).

These findings in addition with the earlier mentioned statement by Babai et al. (2022) have motivated researchers to devote more time and attention towards research focused on selecting the most optimal aggregation level to calculate forecast on. Within hierarchical forecasting literature, two types of aggregation methods can be distinguished. As stated by Syntetos et al. (2016), Fildes et al. (2019) and Babai et al. (2022), data can be either aggregated across time (e.g., aggregating data on a daily, weekly, monthly time-span) or aggregated across a series (e.g., aggregating data on a product-family level). The former one is called temporal aggregation, while the later one is referred to as contemporaneous aggregation. After briefly explaining both of these approaches, often used clustering dimensions and evaluation methods are discussed in more detail.

2.2.1 Temporal aggregation

According to Rostami-Tabar et al. (2013), the general idea of temporal aggregation is to aggregate demand data across so-called "time-bucket frequencies". Babai et al. (2022) refers to it as the process of deriving low-frequency time series into higher frequency time series (e.g., aggregating daily demand data into weekly or even monthly demand series). Athanasopoulos et al. (2017) describes temporal aggregation as a tool to highlight and understand available data from different perspectives. Trend cycles are better highlighted when data is aggregated, while seasonality could potentially be better distinguished on a lower frequency. Research states that benefits originating from temporal aggregation includes reducing demand fluctuations/uncertainty, which could potentially result in an increased forecast accuracy (Kourentzes et al., 2017). Nikolopoulos (2021) even argued that applying a different temporal frequency could in theory improve any statistical or machine learning model.

Overlapping and non-overlapping are often two types of temporal aggregation techniques which are distinguished by academics (Rostami-Tabar et al., 2016). The non-overlapping technique divides demand into successive blocks equal to the selected aggregation level, while the overlapping technique constructs successive overlapping blocks where at each time period the block is moved one period (Babai et al., 2022). The paper of Rostami-Tabar et al. (2022) used an numerical example (See Figure 2.1) to illustrate how both techniques aggregate demand data from a non-aggregated time series.

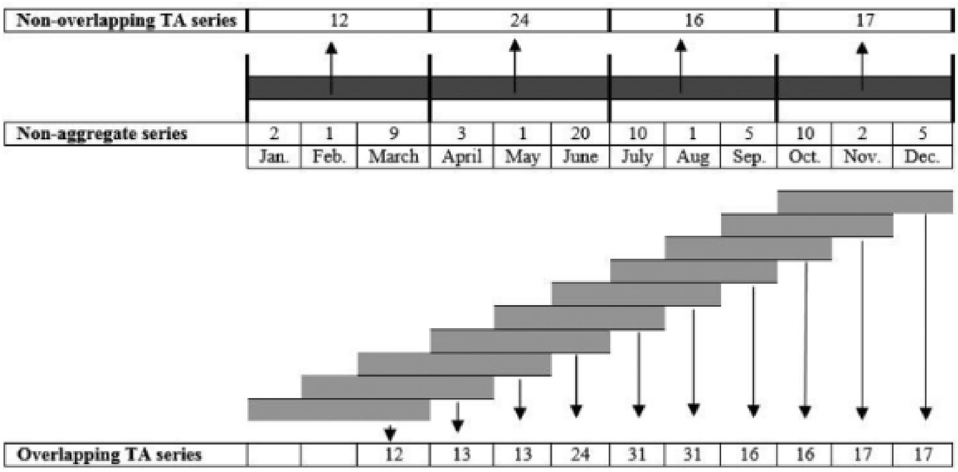


Figure 2.1: The overlapping and non-overlapping temporal aggregation techniques

The disadvantage of the non-overlapping technique is that fewer data-points are constructed in comparison to the overlapping approach. As a result, the information loss (using the overlapping technique) compared to the non-aggregated series is negligible (Rostami-Tabar et al., 2022; Babai et al., 2022). This is an important aspect to consider, especially in situations in which there is not much demand history available. However, unlike the overlapping technique, the non-overlapping approach retains the original auto-correlation structure (Babai et al., 2022).

2.2.2 Contemporaneous aggregation

Contemporaneous aggregation also referred to as cross-sectional or hierarchical aggregation (Babai et al., 2022), is according to Chau (2020) an aggregation process that merges time series data based on different dimensional characteristics (e.g., product, location or customer characteristics). Within the domain of hierarchical forecasting literature, and specifically for contemporaneous aggregation methods, there are three approaches which are often proposed, discussed, evaluated and compared. Those are the Top-Down (1), the Middle-Out (2) and the Bottom-Up approach (3).

Many authors have conducted empirical research in which the forecast performance between two or three of these approaches are compared. Although many authors have tried to argue that one approach significantly outperforms the others, results have often be contradictory (Syntetos et al., 2016). According to Sbrana & Silvestrini (2013), the general question, raised by Dunn et al. (1976), of which aggregation level should be used to obtain forecasts remains unanswered.

Bottom-Up approach

The Bottom-Up (BU) approach is based on forecasts being calculated on an individual demand segment level. Subsequently, whenever a forecasts at an aggregated group level is required, the individual forecasts results are simply summed to the required group level (Fliedner, 2001). That is why this approach is also known and referred to as cumulative forecasting. According to studies from Babai et al. (2022) and Syntetos et al. (2016) the BU approach is favourable whenever the demand characteristics across time series are significantly different. As the BU approach does not aggregate any data, it has a great advantage of causing not to lose any individual information (Athanasopoulos et al., 2009). Meaning that all individual dynamics/characteristics can still be captured and distinguished (Syntetos et al., 2016). However, the downside is that these individual time series are prone to be noisy, making it more challenging to capture these individual dynamics in the first place (Dangerfield & Morris, 1992).

Top-Down approach

The Top-Down (TD) approach works the other way around, as forecasts are calculated on the aggregated data-set and are disaggregated back down on a lower granularity (if required by the organisation) (Schwarzkopf et al., 1988). These dis-aggregated forecasts are also referred to as derived forecasts. Within literature there are multiple dis-aggregation methods proposed and discussed. Gross & Sohl (1990) have examined many methodologies to dis-aggregate aggregated forecast back to an individual level. In total 21 different disaggregation schemes were examined which included simple averages, lagged proportions and combined lagged proportion techniques. The lagged approaches were based on the assumption that a single recent time period were the best predictors for providing proportions for a future period, while the combined lagged proportion technique looked at a weighted combination of all these lags together. While different lag values and weights have been tested on both of these techniques, their research showed that the intuitive mean proportional dis-aggregation methodologies (i.e., simple averages) is still found to be most effective (Gross & Sohl, 1990).

One of the general ideas of the TD approach is that calculating forecasts on a dis-aggregated level is more prone to errors making aggregated forecasts to be more accurate (Dangerfield & Morris, 1992). Furthermore, many authors (Zotteri et al. (2005); Rostami-Tabar et al. (2013)) have argued that if demand is relatively stable, the top-down is found to be favourable. Both through a potential greater accuracy and its lower costs. Moreover, the TD approach is proven to be fairly accurate and especially useful for low volume time series. Such intermittent demand time series are fairly hard to predict, and applying aggregation techniques are proven to be effective in reducing demand intermittency ant therefore results in an improved forecast performance (Lei et al., 2017; Widiarta et al., 2007). However, one of the major weaknesses of the TD method is that a consequence of aggregating data is an loss of information, making it more challenging to capture individual time series characteristics such as individual trends or seasonality.

Middle-Out approach

The Middle-Out (MO) approach is, as the name already indicates, an approach which combines the BU and TD approaches (Athanasopoulos et al., 2009). Forecasts are neither calculated on the lowest aggregation level nor on the highest aggregation level (Babai et al., 2022). In other words, forecasts are calculated at an intermediate level within the hierarchy chain, for which dis-aggregation techniques are used to obtain lower level forecasts and aggregation techniques to obtain higher level forecasts (Hyndman et al., 2011). According to Syntetos et al. (2016), the MO method is not always distinguished as a separate approach since it is often subsumed under the other two methods within hierarchical forecasting literature. When an approach is considered from "above", MO is often defined as bottom-up, while considered from "below" if it is defined as top-down. That is why the drawbacks of this approach also depend on the approach with which it is compared. Whenever it is compared with a TD approach, the MO approach is more vulnerable for individual noise, while it is more prone to a substantial loss of information when it is compared with the BU approach (Syntetos et al., 2016). Zotteri et al. (2005) draws similar conclusions since they observed that their MO method tends to be outperformed. Their MO model seems to fail in capturing individual differences, while on the other hand also still uses rather noisy individual data.

New approaches

In more recent literature, these approaches are considered outdated, as they do not incorporate all available data (Pennings & Van Dalen, 2017). These traditional approaches isolate the data from just one specific hierarchical level, and thus ignore all available data given in each of the other hierarchical levels. That is why the paper of Hyndman et al. (2011) proposed a new approach in which forecast are obtained from multiple levels within the hierarchy. In general, their idea is based on individually forecasting time series from all relevant hierarchical levels, and reconciling all of these into a single optimal forecast by using a regression model. In doing so, Hyndman et al. (2011) argues that an unbiased final forecast including minimum variance across all individual forecasts is obtained. Empirical research in which tourism demand in Australia is forecasted is used as a case study to compare the performance of this newly introduced approach with the traditional TD and BU approach. Results have shown that the "*optimal combination approach*" method outperforms the other two approaches.

Some other potentially promising ideas are summarized by the paper of Babai et al. (2022). These developments are similar to the "*optimal combination approach*" discussed by Hyndman et al. (2011) and are also focused on fully utilizing all available information on different levels of aggregation. This could be done for both the contemporaneous or temporal aggregation approaches, but also through the combination of the two. In other words, effectiveness and performance of both approaches have been studied individually, but not often considered together (Di Fonzo & Girolimetto, 2023). Despite some studies analytically proving benefits of these methods, empirical studies are still lacking behind (Babai et al., 2022).

2.2.3 Hierarchical dimensions

While the previous sections created an understanding of the aggregation concepts through temporal and contemporaneous approaches, this section is devoted to showing the different dimensions that are used to aggregate time series on. For temporal aggregation this is quite straightforward, since data is always clustered based on a specific time-frequency. Daily demand is converted into weekly or even monthly aggregated data. However, for contemporaneous aggregation studies showed that data can be aggregated across a variety of hierarchical dimensions.

All the papers which studied hierarchical forecasting within the context of the Australian domestic tourism (Kourentzes & Athanasopoulos (2019); Hyndman et al. (2011); Athanasopoulos & Hyndman (2008)) have clustered their non-aggregate time series based on geographical locations. Their most aggregated level was the entirety of Australia, which is dis-aggregated down through seven states, 27 zones and 76 unique regions. In the analyses conducted by Zotteri et al. (2005), only the location attribute was used to cluster demand (e.g., demand was aggregated over a total of 38 stores). Aggregating over a different dimension like the product-aggregation was not considered a valid option as only a limited amount of SKU's were considered. Furthermore, in addition to an aggregated model (e.g., demand aggregated over all 38 stores), also an clustered model is introduced. This model clustered stores which had similar demand patterns based on the penetration rate and are defined through a k-means algorithm. As a result each SKU is divided into three unique store clusters. Finally, the paper of Ouwehand (2006), has clustered demand data trough the concept of product-aggregation. Unlike other studies in which standard product characteristics are often used to aggregate SKU's, Ouwehand (2006) constructs product-families consisting out of products having similar seasonal patterns. This idea is based on the assumption that clustering products with similar seasonal patterns will result in a better determination of seasonal indices as noise on the individual level is reduced.

2.2.4 Evaluating hierarchical performance

In order to compare all of these previously described approaches, most studies adopt some kind of an evaluation metric. Moreover, the paper of Perera et al. (2019) mentions the importance of a good evaluation metric as perfect accuracy is impossible in practice, and forecasts methods therefore need to be compared and evaluated through appropriate metrics. While it is emphasized by Armstrong & Forecasting (1985) that there is no universally accepted accuracy measure, the paper of Perera et al. (2019) found that absolute percentage errors such as the MAPE is a frequently used metric by both academics and practitioners. Another metric frequently used within these kind of studies is the mean squared error (MSE) (Rostami-Tabar et al., 2013). However, as the error is squared it is not scaled to the original error, which is often reasoning for practitioners to evaluate forecasts through different measures. The paper of Armstrong (2001) emphasis practitioners to have a good understanding of the advantages and drawbacks associated to the different error measures.

Other studies do not limit themselves in using only one error metric. The paper of Ouwehand (2006), for instance, evaluates forecast performance through the MAD, MSE and symmetric MAPE measures. Armstrong (2001) recommends all academics and practitioners to use multiple error measures, since they claim that error measures must ensure (face and construct) validity. Finally, Armstrong (2001) recommends the usage of statistical significance tests to judge whether accuracy differs among different reasonable forecasting methods. In such comparisons, a conventional null hypothesis is that the methods are equally accurate. That is why, in addition to evaluating forecast performance through the MAPE, Zotteri et al. (2005) used paired-sample t-tests in order to conclude on any significant mean differences between the sample results of their considered approaches.

2.3 Judgmental forecasting

While the last few sections were focused on statistical forecast models, from which performance is further improved through the concept of one of the aggregation techniques, academics have also increasingly acknowledged the importance of human judgment within the field of demand forecasting (Perera et al., 2019). At first, there was a lot of scepticism about the usage of judgmental forecasting as literature argued that it did not add any value towards the forecasting process (Hogarth & Makridakis, 1981; Carbone et al., 1983). One argument often raised against judgmental adjustments was that human decision makers might misinterpreted noise as systematic patterns in the associated time series (O'Connor et al., 1993; Eggleton, 1982; Hogarth & Makridakis, 1981). However in the last few decades, according to Lawrence et al. (2006), this sentiment has changed as many studies showed that statistical generated forecasts can be improved through the expertise of a planner (Armstrong & Fildes, 2006; Perera et al., 2019; Lawrence et al., 2006). The paper of Seifert et al. (2015) claims that people may be very capable in capturing interactions between predictions variables, which are overlooked by statistical techniques. Furthermore, Perera et al. (2019) and J. Alvarado-Valencia et al. (2017) argue that human adjustments are of most value whenever human decision makers possess information that is not incorporated within the statistical models. This could be information regarding price changes, stock-outs, competitor behaviour, promotional activities, holidays and many other factors which influence demand dynamics (Fildes & Goodwin, 2007). Such information is often referred to as "*Contextual information*" and is valuable information which should ideally be incorporated into the forecast model (O'Connor et al., 2000;

Sanders & Ritzman, 1992, 1995; J. A. Alvarado-Valencia & Barrero, 2014). Unfortunately, contextual knowledge is difficult to consequently and accurately incorporate within the forecast creation process.

The paper of Perera et al. (2019) summarizes the concept of judgmental forecasting through the explanation of three different approaches. Forecasting based solely on human expertise (1), adjusting statistical derived forecasts (2) or combining statistical derived forecasts with a forecasts purely created from human expertise (3). The second approach is the one which is predominately used in the field of judgmental forecasting and will therefore be the focus of the remainder of this section. This approach is relatively simple and consist out of two stages (Lawrence et al., 2006; Perera et al., 2019). A planner first has to review the output from the statistical forecast and subsequently needs to decide whether an adjustment is necessary. Subsequently, if a planner has decided that an adjustment is required, they have to decide whether the original prediction need to be adjusted downwards or upwards and by how much. Studies from Franses & Legerstee (2009) and Fildes et al. (2009) have shown that the usage of such a hybrid approach led to many statistical forecasts being adjusted (approximately between 65 and 90 percent).

While academics acknowledged the value of human interference, a recurring message in literate still is that human intervention could both strengthen and harm forecast accuracy depending on the person(s) judgment, available information and various other factors (Davydenko & Fildes, 2013; Fildes & Goodwin, 2007). Many papers have concluded that statistical adjustments are prone to several human biases making judgmental adjustments not beneficial in all situations (Sanders & Ritzman, 2004; Lawrence et al., 2006; Armstrong, 2001). This has motivated both academics and practitioners in their desire to understand how and when human judgment should be interfering with statistical forecasts (Lawrence et al., 2006). A recent published paper from Arvan et al. (2019) argued for instance that the characteristics of the time series data is an important determinant in whether human interference is beneficial. High volatile time series are often hard to be estimated only through statistical estimates and could therefore be improved through a planner's involvement. According to J. Alvarado-Valencia et al. (2017), human responses are affected through personal attributes of the planners, the way of demonstrating information and the number and selection of the responsible planners.

The study performed by Eroglu & Croxton (2010) examines certain forecasts biases. Throughout their paper three types of biases are considered: optimism bias (1), anchoring bias (2) and overreaction bias (3). Within this context, bias is defined as a systematic deviation as opposed to a random one (Harvey, 2001). Optimism bias refers to a planner's tendency to predominantly adjust forecasts in the upward direction, resulting in positive errors. Anchoring bias describes a situation in which a planner correctly judges the necessary adjustments, but is reluctance to deviate from the generated statistical "anchor" value. While the overreaction bias refers to a situation in which the direction is adjusted correctly, but overshoots the actual value such that it yields in a larger error compared to the error associated with the statistical suggested value (Eroglu & Croxton, 2010). Empirical research conducted by Fildes et al. (2009) showed that overall relatively larger adjustments resulted in accuracy improvements, while the smaller adjustments often damaged forecast accuracy. Moreover, negative adjustments were more likely to improve the system generated forecast compared to the positive adjustments (Fildes et al., 2009). Finally, van der Staak & Johanna (2021) tried to identify and summarize planner's behaviours, such that their strengths can be combined with supporting forecasting systems. One of the interesting findings they found is that planners always tend to choose rounded numbers, which are proven to negatively impact forecast performance. Furthermore, similar to Fildes et al. (2009) research they found that planners are great at determining the forecast direction and also the size of a downward adjustments, but perform worse in choosing the size of an upward adjustment (van der Staak & Johanna, 2021).

2.3.1 Evaluating judgmental performance

All the different evaluation methods already discussed in subsection 2.2.4, are generally applicable in the field of demand forecasting. Performance metrics like the MAPE, MAD and the MSE are also frequently used to evaluate performance of planner's (Fildes et al., 2009). Whilst statistical (hierarchical) forecast strategies primarily use such output to compare and evaluate forecast performance, judgmental forecasting also uses such output to enable a planner to learn (Lawrence et al., 2006). In other words, output from these metrics is often translated into some sort of feedback which is shared in order to improve future adjustments. To be able to share such feedback, Eroglu & Croxton (2010) has created several metrics. These metrics expresses one of the biases discussed in the previous section into an error. All these metrics are expressed as functions of the percentage error of the statistical and adjusted forecast (Eroglu & Croxton, 2010). The percentage error is simply the difference between the statistical/adjusted forecast and the actual demand divided by the actual demand. Based on this metric, feedback can be provided on whether he/she has improved the forecast during previous periods.

The metrics associated with these biases provide additional information which could also be used in feedback systems. Over the last few decades, different methods of sharing feedback have been examined, from which *"Outcome feedback"* is the most common and convenient approach (Lawrence et al., 2006). Such an approach just simply informs a planner on their performance throughout the last period. A more sophisticated approach is named *"Performance feedback"* and shares besides accuracy, information on biases which were associated with forecast from the past. Another type of feedback (i.e., *"Cognitive process feedback"*) includes providing information on a strategy which the planner has adopted throughout previous reviewing periods (Lawrence et al., 2006).

Chapter 3

Business understanding

In this chapter, all processes which are relevant to this project are explained in more detail. Since this project is focused on forecasting, this chapter is devoted to describing Ricoh's current forecasting practices. In the current situation, forecasts are created by the demand planning department who are responsible for managing and sharing accurate demand expectations. The aim of this chapter is to create an overall business understanding of the as-is situation.

3.1 System generated forecast

Ricoh Bergen op Zoom is responsible for managing most of the supply chain related tasks within the EMEA region. To ensure that all of these tasks can be executed efficiently, an organisation requires to have insight into what the future will bring. That is why most companies, including Ricoh, create forecasts on a frequent basis. Information originating from these forecasts are crucial in supporting an organisation in its (future) decision making process. Ricoh uses their demand forecasts to create a supply planning for all SKU's and the accuracy of these forecasts have a direct impact on Ricoh's inventory levels. Since manually forecasting every unique SKU is very difficult and foremost very inefficient, many companies are supported by some kind of forecasting software. These forecasting tools systematically derive statistical forecasts which are often primarily based on historical demand information. Depending on the software capabilities, the users preference and common practise, a forecast for a predefined time horizon is generated based on one of the available forecasting algorithm. At the moment, Ricoh's demand planning department is also supported by forecasting software.

Ricoh's forecasting system generates forecast output by closely observing demand history of each SKU separately. Time series data are constructed by aggregating these historical demand volumes across a time frequency (i.e., a monthly time frequency). For example, all individual demand throughout January are consolidated into one single time series value. This aggregation across time is based on the request date of an order and not on the actual date of delivery. The request date is set to be the date on which the demand was originally requested by the customer, while the date of delivery is the actual date on which the order was delivered to the customer. This difference is nowadays of additional importance as these could turn out to be two different dates due to all kind of different supply chain challenges. The system stores and observes a maximum of five years of demand history on every SKU. Resulting in at most 60 (months worth of) consolidated time series data per unique item. In Figure 3.1, a screenshot is provided showing an example of the time series data of a random SKU. On a daily basis the system updates the historical demand database and automatically sorts all new order data into the correct corresponding month bucket (based on the original request date).

Year	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Average	Total
History														
2017 - 2018													0	0
2018 - 2019					4	183	1492	2993	1619	1956	1952	2018	1527.125	12217
2019 - 2020	1680	2439	2049	2871	2532	1899	2427	2479	486	1007	1669	1623	1930.0833	23161
2020 - 2021	1362	2472	1872	2085	2054	1837	1949	2806	2146	2055	1921	1949	2042.3333	24508
2021 - 2022	1383	2657	2201	1666	2075	1612	2091	1714	1040	707	597	237	1498.3333	17980

Figure 3.1: Example of time series data

At the start of every new month, the forecasting software produces a 12-month forecast based on the historical time series data. After the system has removed any outliers from this data-set, it selects the forecast algorithm which best fits the time series data. This best fit selection procedure is also referred to as *"Best Fit Forecasting Tournament"* and it applies and evaluates multiple forecasting algorithms. Each of the algorithms are evaluated with the same performance metric, from which the best performing algorithm is selected. The performance metric which is used to evaluate all algorithms on is the Mean Absolute Percentage Error (MAPE). The algorithms which are included within the system's toolbox are listed below:

- Simple Moving Averages
- Linear Regression
- Simple Exponential Smoothing (SES)
- Double Exponential Smoothing (DES)
- Holts-Winter Additive (HWA)
- Holts-Winter Multiplicative (HWM)
- Croston's

Most of these algorithms require sufficient amounts of available time series data in order to come up with accurate predictions. In general it is often true that the less data is available, the worse statistical predictions get. For instance, an algorithm which incorporates a seasonality component (e.g., both Holts-Winter's algorithms) needs more data compared to an algorithm which only focuses on the most recent time series data (e.g., Moving averages). More specifically, the former often needs multiple seasonal cycles in order to correctly capture a time series seasonal behaviour. Recently launched items which cannot be linked to any predecessor model(s) generally require to be forecasted based on demand expectations. While items which have had significant amounts of demand volume during multiple years, are more likely to be accurately predicted with the aid of one of the forecasting algorithms. More detailed explanation on the above mentioned forecasting system, including its algorithms, will be provided in chapter 6.

3.2 Judgemental forecast

In today's world, future demands are impacted through many factors which cannot all be captured by statistical models. For example, demand volumes are generally impacted by promotional activities, new product introductions, competitor behaviour, price changes and many others (Fildes & Goodwin, 2007). While at the same time, several supply chain challenges also impact demand dynamics (e.g., material/container shortages, Covid-19 pandemic). That is why many companies follow a procedure in which statistical derived forecasts are carefully reviewed by a planner. The latter is referred to as one of the judgmental forecasting techniques and is also part of the forecasting process within Ricoh.

The system derived forecasts are reviewed by team members of the demand planning department. The first week of every new month is dedicated to reviewing the system generated forecasts. During this reviewing process a demand planner is responsible for manually checking the forecast predictions of the items within their portfolio. For each of those items, a planner has to decide whether or not he/she agrees with the system suggested 12-month predictions. If not, a planner must decide on the direction and the magnitude of the adjustment. These adjustments are often based on the earlier mentioned factors or just general basic (work) experience and knowledge and is also referred to as *"Contextual information"*. Such contextual information should ideally be incorporated into a forecasting system, but is very difficult since its impact cannot always be quantified/represented as data (J. A. Alvarado-Valencia & Barrero, 2014). For instance, the exact impact of a new promotional campaign on future demand is hard to quantify. More details on the execution of the judgmental forecasting process within Ricoh are shared in chapter 8.

3.3 Forecast process

After the system proposed forecast is reviewed and adjusted where necessary, it is then presented to several Ricoh representatives. In this meeting the proposed forecast is discussed to achieve a certain level of organization wide consensus between representatives of Ricoh's marketing, RESCM, head-office and sales departments. This consensus process is important as in this way multiple levels within the organization commit to the same demand expectations. This commitment will prevent forecast discussions at a later stage, as it was a shared decision. These demand expectations are officially submitted in the system and automatically shared in doing so.

As a final step, Ricoh examines the forecast accuracy of the submitted forecast. Ricoh calculates its forecast accuracy by comparing the forecast values to the actual demand values. More specifically, the submitted forecast values from N-2 are compared to the actual values of month N. More details on forecast accuracy figures are provided section 3.4.

All of the above mentioned processes are visualized with the aid of a process flow. This process flow illustrates Ricoh's general forecasting creation process and is depicted in Figure 3.2. The forecasting system uses input from aggregated historical demand data to statistically derive forecast figures. These system generated forecast figures are reviewed by a demand planner based on information which cannot be captured by any statistical models. Future demand could for instance be impacted through promotions, product introductions and tenders (i.e., internal dynamics) and/or via general market trends and a variety of supply chain challenges (i.e., external dynamics). A part of Ricoh's demand originates from their tender business, in which they opt to participate in large deals.

These large deals impact normal demand behaviour and therefore need to be carefully incorporated into the final submitted forecast. The forecast procedure as depicted in Figure 3.2 is repated on a monthly basis.

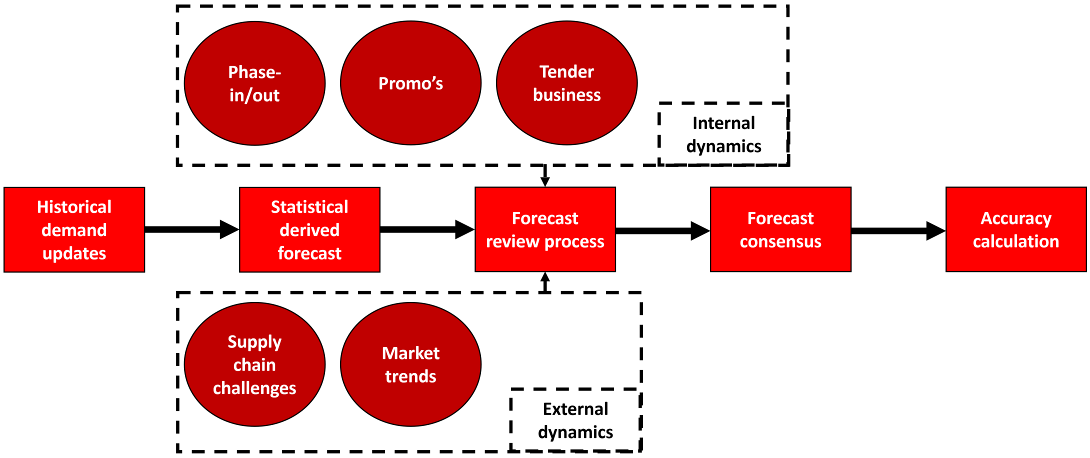


Figure 3.2: Process flow of high-level forecast creation and approval

3.4 Forecast accuracy

In order to get a complete understanding of the as-is situation, this section is devoted to describing the forecast performance of the currently used approach. As mentioned before, forecast performance is expressed as the difference between the submitted forecast values from month N-2 and the actual demand values of month N (See formula below). For example, the forecasted values of March submitted in January are compared with the actual March values. At the moment Ricoh only examines the accuracy of the final submitted forecast values and does not measure forecast accuracy of their system generated predictions specifically. This due to the fact that Ricoh has simply never recorded and/or stored the initial predictions made by the statistical models.

$$\text{Forecast accuracy at month N: } 100\% - \left| \frac{Actual_N - Forecast_{N-2}}{Actual_N} \right| \tag{3.1}$$

Forecast accuracy, within the context of this research is expressed in demand quantities and will therefore also be the focus of the remainder of this section. All accuracy figures mentioned in this section will only reflect on products included within the predefined product scope (i.e, machine SKU's and supply SKU's) during the 2016-2019 period (i.e., the scoped time-period). The next paragraph, section 3.5, will elaborate more on this specific project scope. Figure 3.3 shows two plots providing insight into the forecast accuracy during the 2016-2019 period. The left plot depicts the actual and forecasted demand quantities alongside its absolute deviation, while the right plot shows the forecast accuracy expressed as a percentage (see formula above).

