Aalto University
School of Science
Master's Programme in Life Science Technologies

Esa Turkulainen

Forecasting Blood Demand:

Observations and an implementation

Master's Thesis
Espoo, July 31st, 2020

Supervisor:	Professor Jari Saramäki
Advisor:	Mikko Arvas, PhD

Aalto University
School of Science
Master's Programme in Life Science Technologies

ABSTRACT OF MASTER'S THESIS

Author:	Esa Turkulainen		
Title: Forecasting Blood Demand: Observations and an implementation			
Date:	July 31st, 2020	**Pages:**	54
Major:	Complex Systems	**Code:**	SCI3060
Supervisor:	Professor Jari Saramäki		
Advisor:	Mikko Arvas, PhD		

Reliable blood supply chains are critically important for modern medicine. However, blood inventories are perishable, which frames the issue as an inventory management problem with separable supply and demand components. Inventory management can be improved via multiple avenues, but reliable demand estimation is among the most powerful ones, as it helps parties involved in blood collection to scale the collection based on projected demand, thus reducing the amount of outdating units and alleviating shortages.

The Finnish Red Cross Blood Service (FRCBS) is responsible for maintaining the blood supply chain in Finland. Currently, operational level (donor mobilization) estimates of demand are created weekly by using in-house expertise and planning level (budgeting) estimates by machine-generated statistical forecasts.

This thesis aimed to examine the historical performance of the statistical forecasts used for budgeting and to investigate if they could be improved and expanded to monthly and weekly forecasts for different types of red blood cells. The efforts consisted of reviewing the published literature on short-term and long-term blood demand forecasting, examining the available data, establishing appropriate metrics for evaluation, and trying out better models.

We find that that the mean absolute percentage error of the current forecasting methods can be improved by 22.2% with an additional data preprocessing step and by 50.1% by changing to a better model. The temporal resolution of forecasting was improved by changing the data source. Also, we discovered that the nature of the blood demand signal changes significantly around 2017, underlining the need to develop forecasting systems with the capability to adapt to changes.

Our final implementation is built into an R Markdown file to output an easily accessible HTML for reporting. Further exploration is warranted, especially if the aim is to use forecasting operationally someday.

Keywords:	blood supply chain, blood demand, operative modeling, time series analysis, autoregressive, adaptive
Language:	English

Aalto-yliopisto
Perustieteiden korkeakoulu
Master's Programme in Life Science Technologies

DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Esa Turkulainen		
Työn nimi: Verentarpeen ennustaminen: Havaintoja ja toteutus			
Päiväys:	31. heinäkuuta 2020	**Sivumäärä:**	54
Pääaine:	Complex Systems	**Koodi:**	SCI3060
Valvoja:	Professori Jari Saramäki		
Ohjaaja:	Tohtori Mikko Arvas		

Verihuoltoketjun luotettavuus on kriittisen tärkeä osa modernia lääketiedettä. Veri vanhenee muutamassa päivässä, mikä asettaa huoltoketjuongelman varastonhallinnan viitekehykseen erillisillä kysynnän ja tarjonnan osa-alueilla. Varastonhallintaa voi kehittää useilla eri menetelmillä, mutta kysynnän ennustaminen on menetelmistä tehokkaimpien joukossa, sillä se mahdollistaa veren keräyksen kysynnän perusteella vähentäen erääntyvien veripussien määrää ja riittämättömien varastojen riskiä.

Suomen Punaisen Ristin ylläpitämä Veripalvelu vastaa verihuoltoketjun ylläpidosta Suomessa. Nykyisellään operationaalisen tason (luovuttajien kutsuminen) ennusteet tehdään viikoittaisissa kokouksissa asiantuntijoiden kokemusta hyödyntäen. Pitemmän aikavälin suunnitelmalliset (budjetointi) ennusteet tehdään laskennallisesti aikasarja-analyysillä.

Tämän opinnäytetyön tavoitteena oli arvioida käytössä olevien laskennallisten ennusteiden historiallista tarkkuutta ja selvittää, voiko tarkkuutta parantaa tai ovatko ennusteet laajennettavissa viikottaisiin ennusteisiin ja useampiin verityyppeihin. Tavoitetta edistettiin kirjallisuuskatsauksella verentarpeen lyhyen ja pitkän aikavälin ennustamiseen, saatavilla olevan datan tarkastelulla, sopivien tarkkuusmittareiden selvittämisellä ja muiden mallien testaamisella.

Työn aikana selvisi, että käytössä olevia ennusteita voidaan parantaa 22,2 prosentilla lisäämällä prosessiin uusi datan esikäsittelyvaihe ja 50,1 prosentilla vaihtamalla käytettävää mallia parempaan. Ennusteen aikatarkkuutta saatiin parannettua vaihtamalla datan lähdettä. Opinnäytetetyön päälöydös oli kuitenkin verentarpeen signaalin luonteen merkittävä muutos vuoden 2017 paikkeilla, mikä alleviivaa muutoksiin sopeutuvien ennustejärjestelmien tarpeellisuutta.

Lopullinen järjestelmä rakennettiin R Markdown -skriptin sisälle helppolukuista HTML-raportointia varten. Tarpeen ennustamisen jatkotutkimusta tarvitaan, varsinkin jos tavoitteena on ennusteiden käyttö operationaalisesti.

Asiasanat:	verihuoltoketju, verentarve, operatiivinen mallinnus, aikasarja-analyysi, autoregressiivinen, adaptiivinen
Kieli:	Englanti

Acknowledgements

First and foremost I owe a great deal to my supervisor Prof. Jari Saramäki for support and guidance beyond the call of duty, to Jarkko Ihalainen for trusting my passion in the subject and allowing me to work at the FRCBS under the guidance and expertise of Mikko Arvas, PhD, and Muriel Lobier, PhD, who have helped in all things related to project management and R programming.

I also wish to send my deepest gratitude to all of my friends, who are often in the splash damage range from my nervous breakdowns. Thank you also for all the feedback I received during the writing of this thesis.

Helsinki, July 31st, 2020

Esa Turkulainen

Abbreviations and Acronyms

ANN	Artificial Neural Network
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average model
ARIMAX	ARIMA with exogenous regressors
ARMA	Autoregressive Moving Average
BJ	Box-Jenkins method/protocol
CF	Combination Forecast
cMAPE	Operation critical MAPE
CSV	Comma Separated Values file
ETS	Exponential Smoothing Models
FRCBS	The Finnish Red Cross Blood Service
FFP	Fresh Frozen Plasma
HTML	Hypertext Markup Language
LOESS	Locally Estimated Scatterplot Smoothing
MA	Moving average
MAPE	Mean Absolute Percentage Error
MPE	Mean Percentage Error
NNAR	Autoregressive Neural Network model
PDF	Portable Document Format
PLT	Platelets
RBC	Red Blood Cells
RMSE	Root-Mean-Square Error
RSS	Residual Sum of Squares
SARIMA	Seasonal ARIMA
SMA	Simple Moving Average
STL	Seasonal and Trend decomposition model
TBATS	Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components model
VARMA	Vector ARMA
WHO	World Health Organization

Contents

Chapter 1

Introduction

1.1 Background

Much of the world's health care relies on the availability of fresh blood and its derivatives. In Finland, some 215,000 transfusion events occurred between 2011-2016 in the Hospital District of Helsinki and Uusimaa alone, even though the trend was decreasing [Laurén et al., 2019]. Roughly half of all the transfused blood is required in surgeries, and the rest is used in the treatment of other medical conditions, such as gastrointestinal and cardiovascular diseases [Palo, 2013]. Thus, shortages in fresh blood translate directly to potential loss of life, and the need for a reliable supply chain is apparent. As blood is a perishable commodity, the supply chain reliability is not fully guaranteed via storage solutions. Indeed, with whole blood units expiring in 21/35 depending on the anticoagulant used and red cells in 42 days [American Red Cross Blood Services, 2020], the ethical and cost-related problems arising from wasting voluntary, non-remunerated donations demand the implementation of supply-demand analysis. Voluntary and non-remunerated blood donation is the World Health Organization (WHO) recommendation and the current policy in 28% of countries [World Health Organization and International Federation of Red Cross and Red Crescent Societies, 2009, 2010].