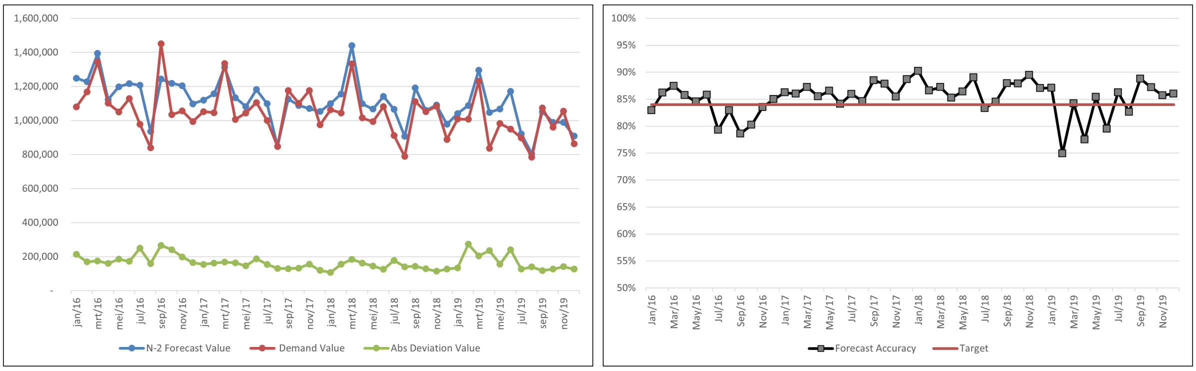


Figure 3.3: Forecast values and accuracy during the 2016-2019 period

During the 2016-2019 period, Ricoh averaged just over one million in demand volume quantities on a monthly basis. Moreover, the average forecast accuracy during this same period was 84%. The machine product type had a average forecast accuracy of 73%, while all supply SKU's were forecasted with a 87% average forecast accuracy. This significant deviation is no surprise since the predictability of supply SKU's is much greater compared to the predictability of machine SKU's. The historical demand patterns of the latter are much more erratic due to the fact that a machine SKU has a much shorter product life-cycle and is also much more often involved in large tender deals. For both product-types the demand data shows that July, August and December are often low volume months, while February, March and September are considered high volume months.

3.5 Project scoping

This final paragraph will elaborate more on the project. More specifically, details are shared on the product types that are included and the time period that is selected.

Ricoh's main business/assortment can be categorized into three different product types (i.e., Machines, Options and Supplies). These three product types contain all together 97.5% of the assortment for RESCM in terms of sales value. Within the machine range (48% of RESCM Sales) there is a wide assortment of printers, copiers, production printers, industrial printers and whiteboards. These machines are mostly sold in configurations together with options (16% of RESCM Sales). Ricoh offers both Ricoh and Third Party options to be installed on their machines. Options can best be explained with an example of buying a new car in which there are plenty of add-ons to pick from (e.g., parking sensors, heated seats). Within the context of Ricoh's business, options are for instance additional paper trays or special finishing option to make booklets. Finally, products like toners and cartridges are categorized as supplies (34% of RESCM sales). Such items are required by the machine to function and need to be replaced once there are depleted. Within the context of this research only the Machine and Supply product types are included. Options are excluded since their current method of forecasting does not follow the procedure as discussed in this chapter.

The plots shown in Figure 3.4 below depict forecast performance during the 2020-2021 period and clearly shows the impact of the Covid-19 pandemic on Ricoh's demand streams. In the first two months of the pandemic (March/April 2020) demand volumes drastically decreased in which forecast accuracy dropped to its lower-bound of 0%. Generally speaking, both the monthly average demand volumes (733,485 instead of over a million) as the average forecast accuracy rates (70% instead of 85%) drastically decreased compared to the period discussed in the previous section (i.e., 2016-2019). All of this clearly indicates that the 2020-2021 period does not represent normal predictable behaviour and that is why more fair and valid conclusions are ensured if this time-period is excluded. In other words, that is why this project is proceeded based on data originating from the 2016-2019 period.

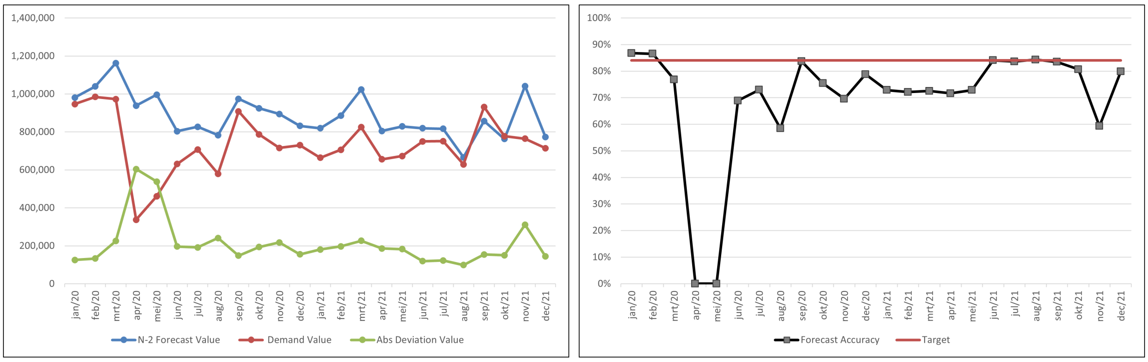


Figure 3.4: Forecast values and accuracy during the 2020-2021 period

Chapter 4

Hierarchical selection

In this chapter, the hierarchical structure of Ricoh will be constructed and discussed. This structure breaks-down Ricoh's business into various dimensions and hierarchical levels. The goal of this chapter is to get an insight into which hierarchical dimensions are of interest within this project.

4.1 Hierarchical structure

Each and every organization can break-down or aggregate its business with the aid of various dimensions and hierarchical levels. Supermarkets, for instance, often categorize their product assortment based on different product characteristics (e.g., beverages, dairy products, meats, fruits, vegetables). Such categorizations do not limit themselves to just only a product dimension, since dimensions separating customers and/or regions are also regularly used by companies.

The hierarchical structure which applies to Ricoh's business is depicted in Figure 4.1, and is constructed based on examples from literature, insights from experienced staff members and the structure of available order-data. Transactional data is analyzed to ensure that the hierarchical structure could also be supported by available data. The constructed hierarchical structure is multi-dimensional, and breaks-down Ricoh's business into three different dimensions (i.e., product, customer and region). Constructing this hierarchical structure is of importance, since the new forecasting system will enable Ricoh to calculate forecasts on different levels of aggregation. At all times, data needs to be (dis)aggregated on one class from each of the three dimensions. For instance, Rioch might decide to calculate forecast based on input data from a Product-Family, Sales Channel and OpCo Cluster level of aggregation. The levels of aggregation on which the system currently calculates statistical forecasts are also highlighted in Figure 4.1. In this moment in time, forecasts are calculated on the most detailed level of aggregation within the product dimension and the most aggregated hierarchical levels within the customer and region dimensions. In other words, each SKU is forecasted separately without differentiating these products on any customer or region characteristics.

Regardless of the selected settings, the organisation eventually requires to have forecast predictions on the highlighted levels of aggregation. This is the level of aggregation that several departments use for their decision making. These departments require forecast figures expressed as demand volumes per SKU without any dis-aggregation on the customer or region dimension. Meaning that if other settings are applied, output has to be dis-aggregated or aggregated to the originally used levels of aggregation. For instance, if the forecast is calculated on a product line level for each OpCo separately, then output has to first be aggregated on the region dimension and thereafter dis-aggregated to a SKU level within the product dimension.

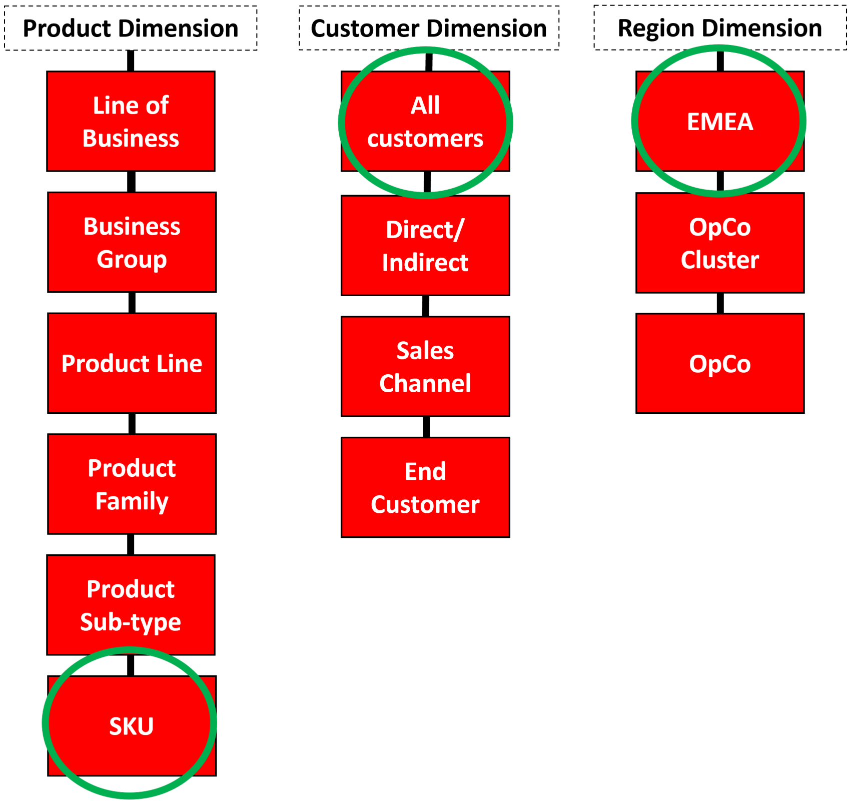


Figure 4.1: Ricoh's multi-dimensional hierarchical structure

Nonetheless it can still be of value to calculate forecast on a different hierarchical level than the one that is required by Ricoh. Hierarchical forecasting in general is a trade-off between having information of an individual time series (i.e., on an granulated level), and reducing noise of an individual time series data by aggregating them. There is no guarantee that the natural level of aggregation (i.e., the level of aggregation on which the forecast output is eventually required), also results in the most accurate statistical derived forecasts. In other words, there might be a different level of aggregation which improves forecast accuracy. The next three paragraphs will briefly explain each of the dimensions. More details on the data distribution behind the different classes and the number of clusters within each class are shared in the next chapter.

4.1.1 Product dimension

The product dimension clusters all items based on their product characteristics and consist out of the most classes (six in total). The most aggregated level in the product dimension is the *"Line of Business"* and only sorts products into either the commercial industrial printing (CIP) or the office printing (OF) line of business. The CIP line of business targets the high volume printing/copying markets (e.g., book publishers), while the office printing line of business is more aimed towards markets which need to print/copy from time to time (e.g., offices). The next class is the *"Business Group"* and is quite similar to the line of business. The CIP group only consist out of one cluster and therefore does not make any extra differentiation. The OP group is further separated by two clusters. The OP line of business differentiates between printers and multi-function printers (MFP). The latter is also referred to as copiers.

The third level of aggregation within the product dimension is the *"Product-Line"* which differentiates each *"Business Group"* into two clusters based on their ability to either copy/print in colour or in black & white. Each of these clusters are subsequently disaggregated into their corresponding *"Product Family"*. A product family represents a range of models that share similar features but are different in terms of processing speeds. The fifth class within the hierarchical product dimension is the *"Product Sub-type"* and has a different meaning for the machines and supply SKU's. More specifically, this class differentiates machine SKU's based on processing speeds (expressed in prints/copies per minute (ppm/cpm)) and supplies are clustered based on the type of supply (e.g., toners, cartridges and staple clusters). The final and sixth class (SKU) does not cluster any products together and therefore treats each SKU separately. A descriptive example showing the clusters to which a random example (i.e., IMC3000) belongs is depicted in Figure 4.2.

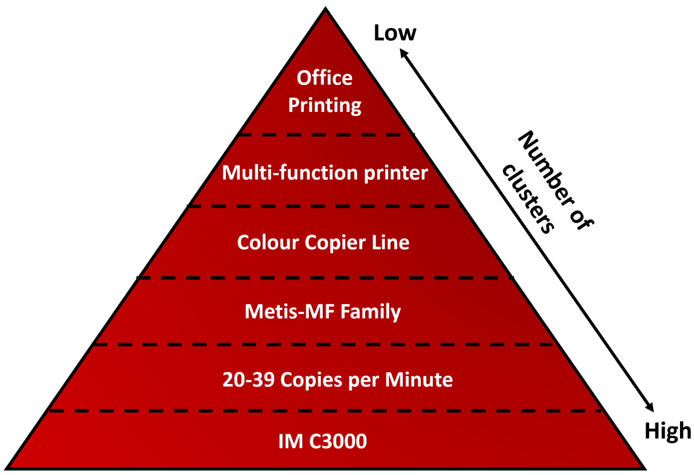


Figure 4.2: Descriptive example of the product dimension

4.1.2 Customer dimension

The customer dimension consist out of four hierarchical classes and clusters order lines based on its customer characteristics. The first class is very straightforward and consist out of only one cluster in which no order lines are differentiated (i.e., All customers). The second class describes whether the products are sold directly to an end customer or indirectly. Direct sales represents sales which are sold to customers who intent to use the products themselves, while indirect sales are sold to customers who's business it is to resell Ricoh products. Such companies also often offer other brands within their product portfolio. Finally, a third (smaller) cluster is dedicated to inter-company sales which represents all products which are sold within Ricoh to be used for marketing purposes (e.g., displayed within a showroom). The direct and indirect clusters are further disaggregated into several sales channels (i.e., the third hierarchical class). The inter-company cluster is split into sales that is done to the EMEA region (i.e., ICO Europe) and into inter-company sales that is done to regions other than the EMEA region (i.e. ICO World). All the direct sales is disaggregated based on the type and size of the customer (e.g., public, private, small, medium). The indirect sales are also differentiated based on the type of customer (e.g., retailer, dealer, distributor, wholesaler). In total there are seven sales channels associated to direct sales and Ricoh defined eight sales channels to differentiate all of their indirect sales. The most disaggregated hierarchical class is once again very straightforward and does differentiate each customer separately (i.e., End Customer).

4.1.3 Region dimension

The region dimension consist out of three hierarchical classes and clusters order lines based on their geographical sales characteristics. Ricoh Europe is active across the entire EMEA region and this region can be disaggregated into multiple smaller regions. The first "EMEA" class is similar to the customer dimension and consist out of only one cluster in which no disaggregated clusters are distinguished. The second class clusters multiple OpCo's together in predefined clusters based on their geographical location (see Table 4.1). These OpCo clusters separate Europe into four clusters, while the remaining OpCo's are either assigned to the Middle East/Africa cluster or to "Others". Ricoh International is the only OpCo assigned to others, since this specific OpCo represents multiple countries which cannot be simply assigned to just one cluster. These clusters are created based on domain knowledge and standardized/often used geographical clusters.

Table 4.1: All predefined OpCo-clusters

OpCo Cluster	**Western Europe**	**Northern Europe**	**Eastern Europe**	**Southern Europe**	**Middle East /Africa**	**Others**
1	Belgium	Denmark	Russia	Italy	South Africa	International
2	France	Finland	Hungary	Portugal	Middle East	
3	The Netherlands	Ireland	Czech Republic	Spain		
4	Germany	Norway	Poland	Turkey		
5	Luxembourg	UK				
6	Austria	Sweden				
7	Switzerland					

Finally, each of the OpCo's as summarized in Table 4.1 are clustered on their own (i.e., OpCo class). In total, 24 different OpCo's are included within this project.

4.2 Scoped hierarchical structure

The structure as depicted in Figure 4.1 illustrates Ricoh's business segmented in different dimensions and hierarchies. Although this structure is considered a complete representation, not all classes are seen as relevant with respect to the conducted research. Some levels of aggregation are perceived as irrelevant since their class either differentiates data by too much or too little. The filtering process conducted throughout this section is mostly based on domain knowledge in which Ricoh considers some levels of aggregation to be irrelevant. That is why by the end of this section, the initial original structure is narrowed down to a structure which will be examined and tested in the next few chapters.

For instance, clustering products within the *"Line of Business"* class is a level of aggregation which is seen as a cluster that differentiates too little. For clarification, calculating forecast on a *"Line of Business"* level requires all SKU's within the same *"Line of Business"* to be aggregated. Since this project only considers two lines of business, it is concluded that the mutual characteristics and demand patterns in between the SKU's of each *"Line of Business"* cluster are too different and therefore expected to result in inaccurate forecast output when disaggregated back down to a SKU level. The same reasoning is applied on why the *"Business Group"* class is excluded from the scoped hierarchical structure. As mentioned before, the business groups are only a slightly more detailed split and it is therefore also assumed to represent groups of SKU's that are too different.

Finally, it is decided to exclude one more class from the original structure. The most detailed hierarchical customer class (i.e., *"End Customer"*) is excluded since the organisation considers disaggregating data down to a customer level far too detailed. Even if there were any benefits on calculating forecast on this specific level, it will result in difficulties from a data perspective. Based on some first data insights it is concluded that most customers do not to have enough consistent/regular historical (order)data to enable the system to generate trustworthy statistical derived forecasts. The scoped structure which will be the foundation for the remainder of this study is shown in section 4.2. In total the structure provides 36 different scenarios between the three dimensions (i.e., 4x3x3=36). A scenario represents a combination of three levels of aggregation (i.e., one level of aggregation within each hierarchical dimension). The forecast performance of each of these scenarios are examined and discussed in chapter 7.

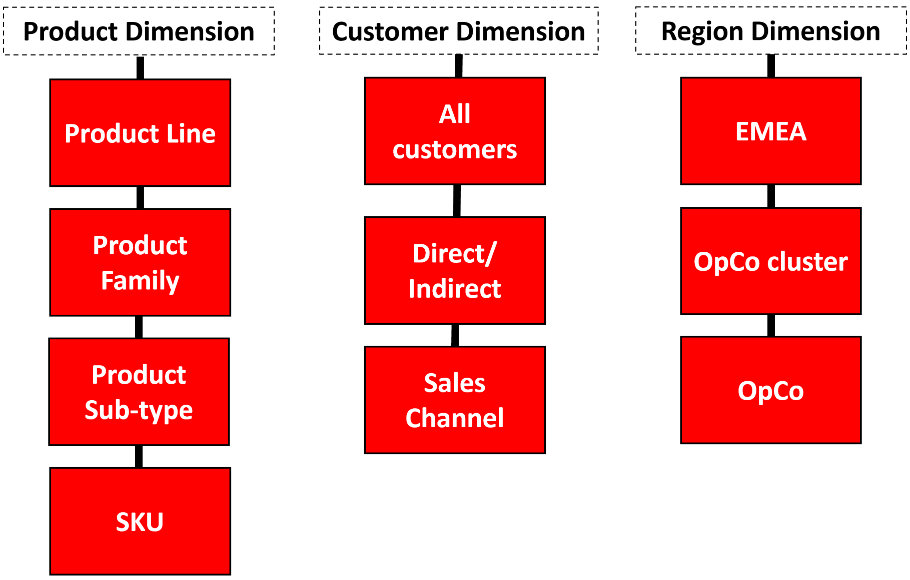


Figure 4.3: Scoped hierarchical structure

4.3 Hierarchical assumptions

As with almost every research project, assumptions and expectations are present before the actual research/analysis are conducted. Based on the defined scoped hierarchical structure as depicted in section 4.2, Ricoh already has shared some assumptions on which levels of aggregation they believe might work well and which they expect to damage a system's forecast accuracy. Ricoh firmly believes that when data is disaggregated to the most detailed levels on all three dimensions, it will not benefit their forecast performance. This is assumed since they consider the associated time series data on this specific scenario to be too random/erratic. Furthermore, they feel like that the product family class might have potential too improve forecast accuracy, while there are also very curious too find out about the forecast impact of disaggregating data based on the region dimension.

Based on the three dimensions that are considered it is assumed that a product level of aggregation might work well in combination with either a more detailed region or customer level of aggregation. It is argued that simultaneously disaggregating demand data into a more detailed customer and region level of aggregation will not benefit forecast accuracy.

Chapter 5

Data gathering

In this chapter, data associated with the scoped hierarchical structure will be gathered and analyzed. Each of the three hierarchical dimensions will be discussed with the aid of some brief data analysis. This chapter will therefore help to create a better understanding of how data is distributed throughout the different dimensions.

5.1 Gathered data

In order to test all of these 36 different forecast scenarios, data needs to be gathered. Data is gathered specifically for machines and supplies (during 2016-2019) since these are the product-types that are considered within this project (see **??** and section 3.5). In addition to collecting the data, this section also aims at providing a brief data insight into each dimension and its corresponding classes. Information is shared on the clusters within each class and its associated data distribution. Furthermore, based on these data insights some first data cleaning steps are conducted. Data is either cleaned due to being flagged as an error or considered as irrelevant. The remainder of this section will discuss the data associated with each dimension. All data shown in the tables/figures throughout this chapter are based on the demand volumes throughout four years (i.e., 2016-2019).

Data is gathered on a detailed transactional order level and includes information on all of the classes distinguished in section 4.2. The closed order bank in which all historical demand volumes are recorded is used to gather the required data. Since the data dimensions are quite enormous (especially for supplies), a direct connection between the database (i.e., SQL) and Python is established. Information on the obtained data-set including its data dimensions in terms of order lines, order volumes and number of unique SKU's is shown in Table 5.1

Table 5.1: Aggregated statistics of obtained data-set

Product-type	**Order-lines**	**Order volume**	**Unique number of SKU's**
Machine	1,182,318	2,197,177	330
Supply	25,499,951	48,149,198	706
Total	26,682,269	50,346,375	1036

Throughout four years, Ricoh processed approximately 27 million orders representing a total sales volume of over 50 million different machines and supplies. 330 different types of machines were sold through 1.1 million orders and collectively resulted in a total sales volume of almost 2.2 million. Furthermore, over 700 different supplies were responsible for over 25 million orders representing a total sales volume of over 48 million. More detailed descriptive statistics are provided in Table 5.2 and Table 5.3.

Table 5.2: Descriptive statistics of demand volume per order line

	N	**Mean**	**St. Dev.**	**Min**	**Median**	**Max**	**Mode**
Machines	1,182,318	1.86	15.30	1	1	2,860	1
Supplies	25,499,951	1.89	12.45	1	1	7,600	1

Analyzing these descriptive statistics, it can be concluded that, except for the number of order lines and the maximum order value, there is not much differences between the order volumes of machines and supplies. The average order-size for both product-types is close to 2 with a standard deviation between 12 and 16. This relatively high standard deviation/mean ratio indicates that there might be some outliers within the data. For machines, the relatively high maximum order volume in combination with the high standard deviation are most probably related to Ricoh's tender business. Another reason explaining this high standard deviation is the different types of sales (i.e., direct versus indirect). Direct sales are, except for tenders, often low volume orders since its bought by customer who intend to use the products, while indirect sales are more likely to be bought in bigger bulks since the products are bought by some type of reseller. The latter is assumed to be the main cause explaining the high standard deviation and maximum demand volume for supplies. Forecast algorithms can be very sensitive to such outliers and since these volumes are not always part of usual business, they need to treated carefully. This is done by outlier correcting the input data on which more details are shared in subsection 6.1.3. Furthermore, the system, as described in section 3.1, does not consider historical demand data on an individual order line level, but instead transforms these into monthly aggregated time series data. To get a better understanding of the time series data as used as input by the system, Table 5.3 summarizes descriptive statistics from a monthly aggregated order volume perspective.

Table 5.3: Descriptive statistics of demand volume per month

	N	**Mean**	**St. Dev.**	**Min**	**Median**	**Max**
Machines	48	45,775	12,230	26,370	42,574	74,068
Supplies	48	1,003,108	134,376	751,231	1,003,341	1,372,304

The average demand volumes on a monthly basis are as expected significantly higher for supplies in comparison to machines (1,003,108 versus 45,775). This makes perfect sense as a machine is considered a long-term investment which requires new supplies every now and then to be able to properly function. Furthermore, it is observed that the min/max statistic values might indicate a high/low pattern between recurring months. To be able to confirm these findings, a stacked bar chart is created and depicted in Figure 5.1.

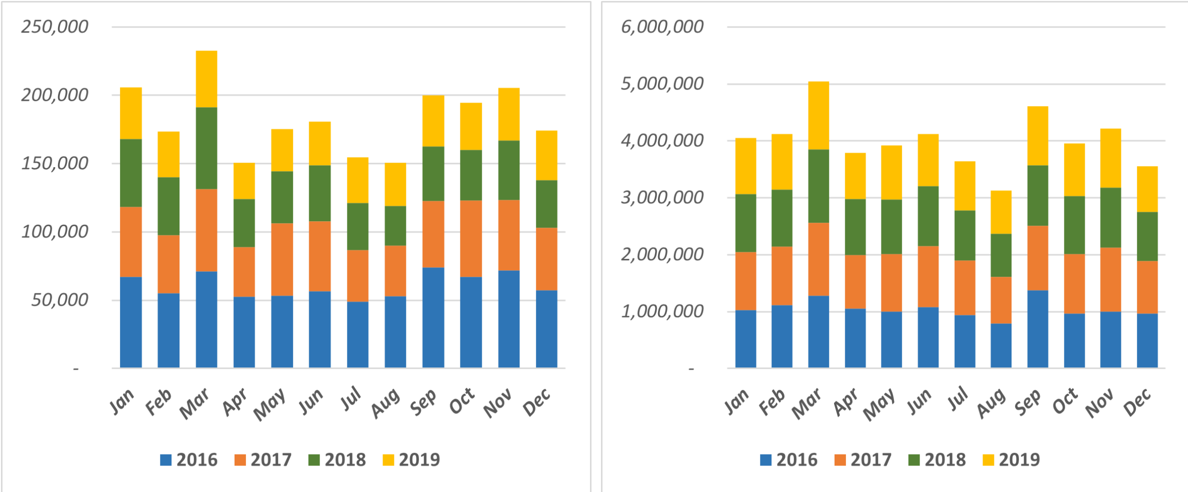


Figure 5.1: Stacked bar charts of monthly demand volumes for machines and supplies

Based on these plots, it is confirmed that the months March, September and November are considered peak months, while April, July and August are found to be low volume months. These peaks are the result of marketing pushing demand in the final month before the ending of each fiscal year (March), while in September most organizations have had their holidays and they start ordering again. The low volume months can therefore be explained as being the first month after the fiscal year ending (April) and the summer holidays (July and August). The remainder of this chapter will provide more insights into the classes per dimension discussed in section 4.2 and depicted in Figure 4.3. Once again, all provided data insights shown in the next three sections reflect demand volumes throughout a total of four years (i.e., 2016-2019).

5.1.1 Data insights into product dimension

After the initial hierarchical structure was scoped down in the previous chapter, a total of four different hierarchical classes are left in scope for the product dimension specifically. The highest class which is kept in scope for Ricoh's product dimension are the different product lines. Some aggregated statistics shown in Table 5.4 provide an insight into the data distribution of the eight different product lines. These eight product lines are considered the most important ones since they represent most of the supply and machine demand within Ricoh. A product-line is included whenever it is considered a relevant product-line by all stakeholders (i.e., the demand planning and business excellence department). As a result, all SKU's associated to one of the eight selected product-lines are included. Supplies only consist out of seven product-lines, simply because the supplies required for the *"Colour Printer"* and *"Col A4 MFP Printer Based"* machines are the same ones and therefore all included within just one product-line (i.e., *"Colour Printer"*).

The *"Colour Copier"* product-line is by far the biggest cluster in terms of order lines and volume for both machines and supplies. With respect to the total demand, this single product-line represents 38.5% of the machine volume and 56.6% of the supplies volume. The smallest clusters in terms of order lines and volumes are both production printer lines (i.e., *"B/W and Colour Production Printer"*). Such machines are used in businesses which require to continuously run large print jobs (e.g., newspapers and book publishers) and are much more expensive compared to other machines. This explains why it has the lowest machine volumes, but still has a significant supply volume.

Table 5.4: Aggregated statistics of demand volumes per product-line

Product-Lines	**Machines**			**Supplies**		
	Order lines	*Order volume*	*Unique SKUs*	*Order lines*	Order volume	Unique SKUs
COLOUR COPIER	618,216	846,068	88	16,165,696	27,276,441	201
B/W COPIER	243,987	435,849	84	3,429,204	9,506,492	89
B/W PRINTER	198,376	480,534	44	1,142,113	2,049,567	58
COLOUR PRINTER	50,861	81,490	22	1,917,877	3,387,999	211
B/W A4 MFP PRINTER BASED	28,734	276,372	41	1,066,222	2,241,984	33
COL A4 MFP PRINTER BASED	28,162	62,596	10			
COLOUR PRODUCTION PRINTER	10,631	10,634	26	1,537,584	2,997,695	76
B/W PRODUCTION PRINTER	3,351	3,634	15	241,255	689,020	38

The next class within the product dimension are the product families. While the product-family class is mainly applicable to machines, Ricoh was also interested in applying this class to supplies. Since the original data structure did not link supplies to any product-families, a more manual set-up was required. Each supply was linked to a product-family based on the product-family of the machine it was linked to. For machines specifically, these product-families also make an distinction between different product-lines. For example, the Metis family is separated into a printer, multi-function and copier family (i.e., Metis-P/MF and C). However, since supplies are often applicable to an entire family, regardless of the product-line, this same distinction could not be made for supplies specifically. Therefore a supply is linked to for instance the entire Metis family. Furthermore, supplies are not always only sold in combination with one specific machine family and if this is the case it is placed into the cluster "Multiple". This could for example be generic used supplies such as staples. Finally, there are also many supplies which could not be linked to any machine families at all and these are clustered as "None". Eventually, all of the included SKU's can be linked to one of the approximately 50 different product family clusters. The top five biggest product families for both product types in terms of demand volumes are shown in Table 5.5.