The Finnish Red Cross Blood Service (FRCBS) is solely responsible for the blood supply chain in Finland. Each working day hundreds of donors are serviced; their donations are exhaustively screened to guarantee the safety of the patient and then produced into several types of products, ranging from red cell products for surgical patients to platelet products for cancer patients, and then stored and delivered to hospitals per demand. Figure 1.1 presents a rough diagram of the blood supply chain as provided by the

FRCBS. In 2019, a total of 198,339 units of blood were drawn from donors, of which 190,437 RBC product units and 31,621 platelet product units were distributed to hospitals [Finnish Red Cross Blood Service, 2020]. This is equivalent to the efficiency of roughly 97 percent, with less than 1 percent expiring in storage. An operation efficiency this high is not given and requires supply management strategies and cooperation between receiving and supplying parties. During the early 2000s, the FRCBS aimed to identify and fix some of the most significant shortcomings in the blood supply chain, including storage management issues, forecasting related problems, and insufficient cooperation with hospitals. While most of the storage management issues could be alleviated with improved logistics and increased cooperation between the blood service and the hospitals [Rautonen, 2007, Sihvola, 2016], forecasting demand remains an issue.

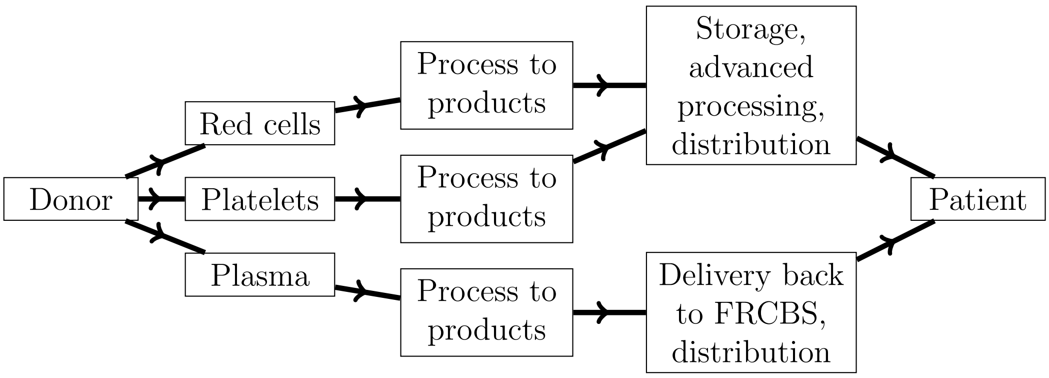


Figure 1.1: The blood supply chain as provided by the FRCBS. The chain begins with an eligible donor donating 0.5 liters of fresh blood. This blood is divided into red cells, platelets, and plasma for further processing into different blood products. RBC and platelet products require the removal of leukocytes, and plasma is frozen and shipped to pharmaceutical laboratories to be processed into medicine. After processing (and screening), the products are either stored, processed further, or distributed to hospitals, in which the products are administered to patients.

Demand forecasts are typically divided into three different levels in the context of supply chain management: operational level (short-term), tactical level (medium-term), and strategic level (long-term). On the operational level, forecasting is usually done on a weekly basis to match supply with current demand. Tactical forecasting is used to help to make budgeting decisions and usually involves the use of monthly data, while the long-term strategic demand estimates are used to facilitate responses to changes in demand trends [Filho et al., 2013]. The FRCBS is currently using simple models to forecast the sales of blood and its derivatives on a tactical level. The supply chain critical forecasts of demand for donor mobilization (operational level) are still made entirely by FRCBS personnel, mainly based on experience. While historically functional, this kind of arrangement can be vulnerable to loss of personnel and inaccuracies introduced by unexpected shifts in trend, and finding working complementary forecasting models could credibly decrease these kinds of risks. With complete information about scheduled surgeries, the number of patients requiring blood products regularly, and the incidence of trauma patients, one can imagine being able to optimize the number of donors mobilized any given day fully. Without this kind of precognition or integrated systems between the FRCBS and the hospitals, the best option currently is to try to find trends, seasonality, and cycles in the blood demand time series.