Table 5.5: Top five biggest product-family clusters per product-type

Machine Product-family	**Demand volume**	**Unique SKU's**	***Supply Product-family***	**Demand volume**	**Unique SKU's**
Metis-C	468,811	45	*Metis*	14,561,912	35
Griffin-C	230,737	15	*None*	10,813,847	349
Gimlet-P	228,967	*5*	*Griffin*	4,328,859	20
Metis-MF	122,232	17	*Multiple*	2,984,249	51
Stella-C	113,628	10	*Athena*	2,172,760	13
Others	1,032,802	238	*Others*	13,287,571	238

From the table above it is clear that the Metis family in general is the biggest family for both product types. The Metis-C family represents over 21% of the total machine volume, while the overall Metis family (used for all Metis machine product lines) represent a total of over 30% of the total supply demand volumes. Moreover it is observed that the top five biggest machine product families represent 53% of all total volumes. This indicates that there are many product families which have had relatively low demand volumes throughout four years. Unfortunately, there are many supply SKU's, including a large portion of the total volume (i.e., almost 11 million in demand volume, representing 23% of all total supply demand), which cannot be linked to any product family specifically (indicated as 'None' in the right side of the table). This issue might make this cluster with respect to supplies result in less accurate forecasting results. Furthermore, the number of SKU's within each family ranges from just a single or a couple to dozens of items.

The third level within the product dimension is the product sub-type class. This class, as mentioned in subsection 4.1.1, has different clusters for machines and supplies. For the machine product-type it disaggregates the different product-lines as defined in Table 5.4 into different processing speeds. Table 5.6 shows the demand volumes for all of the different processing speeds per product-line. The two production printer lines are clustered within the higher processing speeds (i.e., 60+).

Table 5.6: Product sub-type clusters for Machines

Copiers	**Colour**	**B/W**	**Printers**	**Colour**	**B/W**	**MFP A4**	**Colour**	**B/W**
1-19 CPM	28,747	16	*10-19 PPM*	26,893	4,845	*10-19 PPM*	22,101	4
20-39 CPM	630,516	332,396	*20-29 PPM*	164,875	43,995	*20-29 PPM*	199,389	58,709
40-59 CPM	168,618	*83,932*	*30-39 PPM*	72,599	*10,770*	*30-39 PPM*	31,272	3,883
60-89 CPM	23,500	15,286	*40-59 PPM*	207,308	21,680	*40-59 PPM*	23,643	
90+ CPM	5,321	7,820	*60+PPM*	8,859	200			

From this table above it is observed that there are different ranges of processing speeds for the copiers and printers. In general, it seems that the middle-ranged processing speeds (i.e., 20-39 CPM and 20-29 PPM) are the most popular ones in terms of demand volumes. More specifically, the colour copier SKU's which have a CPM in the range of 20-39 are found to be the largest cluster and represent almost 29% of the total demand volume.

For the supply product-type, this class clusters each SKU based on the type of supply. In total there are seven different types of supplies defined within the obtained data structure. The data volume including its corresponding number of unique SKU's per cluster is shown in Table 5.7. This table shows that the sub-type "toners" is in terms of both volume and number of SKU's by far the biggest cluster within this class. This is no surprise since in general toners are the main supply required to be able to print/copy.

Table 5.7: Product sub-type clusters for Supplies

Product Sub-type	**Demand volume**	**Unique SKU's**	**Product Sub-type**	**Demand volume**	**Unique SKU's**
Toner	44,311,943	468	Staple	606,417	49
Other supply	1.437.885	78	Service kit	96,887	42
AIO	947,859	31	Gel cartridge	2,408	8
PCU	745,799	30			

The final and fourth class (i.e., SKU) cluster each SKU separately resulting in a total of 330 and 706 clusters created for machines and supplies specifically. Each SKU included within the initial obtained data-set has had recorded at least one demand volume during the considered time-period (2016-2019).

Table 5.8: Descriptive statistics of demand volume per SKU

	N	**Mean**	**Std.dev**	**Min**	**Median**	**Max**
Machine	330	6,658	13,522	1	1,906	126,077
Supplies	706	68,200	199,530	1	3,118	1,967280

Table 5.8 shows that the SKU that has generated the most machine demand equalled almost 6% of all total machine volume (i.e., 126,077), while the supply SKU with the highest recorded demand volume (i.e., 1.97 million) represents just over 4% of all supply demand volume. Moreover, this table indicates that there are many SKU's which have only recorded relatively low amounts of demand volumes throughout four years. This assumption is checked within the data, and it was found that 23% of all machines represent 80% of the total demand, and 12% of all supplies represent 80% of the total volumes. This clearly shows significant differences among the initially included SKU's and indicates that some SKU's probably have a continuous demand flow, while others only recorded sales during parts of the four years. Generating forecasts with the aid of statistical methods require SKU's to have a sufficient number of historical data points. That is why in subsection 6.1.1, criteria will be formulated which ensure that only such SKU's are considered.

Finally, as discussed in the introduction of this chapter, the data linked to this dimension also required some data cleaning. This due to the fact that some of the SKU's were linked to multiple product lines or product sub-types. These are identified as errors, since in reality they could only belong to a single cluster. All of these duplication errors are manually fixed based on domain knowledge.

5.1.2 Data insights into customer dimension

The hierarchical dimension focused on the customer aspect, is scoped down to a total of three different classes. No additional data insights are provided for the highest hierarchical class (i.e., All customers), since this class consist out of only one cluster.

The second class (i.e., Direct/Indirect), however, consist out of a total three different clusters and some data insights are shared in Table 5.9. The inter-company cluster only represents a low demand volume in comparison to the demand volumes of both the direct and indirect sales types. This is as expected, since this type of sales only consist out of demand required for marketing purposes. Furthermore it is observed that the direct and indirect sales types have comparable sales volumes in which the main difference lays within the average demand volumes per order. More specifically, direct demand most often only consist out of one item, while indirect demand on average consist out of a couple of items. This difference can once again be best explained that a direct customer buys a Ricoh product with the intention to use it, while an indirect customer buys a couple of machine/supplies with the intention to resell them.

Table 5.9: Data insight into the Direct/Indirect customer dimension class

	Machine		*Supplies*	
	Demand volume	**Mean volume per order**	**Demand volume**	**Mean volume per order**
Indirect	1,021,824	3.4	21,067,064	4.3
Direct	828,308	1	23,167,972	1.2
Inter-company	347,045	6.3	3,914,162	3.5

As mentioned in subsection 4.1.2, the third and final class provides insight into the different sales channels which Ricoh utilizes in order to sell their products. The public sales channel within all of the direct demand is found to be the largest cluster in terms of demand volumes (i.e., a demand volume of almost 5,5 million representing 24% of total direct demand). By far the biggest cluster with respect to all of the indirect demand are sales that are done through the dealer sales channel (i.e., a demand volume of over 13 million representing 50% of total indirect demand). Finally the Inter-company sales channel separates the Europe and World channel in which the Europe channel is the only channel which is relevant in terms of demand volume (i.e., a total of 3 million representing 99% of all inter-company demand).

5.1.3 Data insights into region dimension

The region dimension also consist out of three hierarchical classes and clusters order lines based on their geographical sales characteristics. The first "EMEA" class is similar to the customer dimension and consist out of only one cluster in which no disaggregated clusters regions are distinguished.

The second class clusters multiple OpCo's together in predefined clusters based on their geographical location (see Table 4.1). The data distribution in terms of demand volume for each OpCo cluster is summarised in Table 5.10. Somewhat the same data distribution can be distinguished for both machines and supplies. Western Europe is by far the biggest cluster and also consist out of the most OpCo's. Furthermore, it can be concluded that the cluster "others" also represents a significant part of the total volume, despite the fact that it only includes one OpCo (i.e., Ricoh international). Moreover, the data shows that the Middle East/Africa region is by far the smallest cluster.

Table 5.10: Data distribution of demand volume per OpCo cluster

	Machines		*Supplies*	
OpCo_Cluster	**Demand volumes**	**% of total**	**Demand volumes**	**% of total**
Western Europe	841,234	38.29%	20,790,667	43.18%
Southern Europe	426,710	19.42%	9,128,114	18.96%
Northern Europe	298,373	13.58%	9,869,616	20.50%
Eastern Europe	321,807	14.65%	2,593,644	5.39%
Others	222,311	10.12%	4,499,592	9.35%
Middle East/Africa	86,742	3.95%	1,267,565	2.63%

Finally, each of the OpCo's as summarized in Table 4.1 are clustered on their own. In total, 24 different OpCo's are included within this project. The top five biggest OpCo's for machines, put in chronological order, are Germany, France, Russia, International and Italy which all together represent 57% of the total volume. For supplies, the top five biggest OpCo's, in chronological order are United Kingdom, France, Germany, Italy and International and all together represent 61% of the total volume.

Finally, it was found in the data that some demand data was included which was irrelevant. This was demand that was sold to so-called satellites. Since this project and more specifically this dimension is focused on OpCo's it is decided to remove these demand volumes from the data-set. This did not have a major impact on the data.

Chapter 6

Forecast model

In order to examine the forecast performance of all 36 hierarchical combinations, a test environment needs to be created. To ensure that all findings and recommendations are future-proof the decision is made to test these levels of aggregation based on the dynamics of the new forecasting system. This new forecasting system unfortunately does not offer the option to test all these different levels of aggregation, and therefore it is decided to recreate this system with the aid of Python. This chapter will aim at identifying and describing the technical details of the new forecasting system which will subsequently be replicated with a validated Python script.

6.1 Forecast model details

This section is devoted to creating a full understanding of the system's forecasting dynamics. Since it is a new system, not many details are known upfront and realizing a full understanding will thus also improve Ricoh's insights into this system. This improved insight is also considered an indirect benefit as a result from the conducted research. The general dynamics of the new system are visualized with the aid of a process flow depicted in Figure 6.1. The remainder of this section will provide more details on each of these stages separately.

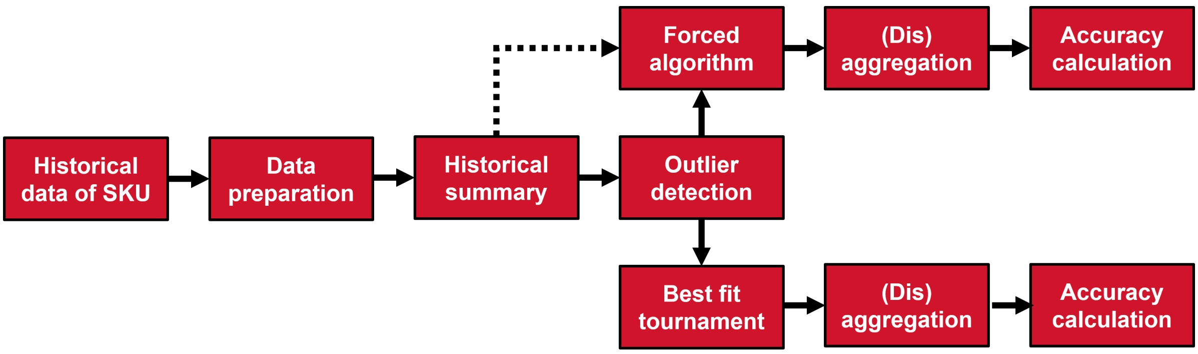


Figure 6.1: The system's forecast dynamics illustrated with a process flow

6.1.1 Data preparation

The first stage (i.e., "Historical data of SKU") of the system is to collect all required data and is already discussed in the previous chapters. The next stage is focused on several important data preparation tasks in which the collected data is made ready to be used in later stages.

One of the limitations of the current system is that it is not able to incorporate any predecessor-successor relations. More specifically, when the system generates forecast of a SKU it does not use any historical demand volumes of any predecessor models and therefore only considers the SKU its own historical demand volumes. This limitation is solved within the new system, since it includes a possibility to link historical data of predecessor(s) to its successor(s). In doing so, the system is able to apply algorithms and calculate predictions as the historical background data is more likely to be sufficient. The new system requires planners to decide on the weights which represent to what extent the demand history of the predecessor is copied to its successor. This is of importance as occasionally an item is succeeded by more than one item. In such situations a planner, supported by marketing, has to decide in which proportions the demand needs to be distributed among the successors. More specifically, marketing will share their expectations to what extent each successor will gradually take over the demand of its predecessor. Whenever a human planner does not interfere, the system by default equally distributes historical demand over the number of successors. In the context of this project, only the machine product-type is impacted by this data transformation process. This due to the fact that multiple generations of machine's use exactly the same supplies and if a new supply is introduced it is considered a completely new one. During the 2016-2019 time-period a total of 181 product/machine introductions took place from which 154 introductions apply to the machines within the data-scope. All these machines were succeeded by only one successor at all times and it successor was occasionally replacing more than one predecessor machine model. As a result, several SKU's are grouped as successors of others and in doing so the data-set now consist out of 176 machines instead of 330 machine SKU's.

One of the project specific preparation steps is to formulate inclusion/exclusion criteria which represents the data requirements from a SKU perspective. As briefly mentioned in the last paragraphs of both section 3.1 and subsection 4.1.1, system derived forecasts requires to have sufficient amounts of historical data in order to accurately predict future demand. More specifically, demand data needs to be recorded over a long enough period to enable a system to statistically derive a forecast. The necessary number of historical data-points depends on the forecasting algorithms, as some algorithms require more data than others. In order to decide on these inclusion criteria, an analysis is conducted which is summarized in Table 6.1 below.

The first criteria is mandatory and is the requirement that each SKU needs to have sales data recorded during the year 2019. This is crucial since eventually all comparisons and conclusions are based on the forecast performance during this particular year. All items (i.e., 105 machines and 439 supplies) which have had sales after the year 2019 are therefore initially included. Thereafter, an additional criteria is formulated and is focused on the length of the time series that each SKU requires to have. Multiple scenarios are tested in order to see how many SKU's are included per time period/window criteria. Based on the results of this analysis showed in Table 6.1, it is decided to continue with items which have had their first demand data recorded before the year 2017. In doing so, approximately 70% of all possible SKU's are still included, while the system still has at least two years worth of data to derive statistical forecasts from.

Table 6.1: Inclusion analysis from a SKU perspective

	All items	**>2019**	**<2018**	**<2017**	**<2016**
Number of Machine SKU's	176	105	84	76	70
% of Machine SKU's included	-	100.0%	80.0%	72.4%	66.7%
Number of Supply SKU's	706	439	326	296	263
% of Supply SKU's included	-	100.0%	74.3%	67.4%	59.9%

The last and final data preparation step involves splitting up the data into a train set and a test set. The train set consist out of all historical data before the year 2019 and is used by the recreated forecast model to fit an algorithm which subsequently outputs twelve forecast predictions for the year 2019. The test set consist out of the data from 2019 and is not considered by the model since it is solely used to calculate forecast accuracy. Eventually, all the forecasting predictions of the different levels of aggregation are compared and evaluated based on the test set.

6.1.2 Historical summary

Another stage characterising the new system is the so-called "Historical summary", which summarizes information/statistics of the historical demand time series. Statistics which are gathered are the Coefficient of Variation (CV) (1), Recent periods with no data (2), Zero after nonzero (3), Average (4), Total quantity (5), Minimum quantity (6), Maximum quantity (7), Total buckets (8) and Buckets with zero (9). The CV and the zero after nonzero statistics will be explained in a bit more detail, while it is assumed that the remaining statistics are self-explanatory. The CV, also referred to as relative standard deviation (RSD) is a standardized statistical measure which defines the ratio of the standard deviation to its mean. Within the context of this research, the CV statistics indicates to what extent the time series data is erratic. The zero after nonzero statistic counts the number of times a gap occurred in the time series data. A gap is observed each time a nonzero demand bucket is followed by a number of zero demand bucket(s).

The aim of this stage is to already decide whether a certain algorithm is most appropriate based on the characteristics of the time series. It either states that demand data is adequate and a Best-fit tournament is applied or it forces the time series to be forecasted through a specific algorithm. For example, when the time series data of an item is considered intermittent, an algorithm widely used for intermittent demand forecasting will be applied (Croston's method). From the statistics the system cam conclude the following.

- If there are less than seven monthly buckets with data, SES or DES must be applied
- If the zero after nonzero value is higher than two, Croston's must be applied
- If the most recent twelve months do not have data, use a 2-month moving average

The latter conclusion ensures that the forecast prediction equals zero as there was not a single demand to be recorded throughout the last year. Furthermore, the statements are in the same order as how they should be interpreted. Meaning that in the situation that a time series has less than seven buckets with data, but also exceeds the zero after nonzero threshold value, the system will force Croston's to be applied. Moreover, if this time series also did not have any data in the last year, a 2-month moving average will be applied instead of Croston's. Whenever none of this three threshold values are exceeded, the data is considered adequate, and the time series will use a Best-fit tournament to determine which algorithm is most appropriate (See subsection 6.1.4).

6.1.3 Outlier detection

Before actual predictions can be calculated, time series first have to be outlier corrected. Similar to most forecasting software, this system also has an automated outlier detection incorporated within their forecasting process.

The procedure of the new system is build on the concept of calculating an upper-limit and a lower-limit. Both limits are calculated through the average +/- three times the standard deviation. The average and the standard deviation are calculated over a maximum of the most recent 36 months (which corresponds with a total of three years worth of historical demand data). If a monthly demand bucket exceeds either the lower or upper-limit, the demand is proportionally adjusted to the limit is has exceeded. However, the outliers which are flagged and adjusted by the system, can still be overruled by a human decision maker. This is also part of the judgmental forecasting process discussed in chapter 8. Moreover, a human decision maker is able to manually flag and adjust or erase any specific order(s) within any of the monthly demand buckets. In the context of Ricoh, those are for instance orders which are included within the tender business. In other words, the system corrects outliers on monthly aggregated data, while an human planner can outlier correct on a more detailed level (i.e., individual orders) and also overrule any decisions made by the system.

One last important detail that needs to be shared is that it is decided that this outlier correction is always conducted on the hierarchical scenario on which the input data is aggregated. For instance, if the data is aggregated on a product family level, the outlier procedure is also conducted on the input data aggregated per product family. While if a scenario requires the data to be aggregated on a SKU level, the outlier procedure will also be executed on the data related to each of the SKU's. In general, it is observed that the more aggregated the input data is, the less outliers are systematically detected and adjusted for. This can be explained since highly aggregated data is often quite stable, while more individual disaggregated data is more likely to be erratic and therefore more sensitive to outliers.

6.1.4 Best fit tournament

All cleansed time series data, classified as adequate will participate in the best-fit tournament. This tournament includes testing all algorithms listed in section 3.1, except for the Croston's method. The Croston's method will only be used in situations in which the time series data is classified as intermittent. Similar to the outlier detection, the algorithms observes and considers a maximum time-span of three years (36 data-points/months at most). Eventually each algorithm is executed on the available data and its predictions are evaluated through the MAPE. The MAPE value is a prediction accuracy measure often used to evaluate performance of forecasting techniques (Armstrong, 2001).

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{6.1}$$

The MAPE takes the mean of the absolute deviations between the actual and predicted values divided by the actual demand values (See formula above). To ensure a full understanding of the best-fit selection, a fictitious numerical example is given in Table 6.2 below. For the sake of simplicity, self-explanatory simple moving average techniques are used within the example as the aim of this example is not too explain any algorithms, but to illustrate the best-fit selection procedure. Exact details on each single algorithms are shared in subsection 6.1.5.

Table 6.2: Best-fit numerical tournament example

Time period	**Actuals**	**Prediction**	**Absolute deviation**	**MAPE**	**Prediction**	**Absolute deviation**	**MAPE**
1	*11*	*2-Month Moving Average*			*3-Month Moving Average*		
2	*13*						
3	*16*	12	4	25.0%			
4	*11*	14.5	3.5	31.8%	13.3	2.3	21.2%
5	*12*	13.5	1.5	12.5%	13.3	1.3	11.1%
6	*12*	11.5	0.5	4.2%	13.0	1.0	8.3%
7	*8*	12	4	50.0%	11.7	3.7	45.8%
8	*9*	10	1	11.1%	10.7	1.7	18.5%
9	*12*	8.5	3.5	29.2%	9.7	2.3	19.4%
10	*13*	10.5	2.5	19.2%	9.7	2.3	19.4%
			Average MAPE	**22.9%**		**Average MAPE**	**21.4%**

In this example, an item is predicted through a 2-month and a 3-month moving average. Over the last ten periods this item has recorded sales volume to some extent for which both algorithms have calculated predictions. Given those predictions, an average MAPE value is computed which is slightly in favour of the 3-month average technique. Meaning that, in a situation in which only those two algorithms are tested, the 3-month moving average will be used in predicting future unknown demand patterns. To conclude, the most appropriate algorithm which has the lowest overall MAPE score will be used in predicting future demand. More specifically, this algorithm will compute twelve predictions corresponding with a forecast of a year.

6.1.5 Forecast algorithms

This section is devoted to explaining each of the forecasting algorithms in a bit more detail. The formulas associated with the algorithms and its parameters are discussed.

Simple Moving Averages

The moving average is one of the more intuitive forecasting techniques and derives a statistical forecast by taking the average of the actual historical demand from a specified number of prior periods. The new forecasting system evaluates the performance of eight different moving averages in the range of 4-11. The moving average final forecast value is calculated by taking the sum of all actuals (y) within the interval range and dividing it by the interval range (k), up to the most recent historical actual (N).

$$Forecast = \frac{y_{N-k+1} + ... + y_N}{k} \tag{6.2}$$

The twelve forecast predictions are all constant and are equal to the result of the formula formulated above. For example, the twelve final forecast values (for a 2-month moving average) in the numerical example shown in Table 6.2 is equal to 12.5 ($\frac{12+13}{2}$). Since all considered historical data-points are of equal weight, this algorithm is most suitable for stable demand.

Linear Regression

The linear regression forecasting model has traditionally been a commonly used technique. Such a model performs a linear regression on a set of historical quantities to determine a trend line that provides statistical forecast predictions. In other words, a regression line is created for the x (time interval values) and y (historical quantities) variables.

$$Forecast\ y = a + bx \tag{6.3}$$

The "a" value is associated with the y-intercept of the regression line, while the "b" value corresponds with the slope value of this line. Based on these calculated constants, a forecast is created by filling in the required x-values. In the situation of Ricoh, in which the system considers 36 historical months, these values are in the range of 37-48 (i.e., resulting in twelve (y) prediction values). Moreover, linear regression is best used for demand that does not include any seasonal component.

Simple Exponential Smoothing

The SES model derives statistical forecast values by combining fitted values with the latest historical demands. This technique considers all available historical data, but gradually decreases its significance to aging data points. A smoothing constant, alpha (α) determines this significance. A larger alpha puts more emphasis on more recent actuals, while a lower alpha gives more weight to more historical actuals.

$$FittedValue_{k+1} = \alpha y_k + (1-\alpha)FittedValue_k \tag{6.4}$$

The equation above determines the fitted values for each period interval and it applies the smoothing constant on both the previous fitted values as the previous actual. The initialization of this equation and also the optimization of the smoothing constant are discussed after all algorithms are briefly explained. The forecast value is a single constant for all twelve predictions and is equal to the final fitted value. This forecasting technique is once again most appropriate in situations in which demand is fairly stable and thus does not show any sings of a trend or seasonality component.

Double Exponential Smoothing

In addition to the smoothed value (F) used in the SES model, the DES model also considers a trend value (T). Both values have a smoothing constant, alpha (α) and gamma (γ), which are values between zero and one. This technique is often used for demand that is considered linear, but includes a upward/downward trend component over time.

$$F_{i+1} = \alpha y_{i+1} + (1-\alpha)(F_i + T_i) \tag{6.5}$$

$$Ti + 1 = \gamma(F_{i+1} - F_i) + (1-\gamma)T_i \tag{6.6}$$

The results from both equations are used to derive a statistical forecast. The fitted values and its final forecast values are calculated based on the formulas below. Statistical forecast values are calculated based on the final smoothing and trend values from which the latter is multiplied with a factor m.

$$FittedValue_i = F_{i-1} + T_{i-1} \tag{6.7}$$

$$Forecast_{N+m} = F_N + mT_N \tag{6.8}$$

Holts-Winter

Holts-winter models, also referred to as triple exponential smoothing is similar to DES in that it also considers a smoothing (F) and a trend value (T). However, this technique also accounts for a seasonal component and are thus best suited for demand which have some recurring seasonal pattern. This seasonal component also requires an smoothing constant between zero and one and is indicated by beta (β). Furthermore, the seasonal value (S) depends on a predefined length p of a season and is set to be a yearly one within Ricoh's business. The equations below are used to calculate each of the three values.

$$F_i = \alpha \frac{y_i}{S_{i-p}} + (1-\alpha)(F_{i-1} + T_{i-1}) \tag{6.9}$$

$$Si = \beta \frac{y_i}{F_i} + (1-\beta)S_{i-p} \tag{6.10}$$

$$Ti = \gamma(F_i - F_{i-1}) + (1-\gamma)T_{i-1} \tag{6.11}$$

Ricoh's new system considers two types of triple exponential smoothing variations. The first one is the Holts-Winter Additive and is often used in situations in which demand has constant variation across a time period. The second variation is the Holts-Winter Multiplicative and should be used when demand has proportional variation across a time-period. The following equations are used to come up with the fitted and statistical final forecast values for both variations. Moreover, both variants require data for all historical months and thus cannot be used if there are any gaps in any of the historical months.

HWA	HWM	
$FittedValue_i = (F_{i-1} + T_{i-1}) + S_{i-p}$	$FittedValue_i = (F_{i-1} + T_{i-1})S_{i-p}$	(6.12)
$Forecast_{N+m} = (F_N + mT_N) + S_{N-p+m}$	$Forecast_{N+m} = (F_N + mT_N)S_{N-p+m}$	(6.13)

Croston's

The Croston's model is a variant on the general single exponential smoothing technique and accounts for intermittent/sporadic demand patterns. It uses three different values, a forecast value (Z), the average length between forecast values (X) and the number of periods between two non-zero demands (q). The forecast value is simply the SES equation which is affected by the other two values.

$$Z_{i+1} = \begin{cases} \alpha y_i + (1-\alpha)Z_i & , y_i \neq 0 \\ Z_i & , y_i = 0 \end{cases} \tag{6.14}$$

$$q_{i+1} = \begin{cases} 1 & , y_i \neq 0 \\ q_i + 1 & , y_i = 0 \end{cases} \tag{6.15}$$

$$X_{i+1} = \begin{cases} \alpha q_i + (1-\alpha)X_i & , y_i \neq 0 \\ X_i & , y_i = 0 \end{cases} \tag{6.16}$$

The results of these equations are used to calculate the fitted values as formulated below. The statistical forecast is often calculated using one of two ways, it is either a constant value or is reported as a sporadic value. Within the business of Ricoh, a constant value is reported.

$$FittedValue_i = \frac{Z_{i-1}}{X_{i-1}} \tag{6.17}$$

$$ConstantForecast = \frac{Z_N}{X_N} \tag{6.18}$$

Initialization & optimization

With the exception of linear regression, each algorithm needs to go through an initialization and optimization process. Many of the used formulas require a initial starting value, while all the different constant values (i.e., alpha, gamma and beta) need to optimized. There are different choices and methods to come up with initial values. The number of prior periods (k) is the variable which needs to be optimized for the moving average algorithm. In order to do so, Ricoh's system generates forecast output based on the k variable that ranges from 4-11. For all the other algorithms, Ricoh's system uses a backcasting methodology in which the initial values are set equal to the final values produced by the backcasting function. Backcasting a function is simply done by executing an algorithm on the reversed time series data. All constant values associated with one of the algorithms are optimized such that its values minimizes the sum of squared errors.

$$\text{Sum of squared errors} = \sum_{i=1}^{Count} (Y_i - f(x_i))^2 \tag{6.19}$$

6.1.6 Dis-aggregation

The next stage, as depicted in Figure 6.1, is to aggregate/disaggregate predictions originating from one of the just explained algorithms. This phase is not part of the system's standard set-up and is created specifically to test and compare different hierarchical levels of aggregation. As mentioned in section 4.1, all prediction values eventually need to be aggregated/disaggregated to some standardized settings (See Figure 4.1). To ensure fair and valid comparisons and conclusions, forecast performance/accuracy will also be evaluated on these standardized settings.

Prediction data needs to be aggregated whenever these are derived on a hierarchical dimension level lower than the desired levels of aggregation. Aggregation of data is very straightforward and is just simply the summation of all predictions within each cluster. Disaggregation, on the other hand, is a bit more complex and is required whenever forecasts are derived on a hierarchical dimension level higher than the desired ones. Many different disaggregation techniques have been examined in literature and according to the research of Gross & Sohl (1990), the mean proportional dis-aggregation methodologies are found to be most effective. That is why, within the context of this research, forecast data is disaggregated based on the historical proportion of each SKU relative to the total cluster volume. For example, if cluster X only consist out of a product A and product B. In which product A represents 75% of the total historical cluster volume and product B recorded the remaining 25%. If forecast are then generated on a cluster level of aggregation, than this volume will also be disaggregated according to these specific proportions. More specifically, in the situation that the system forecasts a demand volume of 100 for cluster X, then this total volume is disaggregated down to a volume of 75 items for product A and 25 items for product B.

6.1.7 Accuracy calculation

The final stage of the forecast model is to evaluate the forecast performance of each of the different levels of aggregation. The paper of Armstrong (2001) states that there is no universally accepted accuracy measure as each error measure has its own strengths and weaknesses. In general, it is recommended to use multiple error measures since the choice of one performance metric can affect the evaluation of results. The first metric that is included in order to evaluate forecast performance is the MAPE. Ricoh system's current set-up also evaluates forecast performance through the MAPE metric (See subsection 6.1.4) and since it is also one of the most commonly used forecast metrics, it is decided to include this metric as well. The MAPE metric is often used as a forecast measure since its output is easy to interpret and its errors are scale independent. It is scale independent since it is expressed as an percentage and its output can therefore be compared across different (sized) time series. However, some of the weaknesses of the MAPE metric are that it is sensitive for high errors if the actual value is relatively small (1), is not able to deal with zero actuals (2) and it is biased to under-forecasting since its error is bound to 0%, while over-forecasting can have an infinite error (3).