1.2 Scope

The scope of this thesis is to review the current literature on blood demand forecasting, determine metrics for evaluating operational and tactical level forecasts, review the existing forecasting methods used at FRCBS, explore the possible sources for demand data, study the nature of the demand time series and attempt to improve the forecasting models and practices based on the observations. Additionally, a user interface is developed for accessing the forecasts.

1.3 Outline

This thesis begins with the literature review in Chapter 2. Research on both operational and tactical level forecasts are considered. Chapter 3 contains the analysis of current forecasts used at the FRCBS and explanation of the metrics used throughout this thesis. We then choose our ultimate source of demand data and explore the behavior of the demand signal in Chapter 4, including also some tangential findings for completeness' sake. Finally, Chapter 5 presents the complete final implementation of the developed forecasting pipeline, along with performance evaluations. Results and findings are discussed in Chapter 6, which concludes with problems with the current implementation and possible future directions for research.

Chapter 2

Literature review

The literature search was performed with Google Scholar and PubMed using the search terms "blood", "demand", "prediction", "forecast", "model", and "short-term" in various combinations and selecting papers from the last two decades. Some older but relevant papers were identified in the citations of the selected papers.

Most of the published literature deals with blood supply chain management, mostly out of the scope of this thesis. Nevertheless, all papers mentioning demand estimation by some means were included in the selection. We will review literature both on the operational and tactical levels.

2.1 Modeling short-term demand

The most often cited early blood demand forecast attempt employed an exponential smoothing model (ETS) with a 10-day ("arbitrarily selected") cyclical weight component [Frankfurter et al., 1974]. Exponential smoothing methods approximate time series using a linear combination of past observations with exponentially decaying weights (more detailed explanation in Section 3.1). The model is used as a part of a "computerized" inventory management system, and the authors do not discuss their reasons for opting for an ETS model. They report that they attained savings using the system, but do not discuss the accuracy of the demand forecasts. A bit later, Gardner [1979] found that ETS models did not predict the monthly demand for blood testing in a clinical laboratory very well, although they outperformed some more complex models, such as autoregressive integrated moving average models (ARIMA) selected using the Box-Jenkins (BJ) method. Instead, simpler multiple regression models with a dummy seasonal variable consistently gave mean absolute percentage errors (MAPE) under 5.5% (9.4% for

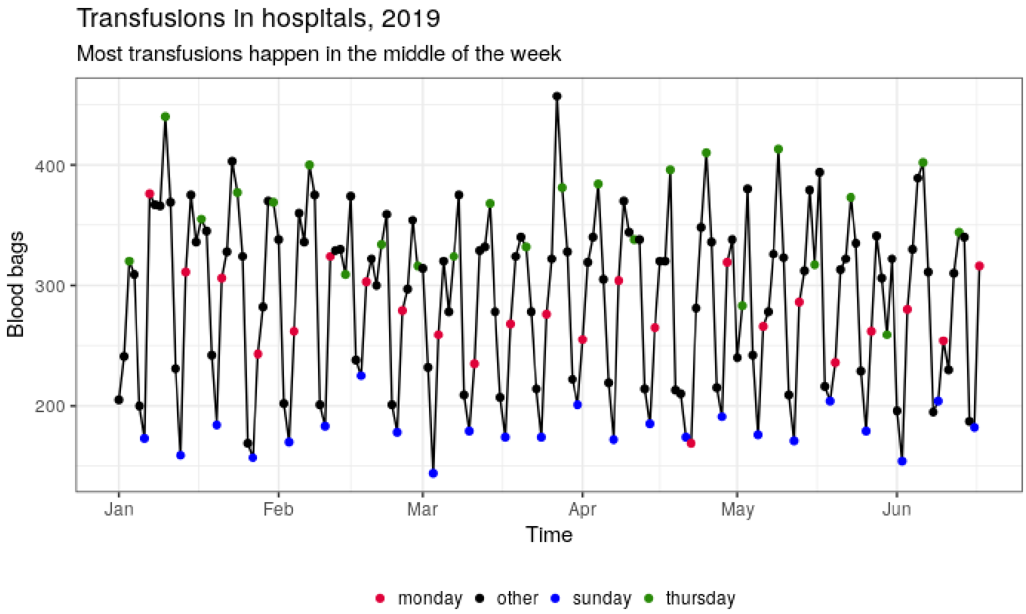


Figure 2.1: Data on transfusions performed in hospitals reveal a pattern: most transfusions are performed in the middle of the week and almost solely on work days.

the best ETS). The author also discusses how the use of complex BJ models is discouraged for being too complicated to interpret and maintain without hiring an analyst. For clarity, ARIMA models are generalizations of autoregressive moving average models (ARMA), which regress the target series using its lagged values (AR polynomial) and take the moving average of the error terms (MA polynomial) to generate a fit. ARMA models require stationarity of the target series, and ARIMA models overcome this limitation by ensuring time series stationarity using differencing. The BJ method is an often automated procedure for selecting the best ARIMA models. Multiple regression, on the other hand, is linear regression with multiple regressors. ARIMA modeling is covered in more detail in Section 5.4.3.

Pereira [2004] attempts to forecast short-term blood demand (12-month and 24-month forecasts) with ARIMA, ETS, and artificial neural network (ANN) models. The author cites Chatfield [2000] when explaining ETS models should model low order moving average (MA) processes well, but no other reasoning is given for the particular choice of models. The neural network input layer consisted of past observations of the series at chosen lags. Parameter selection for all of the models was made automatically using different kinds of software. An ARIMA$(0, 1, 1)(0, 1, 1)_{12}$ model resulted in the best accuracy for 1-year forecasts, and an ETS with a linear trend and multiplicative seasonality performed the best with 2-year forecasts. Both outperformed the neural network in both time windows and all metrics. The performance was measured using the root of residual sum of squares (the root of RSS) and monthly coverage and outdate percentages. Filho et al. [2013] state that blood centers in Brazil are mainly using simple τ-order moving-average models (SMA) to estimate future demand. SMA models are used when the series is expected to follow a slowly changing (noisy) mean, and its forecasts are averaged observations of the series inside a moving window of pre-determined length (Section 5.4.1). They suggest using seasonal ARIMA (SARIMA) models selected using the Box-Jenkins protocol. The authors demonstrate the validity of the method but do not attempt to estimate its performance.

More recently, Fortsch and Khapalova [2016] compared the performances of naïve, SMA, ETS, decomposition, and BJ-ARMA models in predicting blood demand by type and vector autoregressive moving average (VARMA) models in categorical prediction of demand (all types simultaneously). Models were selected by studying the characteristics of the demand series. A naïve model is often used when the modeled series is considered to behave similarly to a random walk, and its forecasts are simply the last observation in the series, as explained in Section 3.4.1. Time series decomposition models attempt to extract the cyclical, seasonal, and trend components of the series and using them in forecasting. The authors do not detail how their decomposition was achieved, but refer to Hanke et al. [2001] for implementation using Excel. The decomposition approach taken in this thesis is covered in Section 3.2. VARMA models are multivariate generalizations of univariate ARMA models. The authors find that ETS models predict the demand for positive blood types better than STL models and that STL models predict the demand for negative types better than ETS models. However, both BJ-ARMA and VARMA outperform them measured by root means squared error (RMSE) and MAPE.

Tanyavutti and Tanlamai [2018] explore forecasting blood demand in Thailand using ETS models and ARIMA models with exogenous covariates (ARIMAX). Thailand can be considered a special case in the context of this review because supply rarely ever meets demand, and that high regional prevalence of dengue fever is a significant driver for RBC demand. The authors try using both the regional platelet demand series and regional dengue fever morbidity as exogenous covariates in their ARIMA models. They find that the ETS models outperform the ARIMAX models on the national level, but the best performance varies between regions. The authors suggest that this is an effect of the high regional variance in dengue fever morbidity.

Most recently, Fanoodi et al. [2019] compare the performance of ARIMA and ANN models in predicting the demand for different blood types. The ANN used lagged observations from the time series as inputs, and the network structure was chosen using a procedure established by Kaastra and Boyd [1996]. Both perform almost equally well against the baseline of averaging, improving RMSE 27% to 66% depending on the model and blood type.

Demand has also been estimated by modeling it as a Poisson process [Bar-Lev et al., 2017], or by drawing from a normal distribution when the average and standard deviation are known and from a negative binomial distribution when they are unknown [Dillon et al., 2017]. Dillon et al. comment on the validity of the chosen distributions by referring to recommendations by Nahmias [2011]. In forecasting terms, these methods are roughly equivalent to averaging with a noise parameter, as we are assuming a constant demand with some variance.