The second measure that is included within this research is the Root Mean Squared Error (RMSE). This metric takes the root of the mean squared differences between the forecasted value and the actual values.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2} \tag{6.20}$$

The major strength of this performance metric is that it punishes outliers since it squares the differences. Another advantage is that its errors are always expressed in the same units as within the original time series data. However, its disadvantages are that its errors are less interpretable since its squares the errors and it cannot be used to compare to other time series models which use different (sized) units (i.e., scale-dependent). Since this project requires the comparison of different sized time series, the RMSE needs to be transformed/normalized. The RMSE is scaled by dividing its total with the average demand volume of the time series data. Since the RMSE is a value that represents the standard deviation of the residuals, dividing it by average demand volume is almost similar to the formula of the Coefficient of Variation (CV). More specifically, the CV is defined as the ratio of the standard deviation to its mean. That is why the scaled RMSE, within the context of this research, will be referred to as the adjusted CV (i.e., CV*). The output of this metric is a percentage that indicates the amount of residual variance.

$$\text{CV*} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2}}{\frac{1}{n}\sum_{i=1}^{n}A_i} \tag{6.21}$$

6.2 Forecast model validation

All of the processes and algorithms discussed in the previous sections are successfully replicated with the aid of Python. To ensure that each step is correctly implemented and its output for the most parts corresponds with Ricoh's forecasting system, it is extensively tested. This process is carefully executed for each phase separately and validation data is gathered from system outputs originating from Ricoh's Asia-Pacific (AP) region. The AP region is one of the few regions in which the new system is already up and running and is therefore perfect in order to fine-tune and validate the Python script.

Based on their data it was found that most of the processes and algorithms are correctly implemented and perform exactly as intended. However, some of the algorithms (i.e., DES, Holts-winter and Crostons) have a slight deviation in their prediction values and after some research, it is concluded that this is most likely caused by a different metric that is used to optimize all constants. Since its deviations are proved to be negligible and all algorithms still react in how they theoretically should, it is not considered a problem. Furthermore, the aim of this research is to conclude on the performance of different levels of aggregation and not to produce an exact copy of Ricoh's new system. This is perceived as another reason of why these small deviations are considered to be acceptable.

6.3 Forecast model output

Before the replicated model is used to test all of the different levels of aggregation, the models output needs to be determined. Eventually, four different dataframes are created which all provide a different insight into the results. The first dataframe/output is most important and describes the MAPE and CV* along with the standard deviation of the MAPE per item per scenario. As mentioned before, a scenario represents a combination of a selected level of aggregation within each dimension as depicted in Figure 4.3. Table 6.3

describes an example of what the result of this first output look like for two specific scenarios (i.e., SKU-AllCustomers-EMEA and SKU-AllCustomers-OpCoCluster).

Table 6.3: Model output example of MAPE & CV* performance per Scenario

		SKU_AllCustomers_EMEA			*SKU_AllCustomers_OpCoCluster*		
Item	**Average sales**	**MAPE**	**St. Dev.**	**CV***	**MAPE**	**St. Dev.**	**CV***
number	**units/month**	*43.2%*	*31.4%*	*45.4%*	*46.8%*	*35.1%*	*48.88%*
404866	*88.08*	14.2%	17.9%	16.8%	13.0%	16.2%	14.7%
404867	*17.58*	30.5%	24.8%	35.0%	40.3%	25.7%	52.4%
407745	*31.50*	48.1%	44.4%	62.0%	29.8%	45.4%	35.2%
408064	*89.83*	34.8%	26.8%	37.9%	28.5%	13.4%	30.9%
408267	*299.17*	68.0%	44.5%	74.2%	137.3%	182.5%	218.1%
408273	*631.25*	65.9%	48.8%	70.2%	42.9%	35.3%	45.6%
408302	*253.58*	19.4%	11.4%	23.1%	21.7%	14.9%	27.0%

The models second output includes all actual demand values alongside with the scenario predictions per item/month. This output is mostly used to validate error measures and to inspect and conclude on potential outlier predictions. The next output provides insight into the outcomes of the historical summary analyses and states which algorithm is executed on each of the time series data. Examples for both of these outputs are shared in Appendix A.1. The final and fourth output provides the same insights as Table 6.3, except that it is now focused on the cluster performance within each hierarchical scenario. For example, within the scenario "SKU_AllCustomers_OpCoCluster" it does not describe the performance of each item, but instead describes the performance of this item across each OpCo cluster (e.g., Forecast performance of item 404866 in Eastern and Western Europe separately). The output of the example as explained above is shown in Table 6.4. All four outputs are created for both machines and supplies separately.

Table 6.4: Model output example of MAPE performance from a hierarchical perspective

Item number	**OpCo Cluster**	**MAPE**	**St. Dev.**
404866	*Eastern Europe*	44.9%	19.0%
404866	*Others*	24.1%	24.8%
404866	*Middle-East*	45.2%	27.3%
404866	*Northern Europe*	27.9%	14.1%
404866	*Southern Europe*	33.2%	42.4%
404866	*Western Europe*	19.4%	12.4%

Based on the outputs of a first test run on the currently used hierarchical settings, it is concluded that there are still some outliers which need to be carefully checked. These outliers are SKU's which have a very poor performance and therefore have a significant impact on the average MAPE and CV* values.

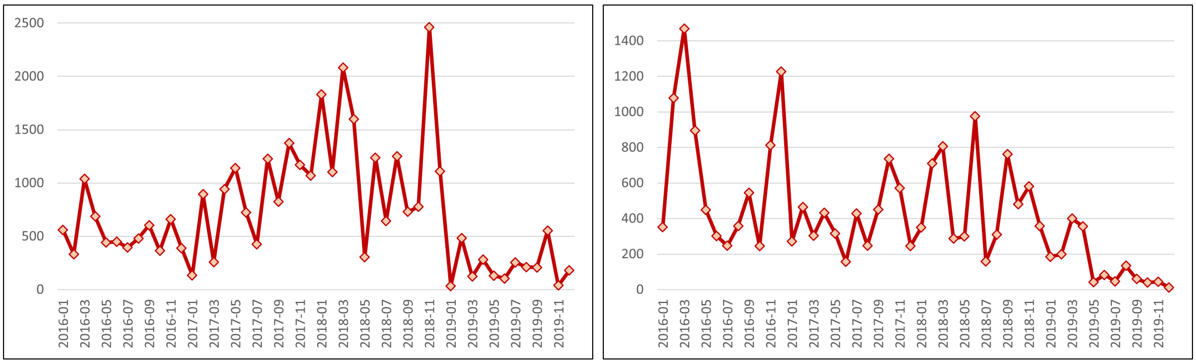


Figure 6.2: Demand patterns of two SKU's which show outlier behaviour

The actual demand patterns of two of such items are shown in Figure 6.2. Both of these plots clearly show that their demand volumes crashed during the final year of 2019. Since forecast performance is measured based on this final year and no statistical forecast will ever forecast anything close to these actual values as depicted in Figure 6.2, such items are concluded to be outliers and are thus removed from the final dataset. Another reason of poor forecast performance can be that some items are just too erratic, making them extremely unpredictable. That is why, after careful consideration, it is decided to remove all items which have a Coefficient of Variation greater than 0.5, while having a MAPE which was greater than 100%. These additional outlier analysis resulted in 19 machines and 64 supplies being removed from the final dataset. The final dataset includes a total of 57 machines and 232 supplies.

All these different machines and supplies are forecasted by all 36 different scenarios from which the results are discussed in the next chapter. The total run-time to calculate forecast on all scenarios was approximately 5 days for machines and just over 20 days for supplies. The most disaggregated scenario has the longest run-time (i.e., 10 hours for machines and 100 hours for supplies) and forecasts each item for every sales channel and every OpCo separately.

Chapter 7

Data analysis

The obtained forecasting results will be discussed in this chapter. All scenario results will first be summarized and discussed before they are analyzed through a statistical approach. This statistical approach includes conducting t-test on the general results from which more detailed analyses are carried out with the aid of ANOVA analysis. These detailed analysis are focused on comparing forecast performance across time alongside with evaluating forecast accuracy for the most important items. The validity of all these results are checked through a sensitivity analysis and the chapter ends by sharing some business insights.

7.1 Scenario results

Forecasting results on the same set of SKU's are obtained for each of the 36 scenarios. Forecasting results are expressed with the aid of two different forecasting metrics (i.e., MAPE and CV*). Each of the scenario's averages and standard deviations for both error measures are shown in Appendix A.2. The hierarchical scenario which is currently in-use is highlighted in green.

Comparing the best performing machine scenarios with the top performing supply scenarios shows that in general the forecasting algorithms can predict supplies more accurately. The MAPE performance of these scenarios are in the range of 25-30% for supplies versus 30-35% for machines. The CV* performance is also in favour of supplies with a score between 28-32% versus 39-43% for machines. This is no surprise, as it was briefly explained in section 3.4 that supplies are less impacted by tender deals and also are a more predictable product-type in general. However, comparing the 36 scenarios within each product-type shows that for machines all scenarios perform relatively similar, while performance across these scenarios for supplies is much more scattered. Without jumping to any conclusions, these differences indicate that for supplies it seems like that some levels of aggregation result in very poor forecasting performance. Furthermore, solely based on the average MAPE values, the currently used level scores best for machines, while the hierarchical scenario "*SKU-Dir/Indir-OpCoCluster*" scores best for supplies. The average CV* scores indicates other scenarios as the top performing ones. The "*Family-Channel-EMEA*" has the best CV* score for machines, while "*SKU-AllCustomer-OpCo*" scores best with respect to the supplies product-type. However, concluding solely based on a single average value is not recommended since averages can be misleading. That is why, in order to provide more precise and valid conclusions, all results are analyzed with more of a statistical approach.

7.2 Statistical approach

The statistical test that is used to compare forecast performance across all 36 scenarios is the so-called (Student's) t-test. The t-test is a statistical method that compares the averages of two groups from which its output is used to conclude whether the averages are significantly different from each other. There are different types of t-tests depending on the kind of data that is used (i.e., one sample, independent samples and paired samples t-test). The one sample t-test is used to check whether a data sample is significantly different than a certain predefined value. This kind of test can for instance be used to test the validity of a claim. For example, to test whether the processing speed of a certain type of printer is more than 20 prints per minute. The independent and paired sample t-tests are both used to compare sample averages of two groups and to conclude whether these are significantly different from each other. The independent test compares two different (independent) groups, while the paired t-test compares sample averages originating from the same sample group.

7.2.1 Student's t-test

Since the aim of this research is to compare forecast performance between different hierarchical scenarios evaluated across the same group of SKU's, results are analyzed with the aid of paired t-tests. More specifically, the *"Two sample, Two tailed, Paired t-test"* is used in order to analyse all results. The t-test are conducted on a two tailed distribution since there is no directional significance assumed. The null hypothesis of this test is that the mean difference between the paired predicted samples are equal, while the alternative hypothesis is that these are not equal. Moreover, the null hypothesis within the context of this research is rejected if the p-value is smaller than 0.05 (i.e., a significance confidence level of 95%). A confidence level of 95% is selected since it is the most commonly used confidence level within most scientific studies. Furthermore, the student t-test is a parametric statistic which means that it includes assumptions on the distribution of the population from which the sample was taken. Before the results of this test are discussed, assumptions with respect to the t-test first need to be addressed.

Statistical assumptions

The paired sample t-test has four main assumptions which need to be tested in order to assess the quality of the results. If one or more of these assumptions are violated, results might be less reliable. This statistical t-test makes the following four assumptions about the data:

1. *The dependent variable must be continuous*
2. *Each observation should be independent of every other observation*
3. *The sample differences should be approximately normally distributed*
4. *There should be no extreme outliers in the sample data*

The first assumption of the t-test requires the sample data to be numeric and continuous. Both error metrics, on which t-tests are conducted, are continuous since they are expressed as an absolute percentage ranging from 0% to infinity. Furthermore, almost all included SKU's within the sample data-set of a single scenario are independent from each other. The only SKU's that might have some correlation are a few toner supplies that for instance only differ in colour. Nevertheless for the majority of all SKU's, the forecast performance of a random SKU is not influenced by the forecast performance of another SKU and therefore the second assumption is also satisfied.

The third assumption requires the sample set to be normal distributed. However, the central limit theorem applies to the used sample data, since both sets include more than 30 observations. The central limit theorem states that the distribution of a sample mean tends towards a normal distribution if the sample size is large enough (n>30), even if the variables themselves are not normally distributed. That is why there is no consequences of violating this assumption if the data is not normally distributed. Finally, the fourth and final assumption states that there should be no extreme outliers in the sample data-sets. Throughout the previous chapters, SKU's which had an extreme poor performance (i.e., an outlier) have already been removed from this sample data-set. Briefly analyzing the performance results of the SKU's which are still left in scope, did not show any additional outliers. None of the SKU's have an average MAPE or CV* score of higher than 100% across all scenarios. Some individual scores on a specific scenario might have a greater error, but these need to stay included since its result indicates the choice of a poor level of aggregation.

Statistical results

The two sample, two tailed, paired t-test are conducted on all scenario and its results are summarized in Appendix A.2. Since this type of test only compares two samples at a time, a total of 630 t-tests per product-type are conducted (i.e., (36 scenarios x 36 scenarios - 36) / 2 = 630). Scenarios need to be compared to each other just once (i.e., division by two), and do not need to be compared to itself (i.e., subtracting 36). The number of significant t-test results between the machine and supply product types vary considerably. For machines a total of only 73 and 25 out of 630 scenarios have significant MAPE and CV* p-values. While for supplies a total of 558 and 549 out of 630 scenarios have significant differences within both performance metrics. This indicates that the selected level of aggregation has much more impact on the forecasting performance of supplies compared to the performance of these levels of aggregation for machines. Since providing an organized view of all individual p-values within a matrix of 36 by 36 is impossible, it is decided to count the number of times each scenario outperforms or under-performs another scenario. These results are shown in Appendix A.3 and Appendix A.4 for both error metrics per product type. Based on the results from both these tables, interesting observations are found. Firstly, some general observations will be discussed and thereafter observations for each of the product types specifically are mentioned. One of the strengths of including two different performance metrics is that it tests construct validity (Armstrong, 2001). Construct validity uses multiple metrics to test to what extent the chosen research metrics are valid and accurate. Construct validity within the t-test results is ensured, since both the MAPE and the CV* show comparable results. For instance, scenarios which mainly outperform other scenarios based on one of the metrics, are also most likely to outperform scenarios based on the other metric.

For machines specifically, a few things can already be observed. In addition to the metrics aligning with each other, it is observed that both performance values fall into a narrow range of values. Meaning that there are only a few scenarios which seem to outperform other scenarios, while most of the scenarios do not show any significant differences between each other. This indicates that there is not one specific aggregation level that significantly improves forecast accuracy, but more interestingly it shows that there are also not many aggregation levels that clearly damages forecast accuracy. Perhaps the most intriguing finding is that the forecast performance thrives best whenever a middle-out strategy is applied. A middle-out strategy within the context of this research is a strategy that does not aggregate data on the most aggregated level nor on the most disaggregated level. Hierarchical scenarios in which at least two out of three dimension aggregate data on the highest or the lowest hierarchical level seem to perform the worst. On the other hand, scenarios which have a combination of aggregation and disaggregation levels within the different hierarchical dimensions seem to be favoured. This finding is emphasised by the fact that the three worst performing scenarios, in terms of being outperformed, are the scenarios that either disaggregate the most (i.e., SKU-Channel-OpCo and SKU-Channel-OpCoCluster) or aggregate historical demand data the most (i.e., Line-AllCustomer-EMEA). Finally, while the currently used scenario (i.e., SKU-AllCustomers-EMEA) has the best MAPE score, it does not outperform the most scenarios based on the t-test results. There are five scenarios which at least outperform the same number of scenarios based on the MAPE, while this scenario does not outperform any other scenarios based on the CV* t-test comparisons.

In contrast to the machine results, the t-test results of the supplies show much more deviation. The table depicted in Appendix A.4 clearly shows that aggregation strategies on a product dimension does not improve forecast accuracy at all. Forecasting each SKU separately results in the best forecasting performance which is emphasized by the fact that it is the only dimension which includes scenarios that are not outperformed by any others. While the product family is the only class that still has somewhat comparable forecast results, it is still most of the times outperformed by at least one of the SKU generated scenarios. The product line and especially the product sub-type classes negatively impact forecast performance considerably. The worst performing scenario (outperformed by all 35 scenarios) based on both error metrics and in terms of average values and number of times being outperformed is by far the SubType-AllCustomer-EMEA scenario. Furthermore, among the nine scenarios which are aggregated on a SKU level, it is interesting to see that the combination with the most disaggregated customer and region classes (i.e., SKU-Channel-OpCo) once again results in the worst performance. Finally, it seems that there are other SKU aggregated scenarios that significantly outperform the scenario which is currently in use within Ricoh (i.e. SKU-AllCustomers-EMEA).

7.2.2 ANOVA analysis

Since these t-test comparisons across all 36 scenarios are quite generic, it is decided to proceed by conducting more detailed statistical analysis. Some more detailed analysis are conducted in order to better understand forecast performance across different scenarios. These statistical analysis are conducted on the best performing scenarios based on the results from the previous section. In order to decide which scenario are considered as the "best" performing scenarios, a decision rule is formulated. The decision rule is aligned with members from the demand planning and business excellence departments and is stated as formulated below.

A scenario is considered in the set of "best" performing scenarios, if it outperforms at least two other scenarios on either the MAPE or the CV metric, while not being outperformed by any other scenario on both of these metrics.*

Since comparisons with the currently used level are very relevant from a business perspective, it is decided to always keep this scenario in scope. Even if it does not comply with the formulated criteria. Applying this decision rule on the statistical comparison tables of both machines and supplies leads to a reduced number of scenarios. For machines a total of 11 out of 36 are included based on this criteria which are all shown in Table 7.1 below. Each and every hierarchical level within the three dimensions are represented at least once in the scenarios that are included.

Table 7.1: Reduced number of Machine scenarios based on t-test criteria

Hierarchical Scenario	MAPE	Out perform	No signficant difference	Under perform	CV*	Out perform	No signficant difference	Under perform
SKU_AllCustomer_EMEA	35.33%	6	29	0	40.23%	0	35	0
SKU_AllCustomer_OpCoCluster	36.87%	2	33	0	41.60%	0	35	0
SubType_Channel_EMEA	35.66%	13	22	0	39.10%	3	32	0
Family_AllCustomer_OpCo	35.46%	9	26	0	39.84%	0	35	0
Family_Dir/Indir_EMEA	36.46%	5	30	0	39.80%	1	34	0
Family_Dir/Indir_OpCo	36.26%	4	31	0	39.78%	0	35	0
Family_Channel_EMEA	36.02%	6	29	0	38.88%	4	31	0
Line_AllCustomer_OpCo	36.77%	6	29	0	39.94%	3	32	0
Line_Dir/Indir_EMEA	38.03%	1	34	0	41.05%	2	33	0
Line_Dir/Indir_OpCo	37.02%	5	30	0	40.12%	1	34	0
Line_Channel_OpCoCluster	36.34%	9	26	0	39.68%	6	29	0

A total of 5 out of 36 scenarios are being included for supplies based on this decision rule. However, despite being outperformed by 7 other scenarios, the currently used scenario is added as the sixth scenario. It is no surprise that these included scenarios are only focused on the SKU class, since the previous section discussed the observation of this class clearly outperforming the other product dimension classes. The classes from the other two dimensions are all represented by exactly two scenarios each. Table 7.2 shows the supplies t-test table, only including these six scenarios.

Table 7.2: Reduced number of Supply scenarios based on t-test criteria

Hierarchical Scenario	MAPE	Out perform	No signficant difference	Under perform	CV*	Out perform	No signficant difference	Under perform
SKU_AllCustomer_EMEA	31.33%	19	9	7	34.20%	19	10	6
SKU_AllCustomer_OpCo	26.25%	29	6	0	28.36%	30	5	0
SKU_Dir/Indir_OpCoCluster	25.82%	31	4	0	28.64%	30	5	0
SKU_Dir/Indir_OpCo	27.07%	28	7	0	30.11%	20	15	0
SKU_Channel_EMEA	26.33%	29	6	0	28.98%	29	6	0
SKU_Channel_OpCoCluster	26.52%	29	6	0	28.77%	29	6	0

Additional statistical analysis focused on these reduced number of scenarios are conducted via so-called ANOVA analysis. ANOVA stands for analysis of variance and its most important difference compared to t-tests is that it is able to compare more than two group samples. More specifically, the one-way ANOVA is used since there is only group variable (i.e., type of hierarchical scenario) and one dependent variable (i.e., forecast performance). MiniTab is the statistical software that is used to conduct the above explained one-way ANOVA analysis. This test is specifically used to conduct more specific and relevant analysis, since all data samples per product type can be tested in one go.

Executing the ANOVA analysis on the same data-set as used in subsection 7.2.1 does not add much value, as both statistical methods are almost identical and will therefore provide similar conclusions. That is why, as a first step, the sample data is slightly re-framed and in doing so the average yearly data is turned into monthly data. In doing so, each SKU is represented through 12 data-points instead of a single average year value per scenario. This is favourable since an ANOVA analysis generally benefits if more data-points are used. However, the downside of this data transformation is that only the MAPE value can be used, since the CV* metric requires input from more than one value. That is why from now on all statistical analysis are conducted only for the MAPE values. Executing the ANOVA analysis on this specific data-set did not show any additional significant differences. The only significant difference that is found is for the supply scenarios between the current used level and all the others. These findings are similar to the previous section and therefore validate the earlier observations made from the t-test results. In the next two paragraphs, some more detailed analysis are conducted in order to try and find aspects in which some scenarios thrive better than others.

ABCD analyses

The first additional analysis that is conducted in order to compare performance among the "best" performing scenarios is aimed at product classification. Ricoh uses an ABCD classification to label their product assortment based on the business impact each product has. SKU's that contribute to the top 80% of total sales volumes (i.e., demand volume x sales value) are labeled as A products, while products representing the lowest 5% of total sales volumes are labeled as D products. The ABCD classification is based on the average monthly sales volume calculated over the last six months of the year 2019. The classification values were unfortunately not incorporated into the obtained data-set and are therefore calculated for both product-types. A total of 107 out of 176 initially included machine SKU's have recorded sales volume throughout the last six months and a total of 473 out of 706 initially included supply SKU's have recorded sales volume throughout the last six months. The number of SKU's per product classification are shown in Table 7.3.

Table 7.3: ABCD classification for both Machines & Supplies

Product Classification	**% of sales volume**	**Machine SKU's**	**Supply SKU's**
A	0-80%	25	55
B	80-90%	13	30
C	90-95%	8	33
D	95-100%	11	114

It is argued that accurately forecasting A classified products is of much more value than accurately predicting D products. More specifically, from a business perspective it is more important to accurately predict products which have a larger impact on a business. That is why an ANOVA analysis is conducted on just these A items specifically. Based on the output from this analysis no additional findings are observed for both product types. The scenarios associated with machines still have comparable performance in which no scenario performs significantly worse or better than any other. Within the scenarios which are considered for the supplies there is once again only a significant difference between the current used level and all the others. While these result do not show anything new, it emphasis the fact that Ricoh might want to look into different levels of aggregation for the supply product type specifically.

Time series analysis

Another aspect that is considered to be relevant is to check forecast performance over time. In all of the previous analysis forecast performance was focused on the average forecast accuracy over all 12 months. This is done primarily due to the fact that Ricoh's current forecast approach submits a 12-month forecast on a monthly basis. While following this procedure is not wrong, it might be of more interest to conduct additional analysis focused on this time aspect. More specifically, for Ricoh's business it is more crucial to accurately forecast the first six months in comparison to the last six months. This due to the fact that business decisions based on forecast data from the first six months are less flexible compared to business decisions that are made for a later stage (i.e., the last six months). Before these analysis are conducted, forecast performance over time for each of the scenarios are first visualized. This is done to check the assumption that forecast performance indeed decreases over time. This assumption is based on the fact that the further in the future someone tries to forecast, the more uncertain these predictions often get. For instance, it is easier to provide weather predictions for tomorrow than to predict the weather for next week.

Figure 7.1 depicts the forecast performance over time for a single scenario with respect to both machines and supplies. For the sake of simplicity, the scenario which is currently in-use is taken as an example. The plots visualizing forecast performance for all the other scenarios are depicted in Appendix A.5 and Appendix A.6. All plots in the appendix are plotted as Xbar-S quality control plots which show the average mean and average standard deviation over time. These Xbar-S plots are specifically created for testing data transformation techniques explained in subsection 7.2.3. However, since these plots show the average MAPE value over time they can also be used within this section.

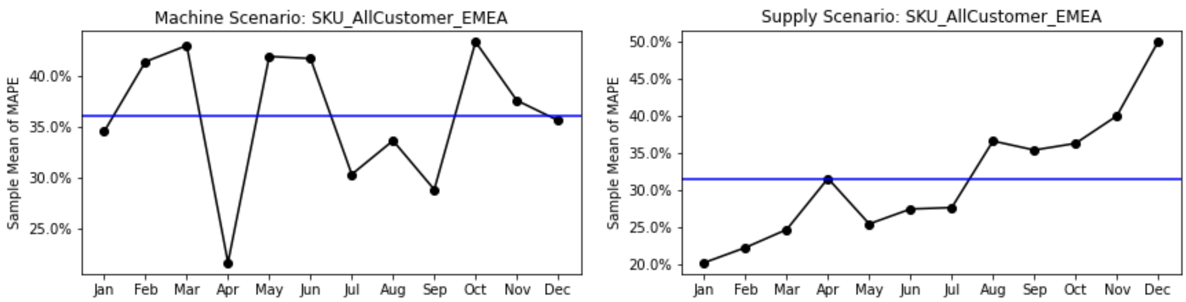


Figure 7.1: Forecast performance over time for the scenario that is currently in-use

After a quick scan, it is concluded that the plots shown in Figure 7.1 represents general behaviour of all machine and supply scenarios. In general, it is observed that the machine scenarios show quite a random/erratic forecast performance over time, while all supply scenario clearly show a better performance in more recent months. ANOVA analysis on both product types are conducted on the performance over time. These analysis did not result in any additional significant differences between the "best" performing scenarios. This indicates that while the graphs show interesting trends, they all proved to not be statistically different from each other. For instance, while the supply scenarios clearly indicates a forecast accuracy difference over time, they all seem to have similar behaviour/patterns. Finally, purely based on the plots as depicted in Appendix A.5 and Appendix A.6, it is interesting to see that for all supply scenarios there are negative outliers to be observed during the months of April, August and December. This is even more interesting if this observed pattern is compared to all machine scenarios. This due to the fact that for machine scenarios the month April is almost always the best performing month in terms of forecast accuracy. Furthermore, most machine scenarios seem to forecast the month of September also with a relatively good accuracy.

7.2.3 Sensitivity analysis

As a final step in this chapter a sensitivity analysis is conducted. Within the context of this chapter, a sensitivity analysis is conducted to check the validity of all previously discussed statistical results. This includes testing all results based on rounded/non-negative predictions, transformed data and through a non-parametric statistical test.

The first analysis that is conducted is to examine the effect of only allowing the system to submit rounded and non-negative values. Ricoh's forecasting system is currently not restricted since it can both predict decimal numbers as negative values. The latter can only be the result of algorithms which include a large enough negative trend component for a specific time series. From a business perspective, negative predictions will always set to be equal to zero, while floats are always rounded to the nearest value. Testing this restriction showed that the impact of this practical implication is negligible and that all previous statistical tests still show the same outcomes. In addition to resolving this doubt that was raised by Ricoh's management, it also showed that the data-set that is used is not sensitive to these kind of changes.

Next up, several data-transformations techniques are applied to the prediction data in order to guarantee data stability and to get closer to a normal distribution. The latter is primarily of importance with respect to the assumptions associated with the ANOVA analysis. Data stability is tested through the earlier mentioned Xbar-S plots depicted in Appendix A.5 and Appendix A.6. The Xbar-S plot are often used control charts to examine data stability of the mean and standard deviation over time. These kind of plots calculate an upper and lower limit (highlighted in red) and if many data-points exceed any of these control limits it might imply that data is not stable enough. Most of the mean MAPE values for all scenarios seem to stay in between the control limits. However, the data-points associated to the standard deviation over time might suggest that the data is not stable. Moreover, the distribution of the MAPE value is checked and is as expected very much left skewed. This makes perfect sense since the MAPE cannot exceed values lower than 0% while its upper bound is infinite. Commonly used data transformation techniques to transform left skewed data into more normal distributed data are the log transformation and the square root transformations techniques. The transformed data distributions are illustrated through several histograms depicted in Appendix A.7 and Appendix A.8. Based on these results it is decided that the log transformation technique works best and is therefore applied on the non-transformed data. Furthermore, all Xbar-S plots are once again created and this time almost no mean or standard deviation values exceed any of the limits. Finally, all previous analysis are conducted (on the transformed data) and do not show any major differences, and therefore once again validates earlier drawn conclusions.