Finally, Wilding et al. [2006] examine the overall forecastability of the blood demand signal in the United Kingdom under the chaos theory domain. The authors find that the demand signal fulfills all the requirements for a truly chaotic system, and only short-term forecasts can be convincingly made. Chaos theory falls outside the scope of this thesis, and no evaluation of the overall forecastability is attempted here.

It is important to note that the demand for blood and its derivatives is likely very different from country to country and possibly from decade to decade, making the majority of the published literature incomparable and explaining some of the conflicts in results in these selected studies.

2.2 Estimating long-term behaviour

The academic discussion concerning the long-term behavior of blood demand has long revolved around the rapid growth of the world's older population [He et al., 2016]. The connection between age and increased demand for blood transfusions has been established in many countries [Wells et al., 2002, Anderson et al., 2007, Beguin et al., 2007], which has led to researchers estimating here in Finland [Ali et al., 2010] and elsewhere [Currie et al., 2004, Seifried et al., 2011, Benjamin and Whitaker, 2011, Drackley et al., 2012, Akita et al., 2016, Roh et al., 2020] that an aging demographic leads to increased blood demand on a national level.

The literature is, however, far from unanimous. Borkent-Raven et al. [2010] challenge the worst estimates of increased demand in the Netherlands

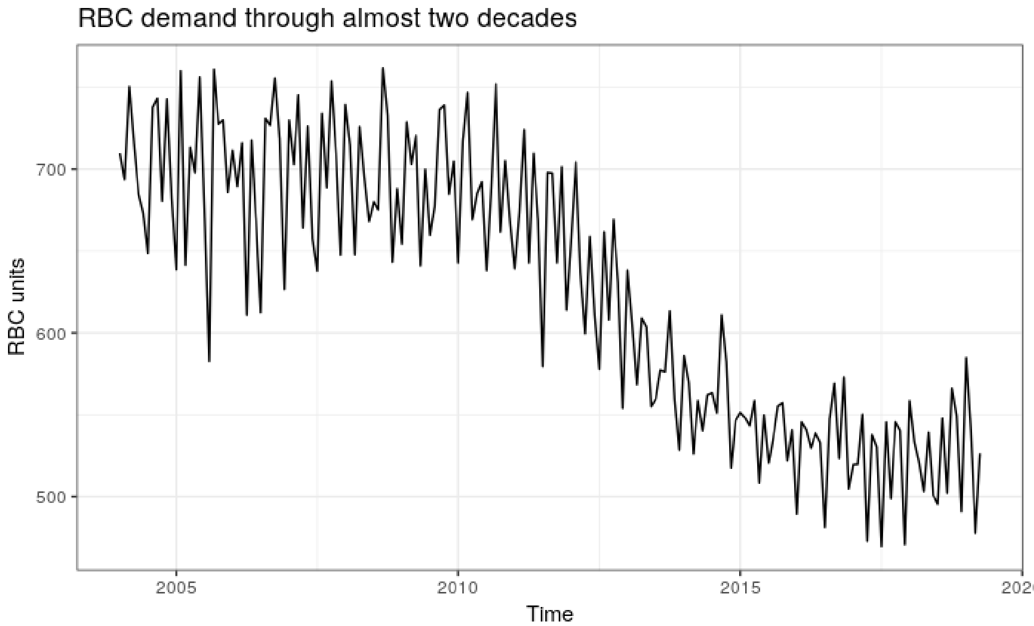


Figure 2.2: Observing the RBC demand on a national level in Finland between 2004 and 2019, we can see a significant decrease in the series level, most likely attributable to hospitals beginning to use 1 unit transfusion schemes more widely. We can also estimate a slight increase in variance towards the end of the series.

by using the current trends in patient blood management (PBM) in their model, thereby ending up estimating a decrease in demand instead of an increase. Volken et al. [2018] follow suit and suggest that based on their models, PBM schemes may entirely counteract the increase in demand caused by the aging Swiss demographic. Moreover, Laurén et al. [2019] note that the RBC demand trend in Finland from 2011 to 2016 is a decreasing one, despite the lack of coordinated PBM programs and an aging population. The authors suggest that this might be the result of improved surgical techniques, decreasing birth-rate, and increasing adherence to single-unit transfusions.