Finally, as a last check, instead of a parametric test (e.g., a t-test or an ANOVA analysis) a non-parametric test is conducted. Non-parametric test are statistical test that do not assume any underlying distribution and are in general a better safeguard against drawing incorrect conclusions. The only downside of such tests is that they have less statistical power, which means that they are less likely to find a statistical effect. The so-called Mood's median test is the non-parametric test that is used and is often used as the non-parametric alternative for the one-way ANOVA. This test examines the equality of medians from two or more group samples. There are no differences found between the results of the parametric and non-parametric tests. All three analysis conducted throughout this section showed that the earlier discussed results are valid/trustworthy and can therefore be used in chapter 9 for drawing relevant conclusions

7.3 Business interpretation

All of the observations and results described throughout this chapter are presented to representatives of the business excellence and demand planning departments. This is done in order to examine whether there are aggregation preferences from a business perspective. The results that were primarily discussed during this meeting are the scenarios shown in Table 7.1 and Table 7.2.

For machines, the eleven best performing scenarios were up for discussion. First up, all representatives favoured scenarios in which not all three of the dimensions are different to the originally used scenario. Following this procedure, a total of ten scenarios were still up for debate (i.e., Familiy-Dir/Indir-OpCo was the only scenario that was different on all three dimensions). From these ten scenarios, everyone preferred the scenarios which on the product dimension were the closest to the SKU hierarchical class. More specifically, from a business perspective more value was placed on retaining individual time series information compared to the benefit of decreasing individual time series noise. They argued that with the implementation of the new system much more attention will be dedicated to manual outlier correction and in doing so it will already reduce the extent of individual noise. In other words, Ricoh expects the scenarios that are more focused on individual time series data to provide more accurate forecasting results in the near future, since the individual data quality is also expected to be improved. As a result, Ricoh considers the scenarios which are either focused on the SKU or the product-family class as the most promising ones. This resulted in a total of five scenarios which are listed below. The product sub-type scenario is skipped since it was concluded that its clusters are more related to the product-line class than to the product family class.

- SKU-AllCustomers-EMEA
- SKU-AllCustomers-OpCoCluster
- Family-AllCustomers-OpCo
- Family-Dir/Indir-EMEA
- Family-Channel-EMEA

For supplies, this same discussion was held and everyone agreed on the fact that applying a different scenario needs to be tested during the implementation of the new system. It was generally concluded that the scenario SKU-Dir/Indir-OpCoCluster seems most promising since it is the scenario, based on both metrics, that outperforms the most scenarios. However, similar to the machine discussion they are interested to see the effect of improving individual data quality on the forecast performance of these "best" performing scenarios. To properly test this impact it is decided to keep all top five scenarios in scope.

- SKU-AllCustomers-OpCo
- SKU-Dir/Indir-OpCoCluster
- SKU-Dir/Indir-OpCo
- SKU-Channel-EMEA
- SKU-Channel-OpCoCluster

Chapter 8

Judgmental forecasting

In the current approach, team members of Ricoh's demand planning department carefully review the statistical derived forecast generated as discussed in the previous chapters. This reviewing process, also referred to as judgmental forecasting within literature, is an important final step and can in general either improve or damage forecast accuracy. This chapter aims at achieving a better understanding of the reviewing process within Ricoh.

8.1 Data collection

In order to get a better insight into the forecasting reviewing process of Ricoh, several members of the demand planning department are asked to participate by filling in a questionnaire. This questionnaire consist out of twelve different questions which all provide an insight into the judgmental forecasting approaches used by Ricoh. The first four questions are more general questions providing an understanding of the planning experience and the current responsibilities of each participant. Questions five up to nine provide insight into the number of SKU's that the participant needs to review, the number of times an adjustments is made and most importantly the participant share their reasons for their decision to agree or disagree with the initial proposed forecast figures. Insights into the latter topic are extracted by asking each participant to write down five examples of situations in which the have not adjusted the system generated forecast and five situations in which they did. The next two questions (i.e., ten and eleven) are more focused on extracting information on whether the planners feel like they are missing some (data) insights that might help them during the reviewing process. The final question asks about whether the participants can come up with any additional (data) variables which could be incorporated into the forecasting system. This could for instance be variables which they currently actively use during the reviewing process and that have the potential to be quantified and incorporated into the system. The questionnaire which is distributed among all participants can be found in Appendix A.9.

A total of seven participants are asked to fill in this questionnaire. These participants represent the team that is responsible for the reviewing process of both the machine and the supply product types. This is done since the previous chapters were also focused on these two product types. Five out of the seven participants are demand planners while the other two are their supervisors. The responsibility of the demand planner within this process is to review the proposed forecast and adjust if necessary. Each demand planner is assigned with a pre-specified product portfolio which represents a range of SKU's within one of the product types. A supervisor is ultimately responsible for the final proposed forecast and therefore often checks the reviewed product portfolios from more of a high-level perspective. One of the supervisors is responsible for the majority of the machine portfolio, while the other is ultimately responsible for the supply portfolio.

As mentioned before, the first week of every new month is dedicated to this forecast reviewing process and during one of these weeks each participant was asked to fill in their questionnaire. In doing so, since they are in the middle of their reviewing tasks, it is assumed that all participants are better able to answer all questions. Furthermore, to ensure a full understanding of all given answers and to further explore interesting ideas a follow-up meeting is planned in which each participant is interviewed. All summarized transcripts of these interviews can be found in Appendix A.10 and Appendix A.11. The next two sections will discuss and summarize the answers that are given by the participants. Each of the product types are discussed in a separate section, since they are too different.

8.2 Reviewing process of Machine SKU's

The first main insight that the interviews provided is that at this moment in time, machines SKU's are forecasted without the support of any forecasting software. Machine SKU's in general have a relatively short product life-cycle and since the current system is not able to link any predecessor-successor relations it is often only able to use limited amounts of historical data. With the implementation of the new system as discussed in chapter 6, these issues are solved and therefore will fulfill the demand planning department's strong desire to get back to a system generated forecast. Until then, Ricoh creates all of his machine forecast figures with the aid of an Excel file. The general idea of this Excel file is that a planner has to decide on an average year value (i.e., a base-rate), which is multiplied with a monthly seasonality percentage. These seasonality percentages are either calculated based on historical demand information or once a year determined by the demand planners themselves. These seasonality percentages are not determined for each SKU separately, but mainly defined on a product family level. The main task of a demand planner during the reviewing process for machines specifically is to decide on the value of this single base rate for each machine SKU. As mentioned before, this manually selected base rate value is multiplied with the defined seasonality percentages and results in twelve unique forecast prediction values. A planner occasionally makes additional adjustments within these twelve prediction values. Furthermore, these twelve prediction values represent regular demand, since all of Ricoh's tender demand is forecasted separately by these same demand planners. In the historical demand however, this separation is not present anymore. That is one of the reasons why before a planner decides on the base-rate, he/she has to manually clean the historical data.

8.2.1 Adjustment motives

From these interviews it is concluded that a planner decides on this base rate by mostly looking at the historical year-on-year averages and the more recent monthly averages.

Depending on the visual interpretation of the historical demand pattern of a SKU, a planner either puts more focus on the year-on-year averages or on the most recent monthly demand averages. More specifically, whenever a planner recognizes a obvious trend-line more emphasis is put on the year-on-year values, while if its pattern is perceived as very random and erratic more value is placed on the monthly averages. Moreover, planners state that they increase the base-rate whenever historical demand showed regular outlier behaviour. This is done even if historical demand is perceived as fairly stable, as it is expected that these outliers will return on regular basis. It is argued by the planners that this will enable the organization to handle potential unexpected demand. Whether this improves overall forecast accuracy is not certain. Another interesting finding is that whenever historical demand patterns of a SKU are perceived as random, planners tend to correct for the seasonality component. This is done by manually adjusting all twelve prediction values closer to the chosen base rate. It is argued by planners that the seasonality component is not reliable whenever demand data is observed to be too erratic.

Based on the output from these interviews, it was very clear that planners do not only adjust forecast values solely based on historical demand data. For instance, in some specific situations, Ricoh's marketing department provides feedback on the initial chosen forecast values. In such situations, marketing generally suggest the planners to increase their forecast values, even if the historic data does not show any sings to do so. Their reasoning to increase forecast is often related to annual budget goals, future promotions or just general expectations. Marketing also supports a planner in deciding the forecast values during a phase-in/out procedure. During a phase-in procedure marketing shares their sales projections for this new product, while during a phase-out process the discussion is more to assure a smooth transition and is therefore focused on how much demand is still expected until the end of product lifetime. Supply chain disruptions such as Covid-19, part shortages and imposed restrictions are also often factors that are taken into consideration by a planner. The pandemic for instance made all demand streams crash and therefore does not represent normal behaviour. A planner considers these kind of disruptions when analyzing historical demand or when trying to predict future demands.

8.2.2 Missing insights

The question whether the planners felt like they were missing any information/data insights resulted in different responses. While some planners felt like additional insights would benefit their review process, others were not so convinced. An insight which was mentioned multiple times is insight into how the historical demand was distributed among the different OpCo's. They argued that this kind of information would help a planner to better explain previous demand behaviour and in doing so it will benefit predicting future demand. For instance, Ricoh international is well known for an erratic demand pattern, while it would also have been of added value to quickly see how much demand of a SKU was historically sold to Russia (due to imposed trade restrictions). A more precise insight into normal demand behaviour versus tender/outlier demand was most probably the most common response. At this moment in time, planners are looking at unprocessed demand data which they have to clean themselves. This made many planners feel like it can be done with much more precision. In other words, they feel like that base rates, but also seasonality percentages are more likely to be wrongly assumed since both are reviewed based on historical demand data that is only purged through a couple of manual initiatives. That is why one of the requested insights by planners is to be notified on historical events that may have affected historical demand patterns. While every planner can link the low historical volumes during 2019-2020 to the global pandemic, there might be some less known other events/ SKU specific circumstances that have impacted demand volumes.

8.2.3 Additional variables

Finally, all participants are asked whether they can think of any relevant data variables that could be incorporated into the statistical forecasting models. While this was a hard question to answer, it still resulted in some interesting given responses. One of the ideas was to provide the system with information about future promotions, while another idea was focused on incorporating general economical trends. While these could indeed potentially improve a systems forecast, the overall conclusion of such variables is that they are very hard to be quantified within the context of Ricoh's business.

8.2.4 Main findings

To summarize, the review process for machine SKU's and even the forecast process is a very manual one. All historical demand data is consolidated into an Excel file in which a planner must decide on a certain base rate value that is multiplied by predefined seasonality percentages. Depending on the observed demand patterns, most planners decide on this base rate value by looking at the year-on-year and/or on the most recent historical monthly averages. Moreover, planners also take into consideration events that had or still have a big impact on (historical) demand volumes when analyzing these historical demand figures. Deviating from these values is only often done in situations in which these supply chain disruptions played a big role or whenever marketing suggest to do so. The latter could be due to established budget goals, planned promotional activities or different sales projections. Moreover, planners indicated that an insight into different levels of aggregation (i.e., OpCo) could be of interest for the reviewing process, while there is also need for a more reliable insight into the normal/standard demand behavior. At this moment in time, data cleaning is mostly conducted based on manual initiatives in which planners must both remove obvious outliers and differentiate tender demand from standard demand.

8.3 Reviewing process of Supply SKU's

Different to the machines, both the forecast calculation as the reviewing process for all supply SKU's are conducted within Ricoh's forecasting system. While supplies are in general more predictable and therefore more eligible to be predicted solely based on a system, they still are reviewed by a team of demand planner. This is strongly requested by the organization since supplies are considered very important to Ricoh's business. All unique SKU's are calculated based on statistical forecasting algorithms from which its predictions can be adjusted by a planner. The main task of planners during the reviewing process is to decide whether the algorithm that is selected for each time series data is the most appropriate one or that they believe that a different one might be a better fit. Each statistical algorithm predicts a total of twelve forecast values and in addition to agreeing on the forecast algorithm a planner can also decide to adjust one of these twelve forecast values specifically. Within the reviewing process of supplies, Ricoh uses a ABC-XYZ classification policy that evaluates each SKU based on their predictability and business importance. This classification is used by planners to devote most attention to reviewing items that are both important to the business while also proved to have a low predictability. Such items are less likely to be accurately predicted with the aid of statistical models and therefore often require the most manual attention. Another important detail for supplies specifically is that a supply is often part of a series which for instance represents all of the different colours of the same toner type. It is mentioned that whenever an adjustment is required for a certain supply, it is often necessary for all of its related supplies within the same series.

8.3.1 Adjustment motives

Based on the interviews it was found that planners often make adjustment to the statistical proposed forecast in a sense that they agree with the direction, but disagree with the magnitude of the forecast. They provided multiple examples in which they feel like that the statistical models correctly interprets the direction of the trend-line, but its corresponding magnitude is too strong. Furthermore, output from the interviews confirmed the usage of this classification policy. However, it was emphasised that still almost all items within a planner's portfolio are reviewed and the classification policy is only used to distribute a planner's time and attention.

The first main finding from these interviews is that during the reviewing process much attention is devoted to checking seasonality. It is stated that seasonality has a big impact on the demand patterns of most supplies and therefore plays an important role during the reviewing process. Planner's often check whether the seasonality patterns as standardized by Ricoh are present in the historical demand data and if so, they will make sure that the forecasted values derived from the statistical models also show this same seasonal behaviour. Similar to the machine forecast review is that planners observe the year-on-year averages in order to conclude to what extend the trend-line is increasing, decreasing or stabilizing. Another factor that is important for supply demand planners to conclude on future trend-lines is the so-called machines-in-field (MIF). The MIF value represents the number of (active) machines sold by Ricoh and therefore provides indirect information about future supply requirements. For instance, if a machine is phased-out at a certain moment, than it is not possible for the MIF value to increase and will therefore stabilize before eventually decreasing. This behaviour will be similar for all of its corresponding supplies, making the MIF value an important piece of information to monitor. More specifically, supplies for machine SKU's which have a relatively low MIF values, while having a high supply consumption per machine, are very much impacted by a change in this MIF value. Ricoh does not register an exact MIF value for every machine, but supply planners instead look at the item status and historical/forecast values of these machines to access whether the MIF is increasing, stabilizing or starting to decrease. Furthermore, Whenever a planner cannot visually identify a clear demand pattern/trend-line and the historical demand is therefore perceived as peaky, its forecast values are often set equal to the average of the most recent months while being slightly adjusted according to standard seasonality protocol.

Since newly introduced supplies do not have sufficient amounts of historical demand available to calculate an accurate forecast, they are handled separately. Finally, supply planners, to a lesser extent, also consider input from marketing. Marketing could inform supply demand planners about hardware availability issues, significant MIF developments, promotional activities and assortment rationalizations.

8.3.2 Missing insights

As concluded from the interviews, many planners argued that the MIF value is an important indicator which is used during the reviewing process. A more quantified value representing this MIF value is often given as a response in order to better support a planner in reviewing and adjusting the statistical derived forecast. Moreover, same as for the machine planners, insight into how demand is distributed across the different OpCo's is mentioned as a potential valuable insight.

8.3.3 Additional variables

Once again, participants found it hard to come up with variables that could potentially improve system generated forecast figures. Nonetheless, some interesting answers are discussed. The first one is to incorporate the relationship between the machine and supply side of business within the system. If it is possible to accurately quantify the MIF value and if there is a good understanding of the impact of a change in this MIF value, then this information should be provided to the system. This could even be combined with information regarding average print volumes for each customer/OpCo/machine for instance. It is believed that all of the above could improve predictions and will make Ricoh's forecasting process more data driven and less based on gut feelings. Another interesting answer is to try and learn/standardize the phase-in procedure. Perhaps there is a standardized demand pattern to be identified for a range of products that a system can apply throughout future phase-in projects.

8.3.4 Main findings

To summarize, all supply SKU's are forecasted with the aid of one of the available forecasting algorithms. The main task of a planner during the reviewing process is to evaluate whether the system selected algorithm is indeed the most appropriate one. A planner primarily focuses on seasonality, year-on-year averages and the MIF value in order to access the latter. Even if a planner agrees with the general output from a statistical algorithm, they can decide to adjust specific months within the forecast. This is for instance done to align more with expected seasonality or since it is concluded that the algorithm slightly overshoots a planner's expectations. Input from marketing is once again occasionally used to deviate from the initial demand expectations. Marketing could for instance provide information in terms of hardware availability, MIF developments and temporal promotions. Next up, it is interesting to observe that insights into demand distributions across OpCo's are also mentioned for supplies as an insight that would help support a planner. Furthermore insight into a more quantified MIF value is also assumed to be of value and could even be incorporated within the forecast that are generated by the system. However, in doing so, Ricoh must be sure that this MIF value is accurate and that there is a good understanding of the impact of a change in this MIF value.

Chapter 9

Conclusion

This final chapter summarizes the main findings and highlights recommendations for the company. A distinction is made between research findings and business findings. The former will discuss more generic conclusions, while the latter concludes on findings that are most relevant for Ricoh specifically. Research limitations and ideas for future research are shared in the final two sections of this chapter.

9.1 Research findings

This section will discuss the findings that contribute to existing literature in rather different ways. First up, within the context of demand forecasting, literature often summarizes the concept of hierarchical forecasting as a trade-off between capturing individual behaviour on a disaggregated level and being better able to manage individual noise on a more aggregated level. The results from this study clearly showed the impact of the two. In scenarios in which data was disaggregated on the most individual levels on all three dimensions resulted in poor performance, while the same is observed whenever data is aggregated on the highest levels with respect to these dimensions. It is concluded that the former results in rather unreliable estimations because for most time series relatively little information is available, while the latter is assumed to be very inefficient since the aggregated time series data is not able to capture heterogeneity among individual demand patterns. These results therefore provided evidence that a more balanced strategy (i.e., the middle-out approach) is often most favourable and provides the most accurate forecast output. However, it must be mentioned that the most appropriate level of aggregation is context specific and very much depends on the underlying data distribution. This conclusion is emphasized by the fact that this study tested two different product types which had rather different performance across the same kind of data aggregations.

Secondly, while the different levels of aggregation have been tested in a company specific context, it is argued that the constructed methodology within this thesis can still be generalized and might motivate researchers to conduct further empirical or theoretical studies on this topic. This method showed that despite the level on which forecast are required, forecast can be calculated on almost any level of aggregation. In order to do so, the first step is to analyse whether the defined levels of aggregation are coherent with available data. Secondly, this methodology showed insight into how these forecasting results can be gathered and more importantly how these can be analyzed. It is claimed that the statistical approaches used within this thesis, including all of its detailed analyses are very much applicable and reproducible across these kind of studies. More specifically, it is argued that determining the most promising levels of aggregation based on the decision rule focused on statistical out-performance is generally applicable in other studies as well.

Moreover, different to most other studies, this study examines the impact of hierarchical forecasting from a multi-dimensional perspective. More specifically, this study considered data aggregation across multiple dimensions (i.e., product, customer, region), while most studies focused on this topic only examine the impact of data aggregation within a single dimension.

Finally, this thesis in general contributes to literature since it draws the attention of both practitioners and researches on the importance of choosing the most appropriate levels of aggregation to calculate forecast on. This is emphasised by the fact that the results of this study are significantly different across various hierarchical scenarios. Perhaps even more intriguing is that for one of the two product types the level on which the forecast are required does not provide the most accurate forecasting results.

9.2 Business findings

First up, it is concluded that the forecast accuracy heavily depends on the data quality of the time series that needs to be estimated. During the analysis of the initial system generated results discussed in section 6.3, it is observed that Ricoh's forecasting software is very sensitive to time series data that is considered volatile. Moreover, in chapter 8 it was also mentioned that during the reviewing process the quality of data is a crucial factor, and that the decision whether or not to make an adjustment is often based on largely unprocessed amounts of historical data. This shows that both a system as a human planner benefit from a standardized process in which better data quality is guaranteed. This emphasizes the importance of an efficient and standardized procedure aimed at pre-processing all historical data. While the outlier detection policy that is incorporated into the new system (see subsection 6.1.3) is a step into the right direction, there is still much progress to be made. This due to the fact that this policy is only focused on correcting outliers within monthly aggregated values, while it is also considered of much value to correct outliers from different, more detailed, perspectives. An outlier might not stand out on an aggregated level, but is considered an outlier on a more detailed disaggregated level. For instance, irregular demand volume for an often low volume demanding OpCo might not stand out on an OpCo cluster level and even less on an EMEA level of aggregation. The correction for such outliers at this moment mostly depends on various inconsistent manual attempts and initiatives. That is why this study recommends Ricoh to define a more standardized and structured data processing method which will result in more reliable and consistent input. This method should structurally be able to separate tender demand volumes from regular demand volumes, identify outliers on a monthly aggregated demand level, and do the same on a more detailed order line level. In order to do so, it is considered crucial to establish a well aligned approach involving a system and a demand planner.

The results with respect to the different hierarchical scenarios also provide Ricoh with a couple of interesting insights. For both product types it was found that forecast performance is most optimal if data is aggregated according to a middle-out strategy. Furthermore, results showed that certain classes within a hierarchical dimension might work well for one product type, but result in poor forecast performance within another product type. For instance, different classes within the product dimension result in somewhat comparable performance for machines, while for supplies there is one specific class (i.e., SKU) that clearly outperforms the others. After conducting and analyzing all statistical results, many scenarios are excluded from the potential list of best performing scenarios. These results were presented to several stakeholders who's input/preferences are used to filter this list down to a top five most promising scenarios for both product types (see section 7.3). Based on the results from this study it is therefore recommended that Ricoh incorporates and subsequently tests these most promising scenarios in order to conclude which scenario is most favourable per product type. This can be done by setting up a pilot phase in which the system calculates multiple forecast based on these different scenarios for which it monitors real-time forecast accuracy. This is especially recommended for the supply product type, since this study showed signs that multiple scenarios other than the currently used one improve the system's forecast performance. Furthermore, another benefit of such a pilot, is that it provides Ricoh with insights into potential technical difficulties from an implementation perspective.

The final business findings are focused on the reviewing process as discussed in the previous chapter. First of all, it can be concluded that within the context of Ricoh's business, no matter how much the system generated forecast are improved, there will always be a need for human interpretation to some extent. This due to the fact that future demand streams are impacted by certain factors that cannot be captured by any statistical models. Such factors are for instance related to Ricoh's tender business, promotional activities, new product introductions, supply chain disruptions and others. Furthermore, based on the output as described in chapter 8, it is concluded that there is much more potential within the data that is currently used by demand planners. More specifically, there might be some data insights within the historical demand volumes that will support a planner during his/her reviewing process. This could be for instance insights into the demand distributions across different sales channels or different regions/OpCo's. These insights are very much associated to the different dimensions that are defined for the hierarchical forecasting topic and that is why Ricoh is recommended to devote time and attention to look into insights across different levels of aggregation that might be of value for demand planners to see. However, to prevent providing a planner with an overload of information during the reviewing process it is recommended to create a demand view in which the aggregated levels are shown first, but allows the planners to zoom in on more of a disaggregated level (e.g., such as demand distributed per OpCo or each sales channel). These insights will support a planner in trying to better understand historical demand volumes and therefore result in adjusting forecast values based on more data driven reasoning. Furthermore, it might also improve the level of discussions that are held with marketing.

Perhaps the most important conclusion for the forecast process in general is that it is highly recommended that Ricoh starts measuring and monitoring forecast accuracy for both their system generated forecast and the final proposed forecast. In doing so, Ricoh can better analyze and conclude what kind of SKU's are already forecasted with a high precision by the system and more importantly Ricoh can measure the impact of the adjustments made by demand planners. The latter is often measured through the forecast value add (FVA) metric which indicates to what extent manual interference have positively or negatively impacted initial forecast performance. Within this topic specifically there are even papers that have constructed specific formulas that are aimed at providing insight into what kind of human bias is most likely the cause of wrongly assumed adjustments. All of this potential information will help Ricoh to better understand when to rely on a system, while it will also offers potential to better inform/train demand planners during their forecast reviewing process.

To conclude, it is strongly recommended that Ricoh takes into account all of the above mentioned findings during their implementation and optimization of their new forecasting system. The implementation and set-up of such a new system is the perfect opportunity to incorporate many of these recommendations.

9.3 Limitations

This section is devoted to describing some of the relevant drawbacks that this research includes. First up, all of the forecast output gathered throughout this study are analyzed through different statistical metrics and approaches. While this is not wrong, studies focused on forecasting often also consider systematical bias. Systematic forecast bias is considered an important piece of information since it provides insight into whether the system has a consistent tendency to over-forecast or under-forecast. Statistical metrics such as the MAPE and the CV* will always be correlated to some extent, while this is much less the case for a metric indicating systematical bias. This could have potentially provided more insight into the differences between the most promising scenarios. It should therefore be considered an important research limitation by both academics and practitioners.

The second limitation of this research is that all input data is only corrected on outliers that were spotted by the system. In a normal business situation, such data might be of better quality, since data is further pre-processed/cleaned by employees. As a result, scenarios focused on individual time series might perform better since data in a normal business setting most likely includes less individual noise. However, within the current situation of Ricoh it is indicated that this data cleaning process is also not very established and mostly depends on individual initiatives. That is why it is argued that the obtained results are still very relevant and applicable to the situation of Ricoh's demand planning department.

The final limitation is that it can be argued that the obtained results originate from a very static approach. Forecast performance of all scenarios are based on a year worth of forecast values obtained during a single fixed moment. Calculating multiple forecast across a longer time-span from which during each month forecast performance is evaluated might be considered a more trustworthy and valid approach within this study. However, this would have meant that more historical data was required, while the calculation of multiple forecast would also have drastically increased run-time of the used forecast model.

To cover for this limitation it is suggested to set-up a pilot in which all promising scenarios are tested across an extended duration (see section 9.2).

9.4 Future research

The final section of this thesis is dedicated to sharing interesting directions for future research. These ideas are both focused on opportunities defined for Ricoh specifically and on ideas that might contribute to the academical area of demand forecasting.

In this study, all hierarchical classes are defined based on well known hierarchical structures found in transactional order data. While this is not wrong, it might still be of interest for Ricoh to conduct cluster analysis on this data. Other studies have provided good results based on groups that were constructed based on cluster algorithms. Adopting a cluster algorithm/approach might result in a better segmentation of data since groups are created according to their similarity instead of solely based on pre-defined features such as their geographical location.

Furthermore, this study was primarily aimed at understanding which of the hierarchical scenarios perform well and which do not. While this study provided answers to this question, the model that generated a forecast for each scenario also provided much more detailed output. For instance, one of the outputs that is generated provides insight into the performance within each hierarchical class (example shared in Appendix A.1). It might be of interest for Ricoh to conduct more detailed analyses on this kind of output for the scenarios that are considered most promising. Doing this will provide Ricoh with an understanding which type of clusters contribute to a reliable forecast and more importantly which do not. For example, it might be the case that the average forecast performance of all OpCo clusters together is very much negatively impacted by one of the clusters specifically (e.g., Western-Europe). Such findings pinpoint which specific clusters still have much potential to be improved and improving those will benefit the average performance of the entire class/scenario.

Finally, since many different scenarios showed similar performance, reconciling them into a single optimal forecast as discussed in the literature review in chapter 2 might be of interest. This approach is focused on obtaining forecast from different levels of aggregation and reconciling all of these into a single optimal forecast by using a regression model. In doing so, it is argued that the forecast is strengthened since it fully utilizes all available information from different levels of aggregation. Moreover, empirical research already provided some promising results in which this specific approach was compared to forecast that are derived on a single level of aggregation.

References

Adesola, S., & Baines, T. (2005). Developing and evaluating a methodology for business process improvement. *Business process management journal*.

Albarune, A. R. B., & Habib, M. M. (2015). A study of forecasting practices in supply chain management. *International Journal of Supply Chain Management (IJSCM)*, *4*(2), 55–61.

Alvarado-Valencia, J., Barrero, L. H., Önkal, D., & Dennerlein, J. T. (2017). Expertise, credibility of system forecasts and integration methods in judgmental demand forecasting. *International Journal of Forecasting*, *33*(1), 298–313.

Alvarado-Valencia, J. A., & Barrero, L. H. (2014). Reliance, trust and heuristics in judgmental forecasting. *Computers in human behavior*, *36*, 102–113.

Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners* (Vol. 30). Springer.

Armstrong, J. S., & Fildes, R. (2006). Making progress in forecasting. *International Journal of Forecasting*, *22*(3), 433–441.

Armstrong, J. S., & Forecasting, L.-R. (1985). From crystal ball to computer. *New York ua*, *348*.

Arvan, M., Fahimnia, B., Reisi, M., & Siemsen, E. (2019). Integrating human judgement into quantitative forecasting methods: A review. *Omega*, *86*, 237–252.

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, *25*(1), 146–166.

Athanasopoulos, G., & Hyndman, R. J. (2008). Modelling and forecasting australian domestic tourism. *Tourism Management*, *29*(1), 19–31.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, *262*(1), 60–74.

Babai, M. Z., Boylan, J. E., & Rostami-Tabar, B. (2022). Demand forecasting in supply chains: a review of aggregation and hierarchical approaches. *International Journal of Production Research*, *60*(1), 324–348.

Carbone, R., Andersen, A., Corriveau, Y., & Corson, P. P. (1983). Comparing for different time series methods the value of technical expertise individualized analysis, and judgmental adjustment. *Management Science*, *29*(5), 559–566.

Chau, N. N. (2020). *Intermittent demand forecasting for inventory control: the impact of temporal and cross-sectional aggregation* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Chen, A., & Blue, J. (2010). Performance analysis of demand planning approaches for aggregating, forecasting and disaggregating interrelated demands. *International Journal of Production Economics*, *128*(2), 586–602.

Dangerfield, B. J., & Morris, J. S. (1992). Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International journal of forecasting*, *8*(2), 233–241.

Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to sku-level demand forecasts. *International Journal of Forecasting*, *29*(3), 510–522.

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, *22*(3), 443–473.

Di Fonzo, T., & Girolimetto, D. (2023). Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives. *International Journal of Forecasting*, *39*(1), 39–57.

Duncan, G. T., Gorr, W. L., & Szczypula, J. (2001). Forecasting analogous time series. In *Principles of forecasting* (pp. 195–213). Springer.

Dunn, D. M., Williams, W. H., & DeChaine, T. (1976). Aggregate versus subaggregate models in local area forecasting. *Journal of the American Statistical Association*, *71*(353), 68–71.

Eggleton, I. R. (1982). Intuitive time-series extrapolation. *Journal of Accounting Research*, 68–102.

Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, *26*(1), 116–133.

Fildes, R., & Goodwin, P. (2007). Against your better judgment? how organizations can improve their use of management judgment in forecasting. *Interfaces*, *37*(6), 570–576.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International journal of forecasting*, *25*(1), 3–23.

Fildes, R., Goodwin, P., & Önkal, D. (2019). Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*, *35*(1), 144–156.

Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, *38*(4), 1283–1318.

Fliedner, G. (2001). Hierarchical forecasting: issues and use guidelines. *Industrial Management & Data Systems*.

Franses, P. H., & Legerstee, R. (2009). Properties of expert adjustments on model-based sku-level forecasts. *International Journal of Forecasting*, *25*(1), 35–47.

Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of forecasting*, *9*(3), 233–254.

Harvey, N. (2001). Improving judgment in forecasting. *Principles of forecasting: A handbook for researchers and practitioners*, 59–80.

Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management science*, *27*(2), 115–138.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational statistics & data analysis*, *55*(9), 2579–2589.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Jin, Y. Williams, B. D., Tokar, T., & Waller, M. A. (2015). Forecasting with temporally aggregated demand signals in a retail supply chain. *Journal of Business Logistics*, *36*(2), 199–211.

Kalchschmidt, M. (2012). Best practices in demand forecasting: Tests of universalistic, contingency and configurational theories. *International journal of production economics*, *140*(2), 782–793.

Kerkkänen, A., et al. (2010). *Improving demand forecasting practices in the industrial context*. Lappeenranta University of Technology.

Kourentzes, N., & Athanasopoulos, G. (2019). Cross-temporal coherent forecasts for australian tourism. *Annals of Tourism Research*, *75*, 393–409.

Kourentzes, N., Rostami-Tabar, B., & Barrow, D. K. (2017). Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels? *Journal of Business Research*, *78*, 1–9.

Lapide, L., & LaPiDe, L. (2009). History to demand-driven forecasting. *Journal of Business Forecasting Methods and Systems*, *28*(2), 18.

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.

Lei, M., Yin, Z., Li, S., & Tan, Q. (2017). Intermittent demand forecasting and inventory control with multiple temporal and cross-sectional aggregation and disaggregation methods. In *2017 13th international conference on natural computation, fuzzy systems and knowledge discovery (icnc-fskd)* (pp. 1969–1977).

Mahalakshmi, G., Sridevi, S., & Rajaram, S. (2016). A survey on forecasting of time series data. In *2016 international conference on computing technologies and intelligent data engineering (icctide'16)* (pp. 1–8).

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting.* John Wiley & Sons.

Nikolopoulos, K. (2021). We need to talk about intermittent demand forecasting. *European Journal of Operational Research*, *291*(2), 549–559.

O'Connor, M., Remus, W., & Griggs, K. (1993). Judgemental forecasting in times of change. *International Journal of Forecasting*, *9*(2), 163–172.

Ouwehand, P. (2006). Forecasting with group seasonality.

O'Connor, M., Remus, W., & Griggs, K. (2000). Does updating judgmental forecasts improve forecast accuracy? *International Journal of Forecasting*, *16*(1), 101–109.

Pennings, C. L., & Van Dalen, J. (2017). Integrated hierarchical forecasting. *European Journal of Operational Research*, *263*(2), 412–418.

Perera, H. N., Hurley, J., Fahimnia, B., & Reisi, M. (2019). The human factor in supply chain forecasting: A systematic review. *European Journal of Operational Research*, *274*(2), 574–600.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., ... others (2022). Forecasting: theory and practice. *International Journal of Forecasting*.

Rostami-Tabar, B., Babai, M. Z., Ali, M., & Boylan, J. E. (2016). On the performance of temporal demand aggregation when optimal forecasting is used. In *International conference on information systems, logistics and supply chain.*

Rostami-Tabar, B., Babai, M. Z., Ducq, Y., & Syntetos, A. (2015). Non-stationary demand forecasting by cross-sectional aggregation. *International Journal of Production Economics*, *170*, 297–309.

Rostami-Tabar, B., Babai, M. Z., & Syntetos, A. (2022). To aggregate or not to aggregate: Forecasting of finite autocorrelated demand. *Journal of the Operational Research Society*, 1–20.

Rostami-Tabar, B., Babai, M. Z., Syntetos, A., & Ducq, Y. (2013). Demand forecasting by temporal aggregation. *Naval Research Logistics (NRL)*, *60*(6), 479–498.

Sanders, N. R., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making*, *5*(1), 39–52.

Sanders, N. R., & Ritzman, L. P. (1995). Bringing judgment into combination forecasts. *Journal of Operations Management*, *13*(4), 311–321.

Sanders, N. R., & Ritzman, L. P. (2004). Integrating judgmental and quantitative forecasts: methodologies for pooling marketing and operations information. *International Journal of Operations & Production Management*.

Sbrana, G., & Silvestrini, A. (2013). Forecasting aggregate demand: analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework. *International Journal of Production Economics*, *146*(1), 185–198.

Schwarzkopf, A. B., Tersine, R. J., & Morris, J. S. (1988). Top-down versus bottom-up forecasting strategies. *The International Journal Of Production Research*, *26*(11), 1833–1843.

Seifert, M., Siemsen, E., Hadida, A. L., & Eisingerich, A. B. (2015). Effective judgmental forecasting in the context of fashion products. *Journal of Operations Management*, *36*, 33–45.

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, *252*(1), 1–26.

van der Staak, B. B. J., & Johanna, P. (2021). Adjust or accept suggestions by algorithms: Uncovering human behavior in forecasting and inventory planning.

Widiarta, H., Viswanathan, S., & Piplani, R. (2007). On the effectiveness of top-down strategy for forecasting autoregressive demands. *Naval Research Logistics (NRL)*, *54*(2), 176–188.

Zotteri, G., Kalchschmidt, M., & Caniato, F. (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, *93*, 479–491.

Appendix A

A.1 Examples of 2nd and 3rd model's output

Item Number	Date	Actuals	SKU_AllCustomers _EMEA	SKU_AllCustomers _OpCoCluster	SKU_AllCustomers _OpCo	SKU_Dir/Indir _EMEA
404866	*Jan-19*	110	95	102	98	89
	Feb-19	53	90	87	94	86
	Mar-19	105	102	102	99	94
	Apr-19	83	88	89	90	71
	May-19	73	73	81	91	65
	Jun-19	80	91	92	95	76
	Jul-19	100	88	94	88	80
	Aug-19	85	91	86	82	69
	Sep-19	97	77	85	89	61
	Oct-19	78	82	84	86	64
	Nov-19	95	92	91	88	63
	Dec-19	98	113	114	94	82
404867	*Jan-19*	14	25	26	25	28
	Feb-19	33	24	19	21	18
	Mar-19	17	16	15	22	18
	Apr-19	20	19	15	18	19
	May-19	15	10	8	18	16
	Jun-19	20	22	14	19	20
	Jul-19	21	9	4	18	11
	Aug-19	14	12	8	19	16
	Sep-19	15	17	13	21	21
	Oct-19	9	15	12	23	19
	Nov-19	11	16	11	19	19
	Dec-19	22	18	7	18	19

Hierachical item	Hierachical conclusion	Best fit	Hierachical item	Hierachical conclusion	Best fit
SKU_AllCustomers_EMEA			*SKU_AllCustomers_OpCoCluster*		
404866_AllCustomers_EMEA	Adequate_data	HWA	404866_AllCustomers_Eastern_Europe	Adequate_data	DES
404867_AllCustomers_EMEA	Adequate_data	HWM	404866_AllCustomers_International	Crostons	CRST
407745_AllCustomers_EMEA	Adequate_data	HWA	404866_AllCustomers_MEA	Crostons	CRST
407884_AllCustomers_EMEA	Adequate_data	HWA	404866_AllCustomers_Northen_Europe	Adequate_data	HWA
408064_AllCustomers_EMEA	Adequate_data	HWM	404866_AllCustomers_Southern_Europe	Adequate_data	HWM
408267_AllCustomers_EMEA	Adequate_data	HWM	404866_AllCustomers_Western_Europe	Adequate_data	HWM
408273_AllCustomers_EMEA	Adequate_data	HWM	404867_AllCustomers_Eastern_Europe	Crostons	CRST
408302_AllCustomers_EMEA	Adequate_data	HWA	404867_AllCustomers_International	Crostons	CRST
408329_AllCustomers_EMEA	Adequate_data	HWA	404867_AllCustomers_MEA	Crostons	CRST
408335_AllCustomers_EMEA	Adequate_data	HWM	404867_AllCustomers_Northen_Europe	Adequate_data	HWM
409074_AllCustomers_EMEA	Adequate_data	HWM	404867_AllCustomers_Southern_Europe	Crostons	CRST
409076_AllCustomers_EMEA	Adequate_data	HWA	404867_AllCustomers_Western_Europe	Adequate_data	HWA

A.2 Forecast performance of all scenarios

Hierarchical Scenario	Machines MAPE	Machines Std.Dev.	Machines CV*	Supplies MAPE	Supplies Std.Dev.	Supplies CV*
SKU_AllCustomers_EMEA	35.3%	15.8%	40.2%	31.3%	25.1%	34.2%
SKU_AllCustomer_OpCoCluster	36.9%	19.8%	41.6%	28.0%	21.4%	30.4%
SKU_AllCustomer_OpCo	40.3%	21.8%	41.3%	26.2%	21.2%	28.4%
SKU_Dir/Indir_EMEA	37.1%	17.8%	39.9%	27.8%	19.7%	30.8%
SKU_Dir/Indir_OpCoCluster	37.8%	20.6%	40.8%	25.8%	18.8%	28.6%
SKU_Dir/Indir_OpCo	40.8%	22.0%	42.3%	27.1%	26.1%	30.1%
SKU_Channel_EMEA	38.2%	17.9%	40.3%	26.3%	18.5%	29.0%
SKU_Channel_OpCoCluster	41.9%	25.0%	44.1%	26.5%	22.0%	28.8%
SKU_Channel_OpCo	43.0%	25.8%	42.4%	32.3%	39.3%	32.5%
SubType_AllCustomer_EMEA	40.2%	23.2%	43.0%	81.6%	32.3%	82.9%
SubType_AllCustomer_OpCoCluster	38.2%	16.7%	40.6%	44.3%	28.9%	44.3%
SubType_AllCustomer_OpCo	38.5%	22.1%	40.2%	60.8%	44.1%	57.2%
SubType_Dir/Indir_EMEA	39.7%	21.1%	41.0%	62.2%	45.1%	58.3%
SubType_Dir/Indir_OpCoCluster	38.6%	17.8%	40.9%	59.1%	42.6%	55.9%
SubType_Dir/Indir_OpCo	38.1%	19.0%	40.2%	59.5%	42.7%	56.0%
SubType_Channel_EMEA	35.7%	15.9%	39.1%	56.7%	39.8%	53.6%
SubType_Channel_OpCoCluster	38.9%	20.7%	40.5%	54.5%	38.5%	51.8%
SubType_Channel_OpCo	39.6%	22.3%	40.2%	54.9%	38.9%	52.2%
Family_AllCustomer_EMEA	37.7%	18.9%	41.2%	34.4%	22.8%	37.5%
Family_AllCustomer_OpCoCluster	37.3%	17.6%	40.5%	31.2%	18.2%	34.2%
Family_AllCustomer_OpCo	35.5%	17.4%	39.8%	30.7%	18.3%	33.5%
Family_Dir/Indir_EMEA	36.5%	18.8%	39.8%	31.8%	18.8%	35.0%
Family_Dir/Indir_OpCoCluster	37.7%	18.9%	40.2%	30.6%	18.1%	33.6%
Family_Dir/Indir_OpCo	36.3%	17.7%	39.8%	29.9%	18.0%	32.5%
Family_Channel_EMEA	36.0%	18.8%	38.9%	34.7%	20.1%	36.6%
Family_Channel_OpCoCluster	39.4%	21.2%	42.5%	32.3%	18.5%	34.6%
Family_Channel_OpCo	40.9%	24.1%	41.9%	31.1%	18.8%	32.9%
Line_AllCustomer_EMEA	39.8%	23.5%	42.7%	65.2%	35.9%	64.6%
Line_AllCustomer_OpCoCluster	38.8%	22.4%	41.7%	54.0%	39.1%	51.8%
Line_AllCustomer_OpCo	36.8%	20.7%	39.9%	55.6%	41.5%	52.9%
Line_Dir/Indir_EMEA	38.0%	22.4%	41.1%	53.1%	40.5%	51.5%
Line_Dir/Indir_OpCoCluster	39.2%	22.9%	41.3%	52.0%	38.0%	50.3%
Line_Dir/Indir_OpCo	37.0%	20.6%	40.1%	53.2%	39.5%	51.2%
Line_Channel_EMEA	38.8%	22.9%	42.0%	52.1%	37.6%	50.0%
Line_Channel_OpCoCluster	36.3%	20.0%	39.7%	48.8%	35.1%	47.5%
Line_Channel_OpCo	38.9%	22.1%	42.5%	49.7%	36.3%	48.3%

A.3 Statistical t-test results for Machines

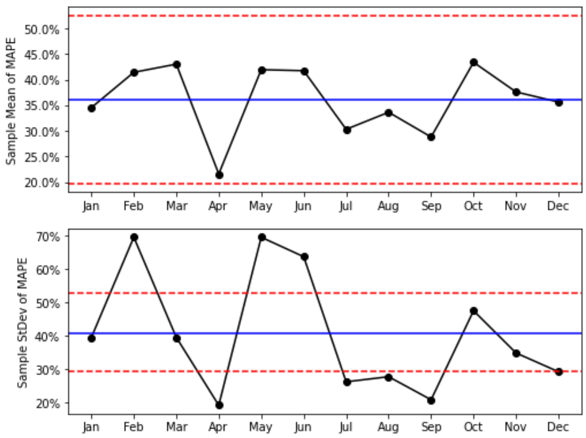
Hierarchical Scenario	MAPE	Out perform	No signficant difference	Under perform	CV*	Out perform	No signficant difference	Under perform
SKU_AllCustomer_EMEA	35.33%	6	29	0	40.23%	0	35	0
SKU_AllCustomer_OpCoCluster	36.87%	2	33	0	41.60%	0	35	0
SKU_AllCustomer_OpCo	40.33%	0	32	3	41.35%	0	35	0
SKU_Dir/Indir_EMEA	37.09%	1	34	0	39.94%	0	35	0
SKU_Dir/Indir_OpCoCluster	37.78%	0	35	0	40.82%	0	35	0
SKU_Dir/Indir_OpCo	40.82%	0	30	5	42.30%	0	35	0
SKU_Channel_EMEA	38.15%	1	34	0	40.29%	0	35	0
SKU_Channel_OpCoCluster	41.92%	0	26	9	44.11%	0	35	0
SKU_Channel_OpCo	42.98%	0	20	15	42.40%	0	35	0
SubType_AllCustomer_EMEA	40.21%	0	29	6	43.01%	0	28	7
SubType_AllCustomer_OpCoCluster	38.24%	0	34	1	40.65%	0	35	0
SubType_AllCustomer_OpCo	38.52%	1	33	1	40.22%	2	33	0
SubType_Dir/Indir_EMEA	39.70%	0	32	3	41.02%	0	35	0
SubType_Dir/Indir_OpCoCluster	38.62%	0	34	1	40.91%	0	35	0
SubType_Dir/Indir_OpCo	38.05%	0	35	0	40.20%	0	35	0
SubType_Channel_EMEA	35.66%	13	22	0	39.10%	3	32	0
SubType_Channel_OpCoCluster	38.90%	0	33	2	40.45%	0	35	0
SubType_Channel_OpCo	39.65%	0	28	7	40.18%	1	34	0
Family_AllCustomer_EMEA	37.67%	1	34	0	41.22%	0	34	1
Family_AllCustomer_OpCoCluster	37.27%	1	34	0	40.50%	0	35	0
Family_AllCustomer_OpCo	35.46%	9	26	0	39.84%	0	35	0
Family_Dir/Indir_EMEA	36.46%	5	30	0	39.80%	1	34	0
Family_Dir/Indir_OpCoCluster	37.67%	1	34	0	40.22%	0	35	0
Family_Dir/Indir_OpCo	36.26%	4	31	0	39.78%	0	35	0
Family_Channel_EMEA	36.02%	6	29	0	38.88%	4	31	0
Family_Channel_OpCoCluster	39.38%	0	35	0	42.52%	0	35	0
Family_Channel_OpCo	40.89%	0	29	6	41.87%	0	35	0
Line_AllCustomer_EMEA	39.84%	0	27	8	42.72%	0	26	9
Line_AllCustomer_OpCoCluster	38.77%	1	33	1	41.71%	1	30	4
Line_AllCustomer_OpCo	36.77%	6	29	0	39.94%	3	32	0
Line_Dir/Indir_EMEA	38.03%	1	34	0	41.05%	2	33	0
Line_Dir/Indir_OpCoCluster	39.17%	0	31	4	41.26%	1	33	1
Line_Dir/Indir_OpCo	37.02%	5	30	0	40.12%	1	34	0
Line_Channel_EMEA	38.77%	0	35	0	42.00%	0	33	2
Line_Channel_OpCoCluster	36.34%	9	26	0	39.68%	6	29	0
Line_Channel_OpCo	38.86%	0	34	1	42.50%	0	34	1

A.4 Statistical t-test results for Supplies

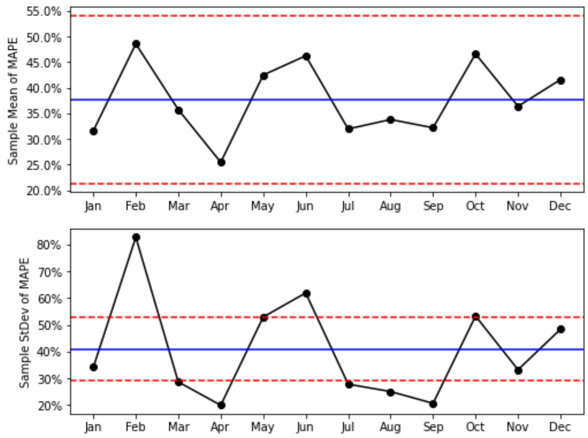
Hierarchical Scenario	MAPE	Out perform	No signficant difference	Under perform	CV*	Out perform	No signficant difference	Under perform
SKU_AllCustomer_EMEA	31.33%	19	9	7	34.20%	19	10	6
SKU_AllCustomer_OpCoCluster	27.98%	27	7	1	30.38%	26	7	2
SKU_AllCustomer_OpCo	26.25%	29	6	0	28.36%	30	5	0
SKU_Dir/Indir_EMEA	27.81%	28	6	1	30.83%	26	5	4
SKU_Dir/Indir_OpCoCluster	25.82%	31	4	0	28.64%	30	5	0
SKU_Dir/Indir_OpCo	27.07%	28	7	0	30.11%	20	15	0
SKU_Channel_EMEA	26.33%	29	6	0	28.98%	29	6	0
SKU_Channel_OpCoCluster	26.52%	29	6	0	28.77%	29	6	0
SKU_Channel_OpCo	32.29%	18	10	7	32.52%	19	16	0
SubType_AllCustomer_EMEA	81.61%	0	0	35	82.90%	0	0	35
SubType_AllCustomer_OpCoCluster	44.33%	17	0	18	44.33%	16	1	18
SubType_AllCustomer_OpCo	60.75%	2	1	32	57.16%	3	0	32
SubType_Dir/Indir_EMEA	62.21%	1	1	33	58.26%	2	0	33
SubType_Dir/Indir_OpCoCluster	59.07%	4	1	30	55.87%	4	1	30
SubType_Dir/Indir_OpCo	59.53%	4	1	30	56.00%	4	1	30
SubType_Channel_EMEA	56.66%	6	1	28	53.56%	6	2	27
SubType_Channel_OpCoCluster	54.51%	8	5	22	51.76%	8	6	21
SubType_Channel_OpCo	54.93%	7	4	24	52.15%	7	5	23
Family_AllCustomer_EMEA	34.44%	18	4	13	37.50%	18	1	16
Family_AllCustomer_OpCoCluster	31.16%	20	7	8	34.18%	20	7	8
Family_AllCustomer_OpCo	30.71%	21	7	7	33.54%	22	6	7
Family_Dir/Indir_EMEA	31.79%	20	6	9	34.98%	20	5	10
Family_Dir/Indir_OpCoCluster	30.65%	22	6	7	33.64%	21	7	7
Family_Dir/Indir_OpCo	29.92%	25	6	4	32.47%	25	6	4
Family_Channel_EMEA	34.67%	18	2	15	36.62%	18	3	14
Family_Channel_OpCoCluster	32.26%	19	5	11	34.58%	20	6	9
Family_Channel_OpCo	31.13%	21	6	8	32.86%	23	8	4
Line_AllCustomer_EMEA	65.19%	1	2	32	64.56%	1	0	34
Line_AllCustomer_OpCoCluster	54.00%	8	3	24	51.78%	8	4	23
Line_AllCustomer_OpCo	55.58%	6	3	26	52.93%	6	3	26
Line_Dir/Indir_EMEA	53.12%	8	5	22	51.50%	7	5	23
Line_Dir/Indir_OpCoCluster	52.04%	13	1	21	50.29%	11	3	21
Line_Dir/Indir_OpCo	53.24%	9	3	23	51.24%	8	4	23
Line_Channel_EMEA	52.13%	11	3	21	50.01%	12	2	21
Line_Channel_OpCoCluster	48.78%	16	0	19	47.53%	16	1	18
Line_Channel_OpCo	49.74%	15	0	20	48.32%	15	0	20

A.5 Xbar-S charts for Machine scenarios

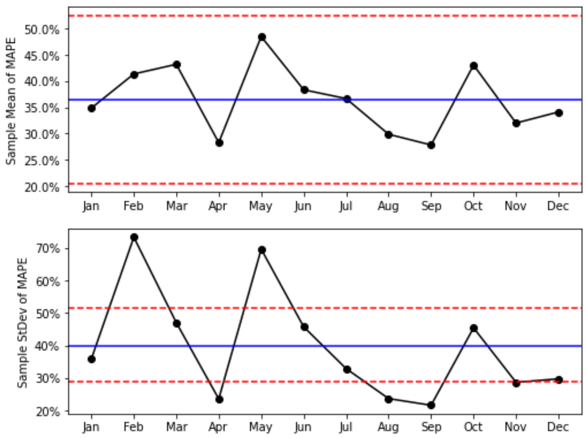
Xbar-S Chart of MAPE of Scenario: SKU_AllCustomer_EMEA



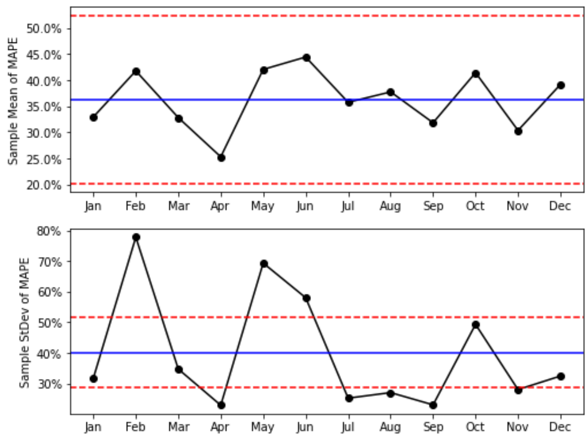
Xbar-S Chart of MAPE of Scenario: SKU_AllCustomer_OpCoCluster



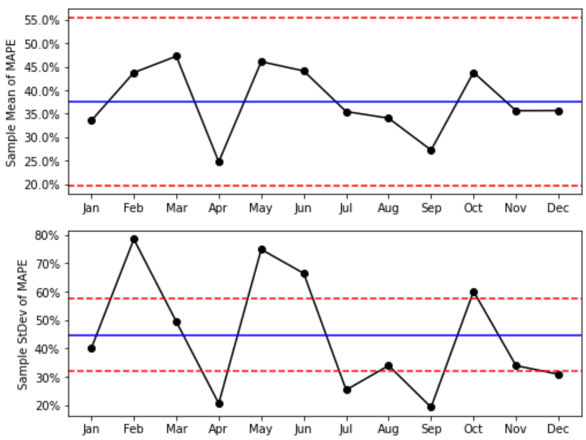
Xbar-S Chart of MAPE of Scenario: SubType_Channel_EMEA



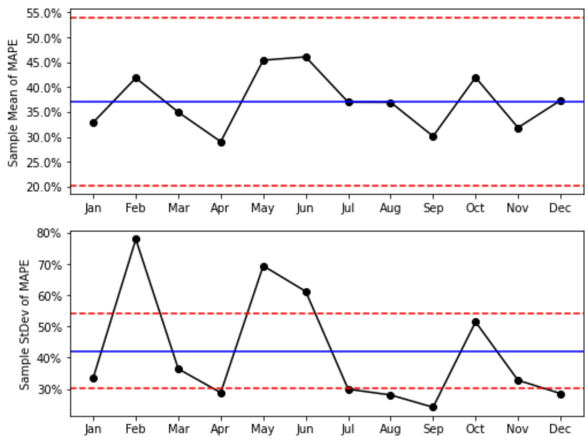
Xbar-S Chart of MAPE of Scenario: Family_AllCustomer_OpCo



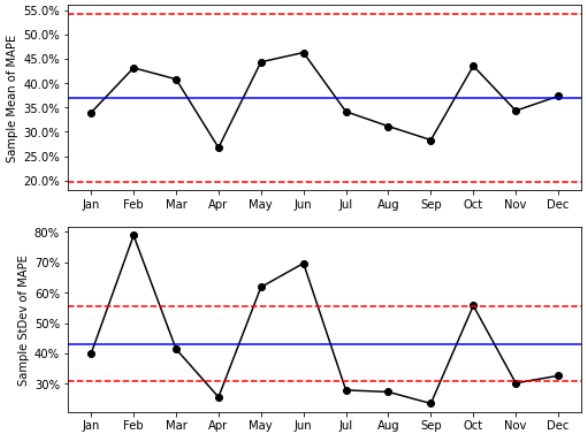
Xbar-S Chart of MAPE of Scenario: Family_Dir/Indir_EMEA



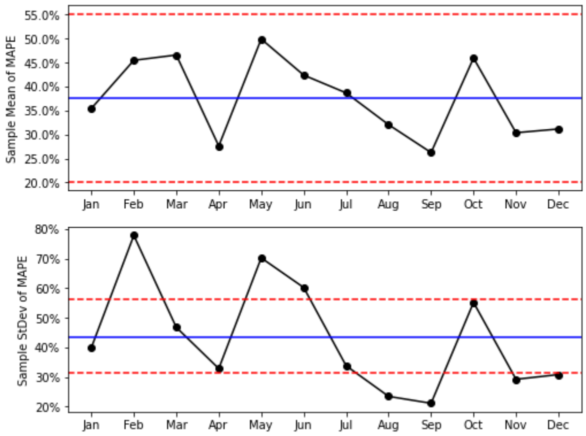
Xbar-S Chart of MAPE of Scenario: Family_Dir/Indir_OpCo



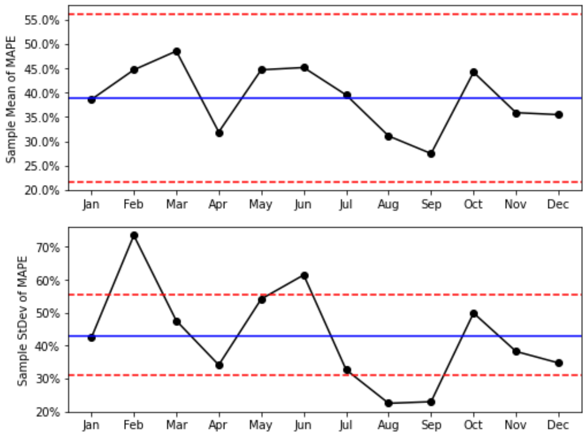
Xbar-S Chart of MAPE of Scenario: Family_Channel_EMEA



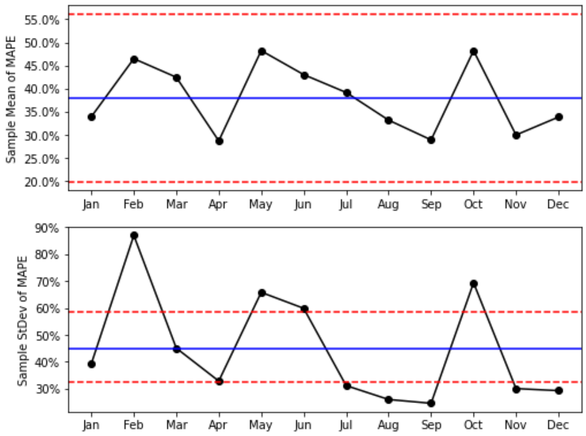
Xbar-S Chart of MAPE of Scenario: Line_ AllCustomer_OpCo



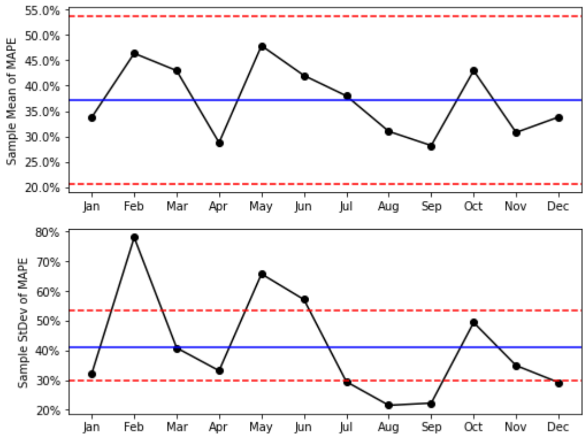
Xbar-S Chart of MAPE of Scenario: Line_ Dir/Indir_EMEA



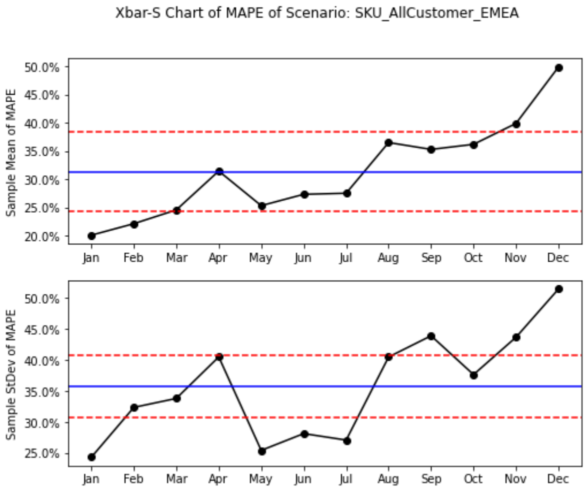
Xbar-S Chart of MAPE of Scenario: Line_ Dir/Indir_OpCo

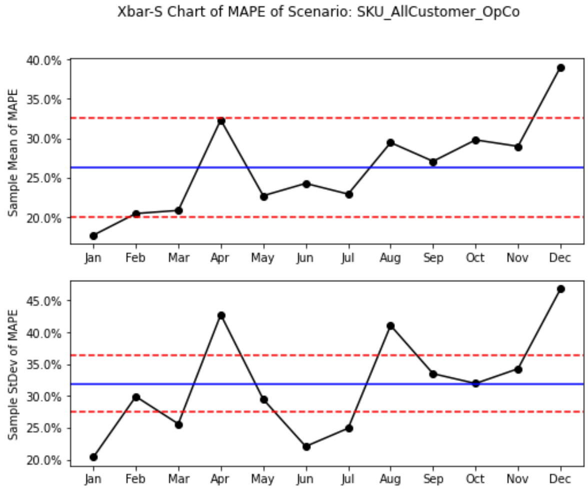


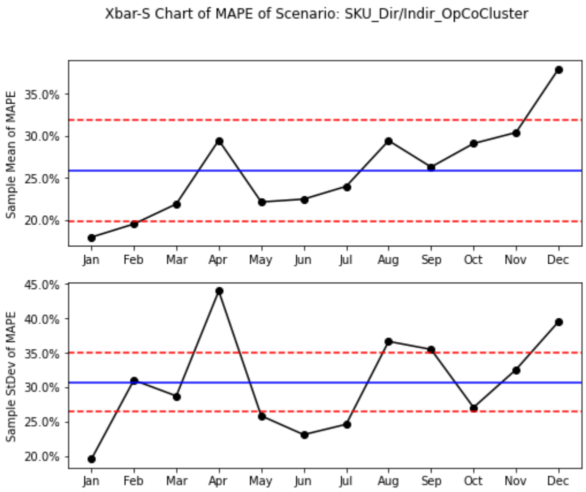
Xbar-S Chart of MAPE of Scenario: Line_ Channel_OpCoCluster

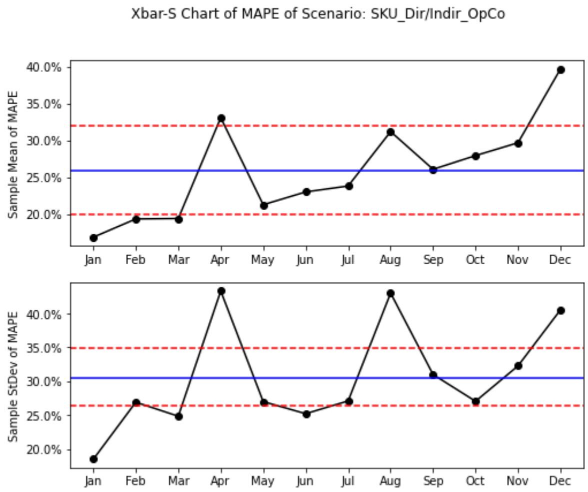


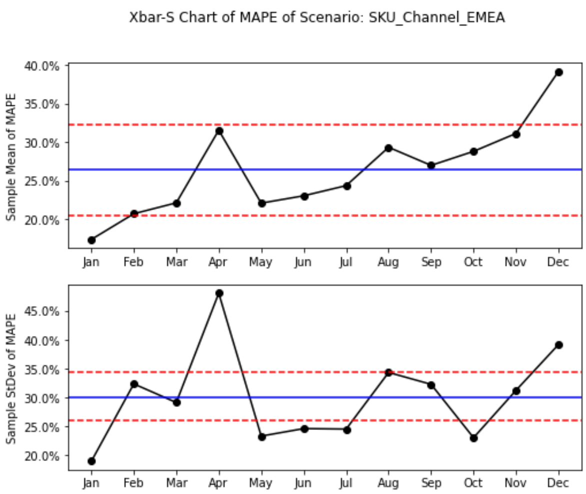
A.6 Xbar-S charts for Supply scenarios

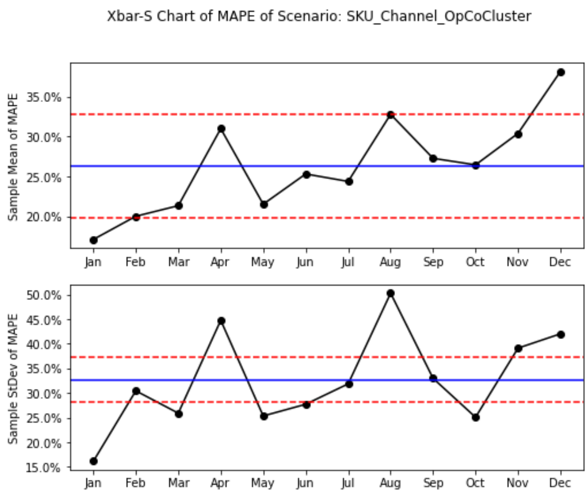




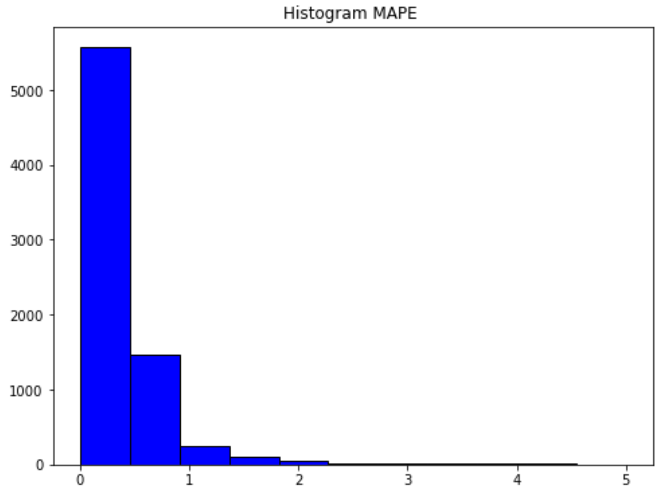


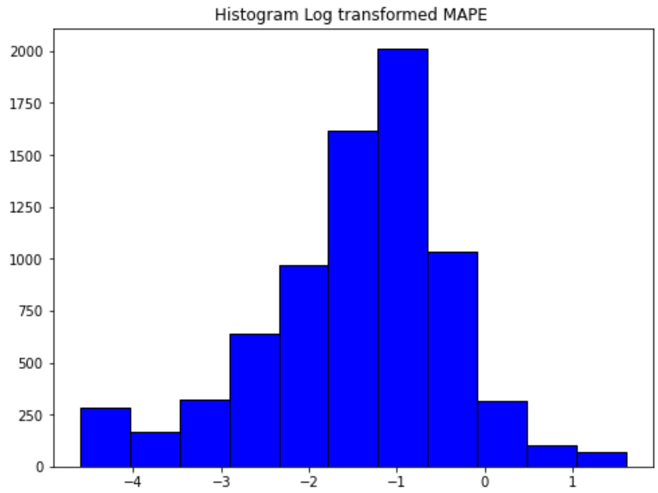


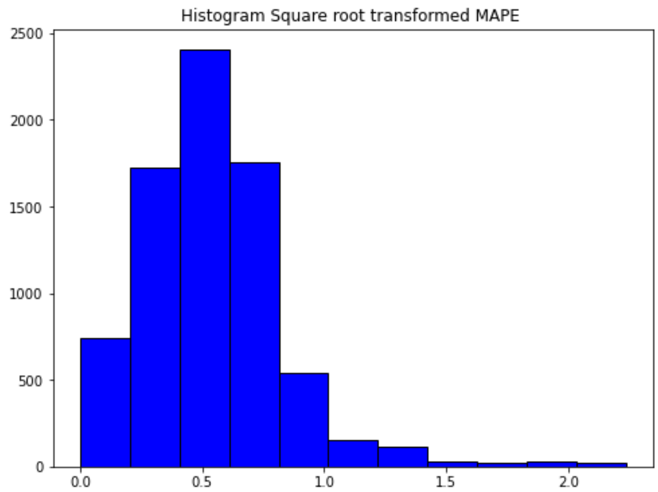




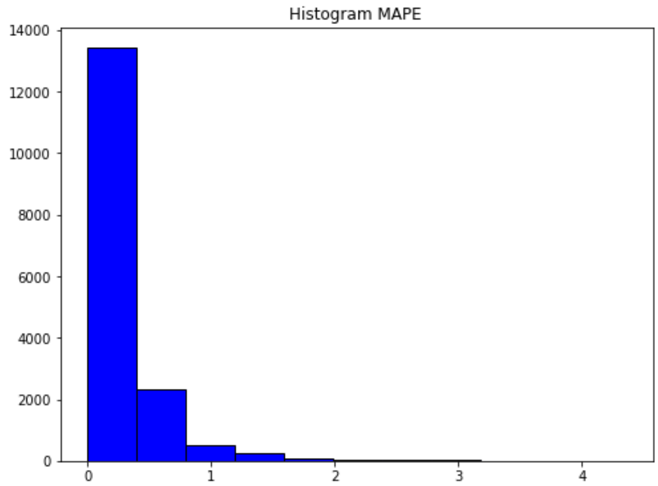
A.7 Data transformation techniques (Machines)

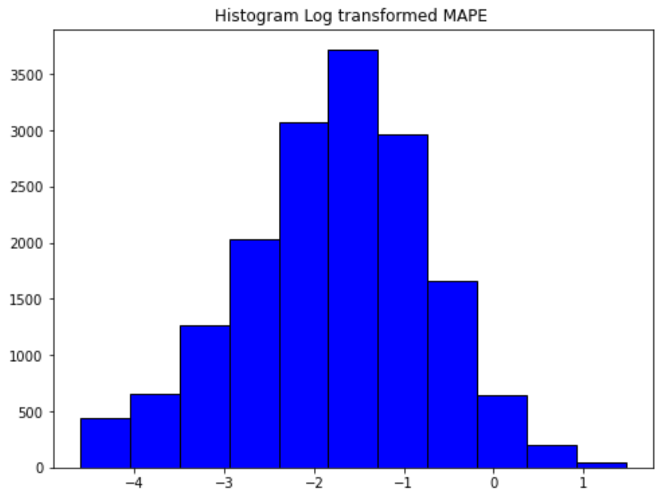


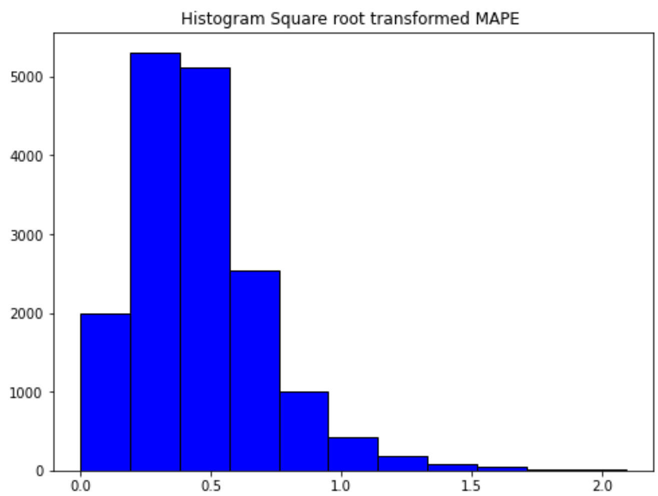




A.8 Data transformation techniques (Supplies)







A.9 Judgmental forecasting questionnaire

1. For how many years have you been working at Ricoh?

2. For how many years have you been active in planning related jobs?

3. For which product type(s) are you responsible during the forecasting process?

4. How many unique SKUs are within your portfolio?

5. How many of those do you need to manually review on a monthly basis?

6. Approximately how many items have you adjusted during the current demand forecasting week?

7. Provide at least 5 examples for which you **have not** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Example 1:

Example 2:

Example 3:

Example 4:

Example 5:

8. Provide at least 5 examples for which you **have** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Example 1:

Example 2:

Example 3:

Example 4:

Example 5:

9. Are there any other (not yet mentioned) elements/reasons which will result in a planner manually adjusting the system generated forecast?

10. What kind of historical demand driven information is missing and would help support a planner in its forecasting reviewing process?

11. What kind of other information is missing and would help support a planner in its forecasting reviewing process? (In addition to the historical demand information)

12. In addition to historical demand data, is there any other kind of information/data which the system should consider when calculating/suggesting forecast figures?

A.10 Judgemental forecast interviews for Machines

Interview with 1st Machine demand planner

1. For how many years have you been working at Ricoh?

Answer: 2 years

2. For how many years have you been active in planning related jobs?

Answer: 6 years

3. For which product type(s) are you responsible during the forecasting process?

Answer: Machines

4. How many unique SKUs are within your portfolio?

Answer: Approximately 55 items

5. How many of those do you need to manually review on a monthly basis?

Answer: All of them, as we forecast machines using an Excel file

6. Approximately how many items have you adjusted during the current demand forecasting week?

Answer: Between 20 and 25 items

7. Provide at least 5 examples for which you **have not** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: One of the first steps I take before being able to properly forecast is to execute some sort of an outlier detection. Within the system there is a flagging system which indicates if orders were flagged as outliers. However, this is not always 100% accurate, as sometimes a customer has spread out its demand over multiple orders. That's why I manually create a pivot table to investigate demand per customer and check on irregularities. Moreover, there is a build-in warning system which makes a historical data bucket red whenever it deviates more than 30% from the average of that same month throughout the past years. Eventually I must manually remove all identified outliers from the unprocessed dataset within Excel. This is something I update each month again.

From the cleaned historical data, Excel calculates a monthly average and a yearly average. Subsequently, each monthly average is divided by the yearly average representing the seasonality percentage of each month. Next up, I must decide on a single base rate value which is multiplied by each seasonality percentages, resulting into a 12-month forecast. I base the value of the base rate on for instance the yearly averages, which helps me in understanding a potential data trend. To be able to see the changes we made with the base rate value, we use a file. This file states the baseline value from month to month for each item. This is something which is sometimes helpful as it states the changes in baseline values and the reasoning behind the chosen baseline value.

Example 1: EDP 418118 (Bric-MF model)
The baseline was kept stable (same as last month), since the historical data is reliable once you have removed the outliers. The chosen baseline was slightly higher than then averages of the most recent months. I agreed with this as it is now able to handle potential unexpected demand which looking at the number of outliers happens occasionally for this item specifically.

Example 2: All Chimay-MF models
For all the Chimay-MF models the baseline has not been adjusted for several months now. This is because the demand of these models has been impacted by part issues, which caused a lot of shortages during the past year. For that reason, we have adjusted the Excel file such that it only uses historical data from before the shortages occurred (2019-2020) and we expect that sales would still be on this level if those shortages never happened.

Example 3: EDP 418839 (Clover-MF model)
For the Clover-MF model 418839 the baseline was kept stable (same as last month), as the 3-month average and the 6-month average have had little variations for the past months. In the current situation, we must check these figures ourselves, but it would be nice if such information would be automatically shared.

8. Provide at least 5 examples for which you **<u>have</u>** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer:

Example 1: Beluga models
For the Beluga models I have lowered my forecast because these models were mainly sold to Russia, and since the war and the sanctions imposed on that country, sales have been dropping for the last months.

Example 2: EDP 417435 (Leffe-MF mode)
For the Leffe-MF model 417435 I had to manually lower my forecast, as this model has not been selling to the levels we expected after the relaunch in June22. We relaunched this item after there were part issues for some time. We expected sales volume to bounce back to the sales levels before these issues occurred. From the most recent months we can conclude that this expected behavior did not happen. Reasoning for this behavior not happening is not 100% clear, but could probably be explained due to the long duration of part issues and therefore customers losing/shifting their interest.

Example 3: EDP 417378 (Bric-C model)
For the Bric-C model 417378 I had to manually increase my forecast, as this model is very fluctuating, and the historical data is not so reliable (even after removing the outliers).

This may be because the main buyer for this model is RIBV and they have a very erratic demand in general, which is very hard to predict. For example, after removing the outliers, the 3-month average is 203 units, the 6-month average is 150 units and the 12-month average is 186 units, which shows the values really change from month to month. Based on the 2 most recent months it seems like the sales volume equals 200+ and therefore I have chosen my baseline as 200. Moreover, I have increased the two lowest forecast values closer to the baseline value such that all values are more even. This since the seasonality percentages are not reliable due to its erratic behavior.

9. Are there any other (not yet mentioned) elements/reasons which will result in a planner manually adjusting the system generated forecast?

Answer:

1. Corona had a big impact in the way we had to forecast machines. Having to manually adjust the expected forecast to the levels we had before the virus, once the situation returned to "normal", as the historical data was "polluted" due to the low sales we had during that period.

2. Phase-in/out situations as we must be more aware for such changes. For instance, we do not want too much excess stock on items which are going to be phased out soon. For phase-in situations we use historical data from predecessor combined with additional information from our marketing team.

10. What kind of historical demand driven information is missing and would help support a planner in its forecasting reviewing process?

Answer:

1. Identification of patterns / trends given by the system. Would be nice if the system could make suggestions regarding the seasonality months. For instance, the system states to please be aware that over the past 3 years March had a 20% increase due to end of fiscal year.

2. An extra insight on different hierarchical levels would be of added value, especially on OpCo level that I could think of. Because we have domain knowledge on which OpCo's for instance have a more erratic demand than others. So, if we know how historical demand was distributed over the different OpCo's, we can use this information to better explain behaviours and predict our future forecast. As an example, a planner should be warned whenever RIBV orders more than 20% of the total demand (as RIBV is an erratic OpCo in general). Same with the Russian OpCo which was explained in the first example of question 8. For this specific item, I had to be told that the Beluga SKU is sold mainly to Russia.

3. Signalling of potential historical facts that may affect the historical demand data, for example Corona virus, global parts issues, etc. The only reason why I consider such information nowadays is since I know it happened for specific items. Perhaps it is an idea to provide information/warnings on such matters. For instance, some time ago a big fire had happened and impacted some of the factories. If you were not aware of this

event it would be very hard to explain the historical data numbers without being able to connect it to such events. Of course, Covid is worldwide known, but perhaps there are some events/circumstances that you don't know of and would help in explaining historical behaviour.

11. What kind of other information is missing and would help support a planner in its forecasting reviewing process? (In addition to the historical demand information)

Answer:

1. Alerts of phasing in or out products that could potentially affect the forecast review. Of course, we are aware of these situations, but we must look them up in a different file. Would be nice if such information like the predecessor(s) and successor phase-in/out date is automatically shared upfront. By doing so, it would save me some time. Moreover, it would perhaps be a good trigger to know if any SKUs are phasing-out in the same product-family such that you can ask the product manager whether demand is going to shift to and/or affect demand of a different item.

2. Alerts of potential issues related to component shortages. So, this one would be more focused on potential shortages in the (near) future. Once a month we get a file which discusses ongoing and potential future shortages. This kind of information I already consider during my forecast review. For instance, if I know that a certain SKU is going to have shortages within 3 months, then I will probably increase my forecast upon that moment to create some sort of a buffer stock.

3. Easy identification of outliers that should be removed from the baseline forecast. As discussed before, the system already marks some outliers, but does for instance not considers demand at a customer level. For instance, when a customer places 100 orders of 1 unit it could be considered as an outlier. Therefore, it would be of interest to provide information on the how demand is spread out over different customers to potentially identify more outliers.

12. In addition to historical demand data, is there any other kind of information/data which the system should consider when calculating/suggesting forecast figures?

Answer: Perhaps something related to tenders would be of benefit for the system to know upfront.

Interview with 2nd Machine demand planner

1. For how many years have you been working at Ricoh?

Answer: 4 years

2. For how many years have you been active in planning related jobs?

Answer: 4 years

3. For which product type(s) are you responsible during the forecasting process?

Answer: Commercial printing and commercial feed Machines/Options

4. How many unique SKUs are within your portfolio?

Answer: Approximately 1,000 SKUs from which 25 are Machines

5. How many of those do you need to manually review on a monthly basis?

Answer: Approximately 90 SKUs from which 21 are Machines

6. Approximately how many items have you adjusted during the current demand forecasting week?

Answer: 7 Machines

7. Provide at least 5 examples for which you **have not** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: In the current format, forecasting numbers are calculated via an Excel sheet which are based on the base rate and seasonality. Both the base rate as the seasonality percentages are linked to the historical demand quantities.

Seasonality is a monthly percentage which is based on the most recent 3 years of historical demand. These percentages are either automatically calculated (for high volume selling machines) or determined by a demand planner (low volume selling machines). For low sales volume machines, I review and decide those seasonality figures at the start of each year myself. However, the logic behind both methods is almost the same. Moreover, the product families which have these low sales patterns are consolidated on the same seasonality figures. This is done as it does not make sense to determine seasonality figures for each low sales item separately.

The base rate is a figure which is determined by us and reviewed and adjusted if needed on a monthly basis. The reviewing process of this figure is mostly based on the most recent 3M, 6M and 9M historical demand averages.

Example 1: EDP 409395 (Charis Pro C5310S Machine)
The historical demand for this item is quite stable with a 3M average of 17 units/month, a 6M average of 16 units and a 9M average of 16 units. Therefore, I have not changed the base rate for this item and kept the base-rate on an average of 16 units/month.

8. Provide at least 5 examples for which you **have** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: Items are either classified as weekly forecast items or as monthly forecast items. This classification determines whether the calculated forecast figures are sent on a monthly or a weekly basis. The classification is determined by our Japan headquarters and is mainly based on portfolio importance and current sales volume. Furthermore, within this Excel set-up most adjustments are done by changing the base rate figure from last month. Occasionally in specific cases, I change one or more of the 12 forecast values separately (See example 2 and 5).

Example 1: EDP 409368 (Andromeda C9200 Machine)
I have reduced the baseline forecast from an average of 19 units per month to now an average of 17 units per month. I made this change based on the demand of the past 9 months. I calculate the 3M, 6M and 9M average demand for this item and have adjusted the forecast to align on this 3-, 6- and 9-month downwards trend. The previous 3/6- and 9-month average is a piece of information which I often look at to decide whether an adjustment to the base rate is necessary.

Example 2: EDP 409394 (Charis Pro C5300S Machine)
We decided to increase our December until March forecast from 438 to 458 units to make sure we get sufficient stock from the factory to reach 90% of the budget provided by Marketing. This increase of 20 units is to cover the underperformance of November (demand in November was lower than our forecast). So, to conclude this adjustment was mainly pushed from a marketing perspective and expectations.

Example 3: EDP 409474 (Charis Pro Lite C5300L Machine)
During October I reduced the forecast for this item from an average of 49 units/month to an average of 40 units/month based on feedback from Marketing team. We saw that the forecast of 49 units was way too high as actual demand was closer to 20 units/month, so our proposal was to reduce forecast to around 20/month. Based on feedback from Marketing team they advised that with the new promotion on this model it was realistic to expect demand of 40 units/month.

Example 4: New Leo and Andromeda introduction
During November we submitted the first forecast for the new Leo and Andromeda machines that we expect to launch in June'23. This forecast was based on the demand of the predecessor models. The change we made is the following: the first 3 months (June until August) we increased the forecast with 20% based on feedback from Marketing team where they expect some higher demand at the start of the launch. For Leo, this was an increase from 136 to 163 units throughout this 3-month period. For Andromeda, this was an increase from 91 to 109 units.

Example 5: EDP 409241
Forecast of December was 43, while order intake was only 6 items on the 8th of December. Considering the Christmas holiday, this order intake already equals 1/3 of the month and it was not expected to reach the demand of 43 items. Consequently, I have decreased the December forecast with a total of 20 units and increased both February as March with a total of 10 units. Meaning that the total demand was kept stable while changes were made within the 12-month distribution. This is done such that the information flow and requirement calculations with involved factories is kept up to date.

9. Are there any other (not yet mentioned) elements/reasons which will result in a planner manually adjusting the system generated forecast?

Answer:

1. New tender information (tender demand is added separately within the system)
2. Marketing input (Could be budget, promotions, expectations related)
3. Phase-in/out processes (both marketing as predecessor data driven)

10. What kind of historical demand driven information is missing and would help support a planner in its forecasting reviewing process?

Answer: Tender demand. In the current set-up tender demand is included within the historical demand data. The Excel sheet and the human decision maker use this data to determine their forecast figures. In other words, as they are included it could be that the base rate and seasonality are wrongly assumed as tender demand is included within the calculation of these rates. Being able to separate both streams would be very helpful as tender demand is often unusual and thus will not re-occur in the upcoming months/years. During our forecasting process we distinguish usual demand and tender demand, but this distinction is not to be found within the historical demand. Moreover, I believe that most answers/improvements are all related to an accurate outlier detection.

11. What kind of other information is missing and would help support a planner in its forecasting reviewing process? (In addition to the historical demand information)

Answer: At this moment I don't really mis any information. Tender information is available via the RUDI download and Marketing input (for example promotions) can be discussed during the Demand Meeting. In case there are any changes in the Phase-in Phase-out schedule for a certain machine family then this will also be communicated by the supplier, so this information is also available.

12. In addition to historical demand data, is there any other kind of information/data which the system should consider when calculating/suggesting forecast figures?

Answer: Hard to tell as for instance local events do not have a major impact on Ricoh's worldwide future demand patterns. Shifting demand patterns due to stockouts as you suggested could be an interesting factor. However, this is information which would be hard to incorporate as its very subjective information most likely to originate from our marketing team.

Interview with Machine/Supply demand planner

1. For how many years have you been working at Ricoh?

Answer: 4 years

2. For how many years have you been active in planning related jobs?

Answer: 2 years

3. For which product type(s) are you responsible during the forecasting process?

Answer: Supplies & machines

4. How many unique SKUs are within your portfolio?

Answer: 163 (20-25 of these are machines)

5. How many of those do you need to manually review on a monthly basis?

Answer: All of them as the machines are forecasted manually via Excel and the portfolio which I have for supplies need to be reviewed due to the nature of these items (continuous feed supplies). These supplies belong to machines worth several hundred thousand of euros and generally have a long lead-time up to 6 months. Only for very few items I rely on our system forecast.

6. Approximately how many items have you adjusted during the current demand forecasting week?

Answer: +- 100

7. Provide at least 5 examples for which you **have not** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: During the reviewing process I check on whether I see the seasonality patterns as defined within Ricoh. If those high/low peaks are present, I will make sure that the forecasted value also shows the same seasonality behaviour. Furthermore, for stable items I put more attention on the trend line of recent months/year, while for the more volatile items I put more emphasis on the year-on-year averages.

Another aspect I look into is the MIF value for which I monitor whether it is increasing, decreasing or stabilizing. In theory it would be possible to check the MIF value and multiply it with an average print volume. However, this makes the review/forecast process too complex.

Example 1: EDP 828286 (Supply)
Forecast is quite stable, as is historical data. There are two slight peaks due to seasonality,

which are also included within the system generated forecast.

Example 2: EDP 828290 (Supply)
Demand history is very spiky and seems random as it does not show any signs of re-occurring high/low volume months. Moreover, the common seasonality months as identified within Ricoh cannot be found within the historical demand pattern. That's why I agreed with the system suggested forecast as they suggested a straight line around the recent monthly averages. By doing so, we will be able to cover for the peaks, but at the same time will not risk getting any excess stock whenever the sales volumes are lower than forecasted. Furthermore, I have checked the year-on-year average and noticed that the suggested baseline value is around the same as the most recent years (which I agree on). Finally, as it is a relatively expensive machine (which is still active), the MIF will not be very vulnerable for any unexpected abrupt changes.

Example 3: EDP 828544 (Supply)
This is the successor of another supply for which we see that it slowly takes over the demand of our predecessor. The historical data also shows this kind of behaviour and therefore the system suggested a forecast which is also very steadily increasing. On top of that, it also correctly includes the seasonality peak during March.

Example 4: EDP 828545 (Supply)
Is an item which is within the same range as the previous item (different colour) and shows that the same behaviour is often present within an item series. The suggested forecast is once again increasing very steadily which is in line with historical data. The suggested forecast also includes a peak during March.

8. Provide at least 5 examples for which you **<u>have</u>** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: Within our machine forecasting file there are always going to be some automatic tweaks, purely since the previous month is added as the most recent historical demand bucket. Adding these will slightly change the seasonality percentages which will then change the twelve forecast values. In other words, machine adjustments are more about tweaking as major base rate changes are not regular. In addition to determining our forecast values, we also adjust throughout the month (weekly forecast maintenance). These adjustments are often based on the order intake for which we determine whether adjustments are required for month N and N+1. Whenever the order intake is out of proportion to the number of days we have left in the month, we might have to adjust these forecast values.

Example 1: Libera MF
As demand in the previous months has been crashing, I decided to adjust the corresponding base rate drastically. As the previous base-rates were too high, I set the new base rate around the 3,6 and 9-month averages. Exact reasoning behind this crash is hard to isolate but is most probably due to not received tender demand. Our major OpCo's requested much less demand than previously (which could have been the result of our charging price).

In this case, it would have been nice to have insight into how requested demand was distributed among the different OpCo's. Eventually, I had to research this myself and asked product marketing about these 'demand collapsing' OpCo's. If I knew this upfront, I could have reacted on this demand crash much earlier which would have led to much more accurate forecast values. This kind of information would definitely be of interest, especially whenever there are significant changes with respect to these distributions.

Example 2: EDP 905398
We were approaching last buy moment due to end of life of this item at our supplier. So, we are slowly forecasting the last units that are coming in until we have 0 left. To determine the required forecasted amount, we use external information from our supplier analyst who provides use with customer requirements.

Example 3: EDP 905399
We were approaching last buy moment due to end of life of this item at our supplier. So, we are slowly forecasting the last units that are coming in until we have 0 left. To determine the required forecasted amount, we use external information from our supplier analyst who provides use with customer requirements.

Example 4: EDP 910152/910154/910155
We were approaching last buy moment due to end of life of this item at our supplier. So, we are slowly forecasting the last units that are coming in until we have 0 left. To determine the required forecasted amount, we use external information from our supplier analyst who provides use with customer requirements.

Example 5: EDP 910157 (Supply)
This is a supply which will be phased out soon and therefore should have a slowly decreasing trend. The reason why I overruled the system and kept it stable is because we have one customer which has this machine who made it a very clear that they would like to receive significant order volume until the phase out date.

9. Are there any other (not yet mentioned) elements/reasons which will result in a planner manually adjusting the system generated forecast?

Answer:

1. For the items within my portfolio, I have a lot of input from Marketing/Sales/OFD about potential or signed deals. This will have a direct impact on the ink usage. Most of the continuous feed supplies have a long lead-time, up to 6 months. So, I must adjust the forecast timely based on calculation of expected usage. This will be put on top of historical demand data which is related to the MIF.

2. Besides that, end of life of items (phase out) or phase in are also specific situations in which we often need to manually adjust/determine the forecast values.

10. What kind of historical demand driven information is missing and would help support a planner in its forecasting reviewing process?

Answer: In general, I would say that we already have a lot of information at our disposal. I don't think that we need more historical demand insights and that in our case future related information is more relevant.

1. The only insight I could think of is the distribution split of demand per OpCo. This is linked to the first example of question 8.

11. What kind of other information is missing and would help support a planner in its forecasting reviewing process? (In addition to the historical demand information)

Answer: Not all upcoming deals are forecasted at all, or they are but not in time (in RUDI). This is not related to historical demand data but related to data that I need to receive upfront. For instance, it has had happened that we receive a new tender for next month which has a win probability of 95%. This is too short notice, for which we will never be able to react as we need to make sure we receive enough volume to support these kinds of tenders.

12. In addition to historical demand data, is there any other kind of information/data which the system should consider when calculating/suggesting forecast figures?

Answer: Is hard to tell and I cannot think of any factors at this moment in time.

Interview with Machine supervisor

1. For how many years have you been working at Ricoh?

Answer: 12 years

2. For how many years have you been active in planning related jobs?

Answer: 12 years

3. For which product type(s) are you responsible during the forecasting process?

Answer: I'm the supervisor for Office Printing HW (machines & options) + Communication Services (Whiteboards, meeting rooms etc).

4. How many unique SKUs are within your portfolio?

Answer: 2001 (Active, PO-Hold or Prelaunch items) + 1810 B2B items

5. How many of those do you need to manually review on a monthly basis?

Answer: I only review the machine forecast which are 226 EDP codes. From these machines I either check them on an aggregated level such as product-family level or check them individually (applicable for the more important machines).

6. Approximately how many items have you adjusted during the current demand forecasting week?

Answer: None, I review and provide feedback whenever I would make changes. I'll leave the actual change up to the buyer (unless I really do not agree and see risks)

7. Provide at least 5 examples for which you <u>**have not**</u> adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: Throughout the demand week I mostly review the forecast value on an aggregated/high level. Whenever I check an item individually, I look at the monthly/year-on-year averages and trend-lines. Occasionally, even if I agree with the chosen base rate and corresponding seasonality percentages, I still want a single month forecast value to be slightly adjusted. This could be the case whenever there were certain peaks and/or lows in the first years which are not to be found or less significantly re-occurring in more recent years. Using averages calculated over multiple years could result in a distorted view and is something we must be aware on.

Another aspect my team looks at is the order intake used within the weekly forecast maintenance. The order intake is the number of received orders upon that date and should be compared to the forecast value of the current month and the next month. Whenever these comparisons are out of proportion (i.e., the order intake is significantly higher or lower) compared to the number of days we still have left, an adjustment should be considered.

8. Provide at least 5 examples for which you **have** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: The weekly forecast maintenance should be done for both monthly as weekly items but is especially of value for weekly forecasted items. This since the requirement calculation for these items is re-calculated on a weekly basis. The better updated these values are, the better the supply chain can respond on abrupt demand changes. Generally speaking, a human decision maker is often more prone to over forecast their future demand instead of forecasting too less. This since human behaviour is more focussed on short-term risks like product availability, then long term risks in terms of excess stock.

From a business perspective it is very hard to quantify to what extent new introductions are going to impact the current business demand streams. It could either happen that it indeed impacts the demand streams of other (similar) models or that it does not interfere with other products at all and should be seen as completely new additional demand.

Example 1: Tender management
This is an aspect which is more related to weekly forecast maintenance, then it is to the actual forecast values. In case, there is a tender forecasted for the current month and there are still no (tender) orders to be seen within the system during the 3rd week, I advise our buyers to move this tender to the next month. Moreover, there are huge differences in between our OpCo's with respect to their reliability in terms of promised tender demand.

Example 2: Excess stock
Occasionally, I observe that there is a risk of purchasing excess stock due to a significant decreasing trend, which I act on if not reacted on by our buyers. These drastic (future) downward trends are often the result of newly introduced successors. In such situations we have be aware as demand will start shifting towards these new items.

Example 3: Phase out process
During phase-out situations, we need to share our forecast values until the moment the factory stops producing the items. In doing so, we must be sharp on preventing both excess stock (result of over-forecasting) as being sold out too soon. Preferably the latter one is more ideal as customers could be offered a similar/successor item, while excess stock must either be heavily promoted (potentially impacting other demand streams) or becomes obsolete.

Example 4: Phase in process
Our buyers should be aware that most items during a phase-in process have a gradual increase of their demand. This makes sense as those items first must be be properly launched and promoted at each OpCo before all customers are informed and getting interested. In case there is an incorrect phase in forecast (i.e., too much purchased stock at the start), I will advice my team to adjust accordingly.

Example 5: Order intake
Whenever the order intake is out of proportion compared to the forecasted value, I always advice my buyers to adjust accordingly. An unexpected tender or a not yet received promised tender could result in an order intake which is disproportional to the forecast value decided on at the start of each month.

9. Are there any other (not yet mentioned) elements/reasons which will result in a planner manually adjusting the system generated forecast?

Answer:

In theory a planner will need to adjust the forecast based on all new info he/she gets. A forecast is created in the 1st week based on available historical demand and input for tenders and any special actions/deals/promotions (all discussed during the demand week). If during the month impact is not seen or the impact is different from expectations, forecast needs to be adjusted.

10. What kind of historical demand driven information is missing and would help support a planner in its forecasting reviewing process?

Answer: 1. For the machine buyers, our current system is missing a link between different versions of the machine #1 linked to #2, #3 and so on. By linking them, the system would be able to generate a forecast based on the historical demand of all machines together. In the Excel set-up we have included this methodology and this link will also be included in our new forecasting system.

2. Next to that, our system doesn't have an opportunity to purge demand at this moment. There is not a clear methodology to exclude outliers form history which will then define a correct new baseline demand. Currently we have some manual initiatives cleaning the historical data and our new system will have some automated features of filtering out the tender demand/outliers.

3. Our system does not have a way to identify received orders as tender forecast. Our system detects all received demand as regular demand and will therefore, without manual interference, result in our shared requirement calculation being too high. In this moment in time, we have too manually shift this tender quantity to our baseline quantity to make sure that our requirement calculations are still correct.

11. What kind of other information is missing and would help support a planner in its forecasting reviewing process? (In addition to the historical demand information)

Answer: The link to promotions. RE Marketing & sales start promotions, but there is no link or no review afterwards what impact on demand really was. During my years I question if promotions are placed to increase demand on models or are required to keep the year over year trend.

12. In addition to historical demand data, is there any other kind of information/data which the system should consider when calculating/suggesting forecast figures?

Answer:

1. Perhaps standardizing the relation between the type of promotion and quantified gain. So, if we know a certain type of promotion is going to happen in the (near) future then we can provide the system with input on the impact of this promotion with respect to the demand stream. However, before we can include such information, we need to define all kinds of promotions including all impacts of these promotions. This is the one thing which I could think of when I'm thinking about stuff that could be automated/given as input to the system.

2. Another aspect could be to include economic trends. For instance, the economic situation in countries we are selling to probably impacts our demand streams as well. However, such data is very hard to include as it is complex to determine and define the impact of these trends. So, to conclude, yes there are features which could potentially improve system generated forecast accuracy, but those are often very hard to quantify.

A.11 Judgemental forecast interviews for Supplies

Interview with 1st Supply demand planner

1. For how many years have you been working at Ricoh?

Answer: 6 years

2. For how many years have you been active in planning related jobs?

Answer: 6 years

3. For which product type(s) are you responsible during the forecasting process?

Answer: Supplies (RPL, RPLS, Asian + double-sourced)

4. How many unique SKUs are within your portfolio?

Answer: 391 (Active + Pre-launch status)

5. How many of those do you need to manually review on a monthly basis?

Answer: In theory based upon quadrant, but I am reviewing all. However, the distribution of attention is based on the quadrant status. In other words, I put more time into reviewing AA items compared to DD items.

6. Approximately how many items have you adjusted during the current demand forecasting week?

Answer: 50

7. Provide at least 5 examples for which you <u>**have not**</u> adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: Seasonality has a big impact on the demand patterns of Supplies. Within Ricoh we recognize three re-occurring trend months. These months are March (Peak month due to end fiscal year in which OpCo's are pushing orders to achieve budget goals), August (Low month due to holiday break) and September (Peak month due to a rebound from the vacation months). However, some product lines are more heavily influenced by seasonality than others. Production printers are for instance constantly in use and their supplies are also relatively expensive making them less sensitive for these seasonality months.

Overall trend lines are in addition to seasonality, also an important factor to consider. While reviewing each item one must conclude to what extend the trendline is increasing, decreasing, or stabilizing. Based on the yearly averages one can for instance decide on this matter.

Another factor to consider during our reviewing process is the MIF. Unfortunately, within finished goods we do not have a straightforward figure which states the MIF. However, as long as the corresponding machine(s) are active one can assume that the MIF trend will also increase (resulting in an increased supply demand trend). Whenever these machines get a Discon item status, the MIF will first stabilize before eventually decreasing and this behaviour will be similar for the corresponding supply demand. The status for all corresponding machines can be found in Optimiza and is definitely something I check whenever I do not know this by heart.

Example 1: EDP 817104 (JP12 Ink Priport from RPLS)
Current statistical forecast shows seasonality, which is also in line with historical demand data.

Example 2: EDP 407324 (PCU SP 4500 (bp))
Statistical forecast is once again in line with historical demand and the system correctly build in the re-occurring seasonality patterns.

Example 3: EDP 407716 (Black toner SPC252E (Double-sourced/BP)
Forecast trend reflects decreasing trend + seasonality. Overall forecast trendline in balance with historical demand.

Example 4: EDP 842080 (Yellow MP C305 (RPL))
Forecast algorithm shows seasonality + matching demand trend.

Example 5: EDP 828601 (Black C5300 (RPL))
Forecast algorithm shows increasing trend + seasonality.

8. Provide at least 5 examples for which you **have** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: Adjustments within the system generated forecast could be both just an adjustment for only one month up to changing every month by changing the statistical forecasting method. Furthermore, I also check the quantity differences between the three seasonality months and try to conclude/adjust future seasonality months based on this information.

Example 1: EDP 817225 (Black ink HQ40 from RPLS)
There is a peaky historical demand, meaning that there is not a clear demand pattern to be identified. We do not see the same demand peaks and lows re-occurring on a yearly basis. It is also not linked to only one dealer. So that is when and why we proposed to build in a stable future demand trend (due to its unpredictability). One extra adjustment was made as I increased March to prevent (potential stock-outs) due to seasonality. Furthermore, potential stock-outs can be anticipated fairly quickly as this item is produced in Europe and lead times are relatively short compared to items produced in factories located in Asia. That's why for such items I prefer to anticipate instead of trying to guess when and where another peak is going to occur (to prevent excess stock).

Example 2: EDP 407318 (SP4500HE Black (Double-sourced/BP))
This item has an increased forecast trend in which the statistical forecast proposal is

considered way too low when compared to the historical demand data.

Example 3: EDP 407634 (SPC 310 HE Black (Double-sourced/BP))
A relatively 'old' item, so we know it has a decreasing MIF. Compared to yearly averages from two years ago there was a decrease of 500 units and only 200 units last year. Thus, one can conclude that demand is still decreasing, but with a more stabilizing behavior. That is why I increased the 'Box Jenkins' statistical forecast with + 200x p/m as overall forecast line including seasonality patterns were in line.

Example 4: EDP 828314 (pro C9100 (RPL))
Similar to the last example as the yearly averages show a decreasing trend which is stabilizing in more recent years. Statistical forecast algorithms put too much focus on the decreasing trend and therefore proposes a too low forecast. It does show seasonality correctly, but overall forecast trend too low.

Example 5: EDP 841817 (Black 14p C3503 (RPL))
Statistical forecast algorithm proposes too low + too quick decreasing. MIF is declining, but not in this rate. Similar as the 3rd and 4th example.

9. Are there any other (not yet mentioned) elements/reasons which will result in a planner manually adjusting the system generated forecast?

Answer:

1. Phase-in/out situations must be handled separately as there is not enough historical demand available for the system to calculate a realistic and accurate forecast. These items are checked with the supply calculation sheet which considers expected corresponding machine sales, sales channel split, toner sales split (i.e., yellow is the least used colour), toner yield, whether a start toner is included in the machine sale and average print volume. Based on all information above we come up with a best guess which will be the forecast for this item specifically.

2. Tender demand which is added separately to the system generated forecast numbers.

10. What kind of historical demand driven information is missing and would help support a planner in its forecasting reviewing process?

Answer: No additional types of information were mentioned.

11. What kind of other information is missing and would help support a planner in its forecasting reviewing process? (In addition to the historical demand information)

Answer: Exact number of MIFs per product-family. This is a figure which could be linked to average usage/duration for each supply, which will subsequently help in predicting these supplies. This would be especially helpful for our important AA items.

12. In addition to historical demand data, is there any other kind of information/data which the system should consider when calculating/suggesting forecast figures?

Answer: One example which I could think of is for one of our Production printers which had an unexpected peak in demand throughout several months. This peak was the result of a fire which happened in one of the factories of our competitor. Therefore, our customers who had both parties as their supplier placed all their orders within Ricoh.

Interview with 2nd Supply demand planner

1. For how many years have you been working at Ricoh?

Answer: 9 months

2. For how many years have you been active in planning related jobs?

Answer: 4 years

3. For which product type(s) are you responsible during the forecasting process?

Answer: Supplies

4. How many unique SKUs are within your portfolio?

Answer: 506

5. How many of those do you need to manually review on a monthly basis?

Answer: All. Some are quickly scanned through while others require a bit more attention. More attention is required whenever the forecast line seems off compared to the expected trend from a visual perspective. Moreover, the kind of portfolio and the ABC classification also indicate which items require more attention than others. For instance, industrial and commercial printing require more attention as their business depends on having a continuous availability/feed of supplies.

6. Approximately how many items have you adjusted during the current demand forecasting week?

Answer: At least 100

7. Provide at least 5 examples for which you **have not** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: Usually when I adjust a certain EDP code I adjust the entire series. A series comprises all the different colours of the same toner type for instance.

For the items which are quickly checked, I mostly focus on the graph and check whether the trend and seasonality are correctly included in the suggested forecast. For the more crucial items or whenever a more in detail check is required, I will check the background numbers which reflect the monthly sales quantities + yearly averages. Based on those values I could for instance decide that another algorithm fits the data much better or whether I need to make manual adjustments to the forecasted monthly values.

During the demand meeting we receive information on changing item statuses of the machines. These changes are going to impact the corresponding supplies demand. The item status is also something we can check within Optimiza in the BOM tab.

Example 1: EDP 828551
Looking at the recent historical months/years there is a very stable demand, which is not increasing nor decreasing. The system suggested a forecast equaling the average of the most recent months (which I agreed on) and also correctly anticipates on the seasonality peaks (September and March). Moreover, the corresponding machine has a discontinued item status which explains why the demand is stable and will for sure start decreasing in the (near) future.

Example 2: EDP 408453
It's difficult to forecast this series, as it is a relatively new item. Due to a lack of historical data, the system was not yet able to calculate reasonable predictions. That is why previous forecasts were based on another file which states expected demand according to the corresponding machine forecast. However, these expectations were way too optimistic when compared to the sales volume of the most recent months. As from last month onwards we are starting to get enough data, I decided to let the statistical forecast calculate the forecast and made small adjustments to comply with general seasonality patterns.

Example 3: EDP 405532
Is an item which has low sales volume and is quite stable and therefore did not require any adjustments.

Example 4: EDP 408314
Within the graphical representation of the forecast one could clearly see that it follows the trend correctly and I therefore did not change anything.

Example 5: EDP 842164
Follows (decreasing) trend correctly which you could clearly see within the graph and therefore I did not change anything.

8. Provide at least 5 examples for which you **<u>have</u>** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: As our submission date is not on the first day of the month, we use the extra days to look at the order intake value for certain items. Whenever the order intake compared to the forecast value for that month is out of proportion, I need to adjust the forecast accordingly. However, from this order intake value we must look closely at which OpCo's already have placed their orders, as some OpCo's already place all their orders at the beginning of the month. Therefore, it is not always true that if on day 1 we already have 25% of forecasted demand that we then should automatically increase our forecast. The order intake is a figure which we can check within Optimza.

Example 1: EDP 257059
I changed the forecast because the last 6 months the demand has been decreasing which was not suggested by the system.

Example 2: EDP 405761

For this item I looked at the most recent months which clearly indicated that the demand is decreasing and therefore it was better to follow the average of the recent months. In previous months I forecasted these items based on the demand stream of the same month of last year, but it was considered more accurate to put more emphasis on sales volumes from recent months.

Example 3: EDP 413013
This is a highly volatile item, which is difficult to predict. For such items, I often take the average of the most recent months and add seasonality peaks to it.

9. Are there any other (not yet mentioned) elements/reasons which will result in a planner manually adjusting the system generated forecast?

Answer:

1. Sometimes none of the forecast options the software offers is aligned with the historical demand. If that is the case, I have too manually adjust the forecast and freeze it. Would be helpful if we get a warning, that we froze such items.

2. September is considered a high selling month, this year I noticed that for some of my items the demand was low, however in October it was high. We as planners must decide whether this kind of behaviour will take place whenever a peak month does not result in high volumes.

10. What kind of historical demand driven information is missing and would help support a planner in its forecasting reviewing process?

Answer:

1. Within the historical data buckets, we are not able to distinguish tender demand from the average demand, making it hard to correctly interpret trends.

2. With respect to the order intake, it would be nice to have information on how the orders are distributed over the different OpCo's. This information in combination with domain knowledge will help me in making more calculated decisions on whether a forecast adjustment is necessary. Currently I have to switch between Optimza and Excel too find those figures.

11. What kind of other information is missing and would help support a planner in its forecasting reviewing process? (In addition to the historical demand information)

Answer: Could not think of any other sorts of information.

12. In addition to historical demand data, is there any other kind of information/data which the system should consider when calculating/suggesting forecast figures?

Answer: Perhaps sales projections/expectations from a marketing/sales side of business could be useful for the system to consider. However, before adding them their predictions should become more reliable. It often happens that an OpCo places a significant order without this ever being communicated with us.

Interview with Supply supervisor

1. For how many years have you been working at Ricoh?

Answer: 2.5 years

2. For how many years have you been active in planning related jobs?

Answer: 8 years

3. For which product type(s) are you responsible during the forecasting process?

Answer: Annuity Consumables which includes several supplies / toners / etc.

4. How many unique SKUs are within your portfolio?

Answer: Approximately 1,000 SKUs (i.e., the entire supply portfolio)

5. How many of those do you need to manually review on a monthly basis?

Answer: 1x per month I will check all items in full and I quickly scan through the forecast maintenance view on a weekly basis. This view provides information on the order intake upon that moment and could indicate any required forecast adjustments within that month.

6. Approximately how many items have you adjusted during the current demand forecasting week?

Answer: +/- 5-10

7. Provide at least 5 examples for which you **have not** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: The main difference between machines and supplies forecast is that whenever a customer decides to buy a machine a certain lead-time is acceptable, while when supplies are ordered they should be delivered the next day as supplies should always be available for our customers.

One of the key aspects which significantly impacts future demand patterns is the so-called MIF (active machines in field). This since the trend line of the MIF correlates with the trend line of the corresponding supply demand. For example, if the MIF is decreasing than we can assume that the supply demand will also stabilize/decrease over time.

Portfolio's which already have a significant MIF are less impacted with an increase/decrease in the MIF in comparison to items which still have a relatively low MIF. Therefore, we must be extra careful in predicting the supply demand within such (low MIF) portfolios. That is why such portfolio are often discussed during the demand week with representatives of our marketing team. These discussions are about whether projected installations are still likely to happen and if there any other installations to occur.

Furthermore, we distinguish two kinds of supplies: toners and ink, from which ink has an expiration date. For such items we must find a balance between being able to support the MIF while at the same time we should try to avoid exceeding the expiration date for our inks on stock as much as possible.

Example 1: EDP 842339
Chose to follow system suggested statistical forecast as it includes a proper (decreasing) trend and correctly forecasts the seasonality trend. I have checked whether the indicated highs and lows are matching our general seasonality patterns (which it does) and looked at the historical demand of the past 6 months to tell me something about future trends and overall expected sales volume. In the current set-up we do not see the number of MIFs for this item, I can only guess/conclude that the MIF is decreasing as our supply demand is also decreasing combined with the fact that I know it is a relatively 'old' item through domain knowledge.

Example 2: EDP 842349
I again chose to follow the system suggested statistical forecast as it includes a proper trend and correctly forecasts the general seasonality trends. Whenever I noticed a decreasing trend, I always check the item status of the corresponding machine. Whether the MIF of a machine is increasing, stabilizing, or decreasing is often based on where in the life cycle that same machine is.

Example 3: EDP 418478
This is a supply for a machine which is still active but is classified as a relatively 'old' machine and therefore has a stabilizing MIF. The system suggests a similar kind of stabilizing behaviour for the corresponding supply as it shows a flattening demand trend which is set almost equal to last year sales. In other words, it does not make the mistake to focus too much on the increasing year-on-year sales resulting in an increasing trend. In addition to the average demand, this item also includes some additional tender demand. Regarding tenders I always ask the responsible demand planner whether they really think these tenders should be seen as additional demand on top of the average demand.

Example 4: EDP 821279
This supply is currently within the saturation phase of the life cycle (corresponding machines are not active anymore). Choosing simple move average is more than sufficient as it has predictable (slowly decreasing) demand. The phase-out strategy/portfolio is determined by our product marketing department in which we as a team occasionally share our opinion.

Example 5: EDP 418240
This is an item which requires domain knowledge from a buyer as the basic tournament selection won't fit. The MIF of this item was increasing until it stagnated due to hardware availability issues which also resulted in lower supply sales. As over time these availability issues were improving, the sales and thus the MIF also increased again. Based on domain knowledge the buyer knew that the tournament selected algorithm won't be a great match and had to manually select a different algorithm (exponential smoothing in this case).

This kind of information was shared during the demand week and is one of the reasons why it is important we are participating in the demand week regarding the corresponding machines.

8. Provide at least 5 examples for which you **have** adjusted the system generated forecast based on the historical demand data? Please mention which exact elements/reasons have led to this decision.

Answer: One of my tasks and responsibilities as a demand planning supervisor is to monitor inventory levels and provisions. For instance, as we had a period in which items had to deal with shortages, planners became more prone to over forecast their items just to be on the safe side. This is one of the subjects I'm currently trying to put more focus on as sometimes this leads to unwanted risks. Furthermore, adjustments are often made on entire series. A series is for instance a series of colours within the same toner type. If an adjustment is needed for the black toner, I always advice to do the same adjustments for all colours. Unfortunately, each item within the series must be manually adjusted separately.

Example 1: EDP 719317 + series (VC4000 ink.)
Current forecast based on (very low) MIF. There are new Installations planned which will result in a higher corresponding supply demand. The information regarding these installations is shared by product marketing who tells us that in for instance January these installations will take place, resulting in exponential increase in supply demand from February onwards.

Example 2: EDP 408517 + series
This is a recently new launched supply, for which we must manually forecast the first 6 months ourselves. After 6 months the system is able to identify trends and thus start proposing reasonable forecast values.

Example 3: EDP 937481 + series
We stopped supplying our highest selling OpCo (Russia) which had a huge impact on our expected future sales. In such cases, we must give huge trustworthy to the knowledge of the buyer to adjust accordingly.

Example 4: Metis MF supply + series
The currently active Metis MF model is planned to be phased out and therefore we must be aware that there is not going to be an increase in the MIF anymore. As a result, the demand forecast trend needs to be a stabilizing trend, rather than an increasing one. The responsible buyer needs to anticipate and correct the forecast based on this knowledge manually. Eventually, the new Metis generation will slowly take over the demand of this predecessor.

Example 5: EDP 408060 / 408061 / 408062
These are all toners, which are the same and only differentiate in their yield. Product marketing wants to offer all three yields to provide the customer with options. We currently have discussions with product marketing to eliminate one or even two versions. If this change will take place than we must be aware that we need to do manual changes within the forecast as demand of these rationalized items will most likely shift to the

remaining one.

9. Are there any other (not yet mentioned) elements/reasons which will result in a planner manually adjusting the system generated forecast?

Answer:

1. Phase-in/out situations as during these events more manual interference is required.
2. Seasonality as supply demand is often impacted during certain months
3. Additional tender demand
4. Availability of hardware in combination with the impact on the MIF development.
5. Promotions which will (temporality) push certain supply demand streams.
6. Rationalization of items within the current portfolio (See question 8, example 5)

10. What kind of historical demand driven information is missing and would help support a planner in its forecasting reviewing process?

Answer: Tender quantity versus average demand on an OpCo level and a total level. Share information on how big the assumed tender demand is compared to the total demand of all OpCo's together and within the demand stream of that OpCo. If it is indeed a significant volume on both aspects, then it should be considered a tender and added as additional demand. As an example, a tender of 100 units on an average demand of 10.000 should not be considered additional tender demand. Even if it is a significant volume for that specific OpCo.

11. What kind of other information is missing and would help support a planner in its forecasting reviewing process? (In addition to the historical demand information)

Answer:

1. A link between machine Lifecycle status / total MIF vs supplies forecast. If we for instance get insight into the corresponding historical MIF value for each supply, we can get a better understanding on the relationship between our total MIF and our supply demand. Having this understanding will help us in better guessing the impact on our supply demand due to future changes within the MIF (Influenced by tender demand and/or the lifecycle status of our machines).

2. The average ink consumption/print volume of the MIFs for different OpCo/region would be something which would be helpful. Perhaps an UK machine has a lower average print volume than machines located somewhere else. Especially interesting for portfolio's which have a relatively low MIF. For instance, 5 new machines on a total MIF of 5 machines does not always result in a double amount of supply demand. Certain customers have a higher average print volume than others. Although this information is valuable, I can imagine it would be hard to quantify such information as this is mostly a gut feeling from product marketing.

3. Learn/identify and standardize demand patterns during phase-in situations. Perhaps we are able to identify a certain standardized demand pattern (within a product-family/line) which can be applied throughout future phase-ins.

12. In addition to historical demand data, is there any other kind of information/data which the system should consider when calculating/suggesting forecast figures?

Answer: The relationship between our machines and supplies would be very interesting. If it is possible to accurately quantify the (future) MIF value for all supplies and get a good understanding of the relation between changes in the MIF and the supply demand, then we should provide our system with this kind of information. This could again be combined with average print volume for each customer/region/OpCo. Using this kind of information will make our forecasting much more data driven and less based on gut feeling.