The convoluted nature of the long-term trend estimation is underscored by a qualitative review by Sasongko et al. [2019], in which they enumerate (often conflicting) expert opinions and multiple possible factors affecting the long-term trend. The authors end up concluding that the most prominent trend seems to be the decline of RBC demand.

Chapter 3

Current forecasts

The current forecast setup at FRCBS was implemented in 2013. The pipeline (Figure 3.1) consists of a manually amended monthly sales text document and an R script that runs on the same machine. The text document has three tab-separated columns: red cell products, platelet products (PLT), and fresh frozen plasma (FFP). The script fetches this document by name, extracts the data, removes missing values and feeds it into `ets()` and `stl()` modelers from the `forecast` package [Hyndman and Koehler, 2006, Hyndman and Khandakar, 2008] with a 6-month time window, which then output a 12-month forecast. A third 12-month forecast is also created by repeating the last six months of the series twice. The script then produces bar plots from the sums of the forecasts by blood product (RBC, PLT, FFP) and line plots by joining the history of the series and the forecast year for each product and forecast type. These figures are compiled into a bare-bones report in Portable Document Format (PDF) and uploaded to a destination folder. This report is used in financial year budgeting, but not operatively in donor mobilization or storage management efforts.

This thesis proceeds with explaining the functionalities of the used modelers, establishing metrics for performance evaluation, and finally evaluating the performance of the current forecasts. The third forecasting method (twice repeated 6-month history) is omitted as uninteresting and trivially inaccurate.

3.1 ETS modeler

The `ets()` modeler from the `forecast` package finds the best model (by minimizing Akaike Information Criterion [Akaike, 1973]) from the family of exponential smoothing models [Holt, 1957, Winters, 1960]. Forecasts made

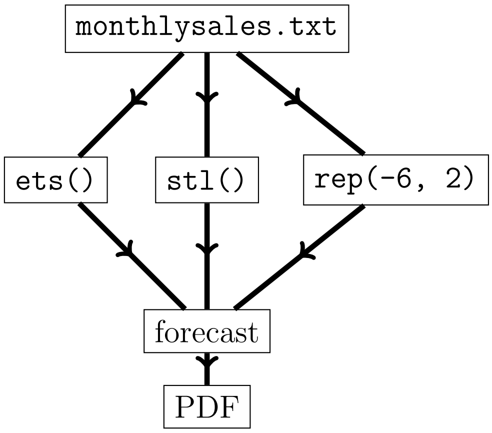


Figure 3.1: Diagram of the current forecasting pipeline at FRCBS.

using exponential smoothing methods are linear combinations of past observations so that the weights of the observations decay exponentially as we move further back in time. The simplest ETS model (without trend or seasonality parameters) is defined as

$$\hat{y}_{i+1} = \alpha y_i + \alpha(1-\alpha)y_{i-1} + \alpha(1-\alpha)^2 y_{i-2} + \cdots ,$$

where $0 \leq \alpha \leq 1$ is the smoothing parameter. The greater α, the less important distant observations become.

The ETS family is expanded by including components for trend and seasonality. These components can be applied either multiplicatively or additively, which in practice means that their effect either changes with the level (of the series) or it does not. Also, all of these models can be applied with either multiplicative or additive errors. The ETS family models can be specified by writing ETS($\cdot, \cdot, \cdot$), using A, A_d, M, or N (additive, additive damped, multiplicative, and none) for Trend, Seasonality, and Errors like so: ETS(A, N, N) [Pegels, 1969, Hyndman et al., 2008]. An ETS model with no trend or seasonality corresponds to the simplest ETS model defined earlier. An example of applying ETS models to an RBC sales series is presented in Figure 3.2.

3.2 STL modeler

The `stl()` modeler in the `forecast` package decomposes the series into a trend, a season, and an error component using locally estimated scatterplot smoothing (LOESS, known more commonly as a the Savitzky-Golay filter [Schafer, 2011]), removes the seasonal component from the series and feeds

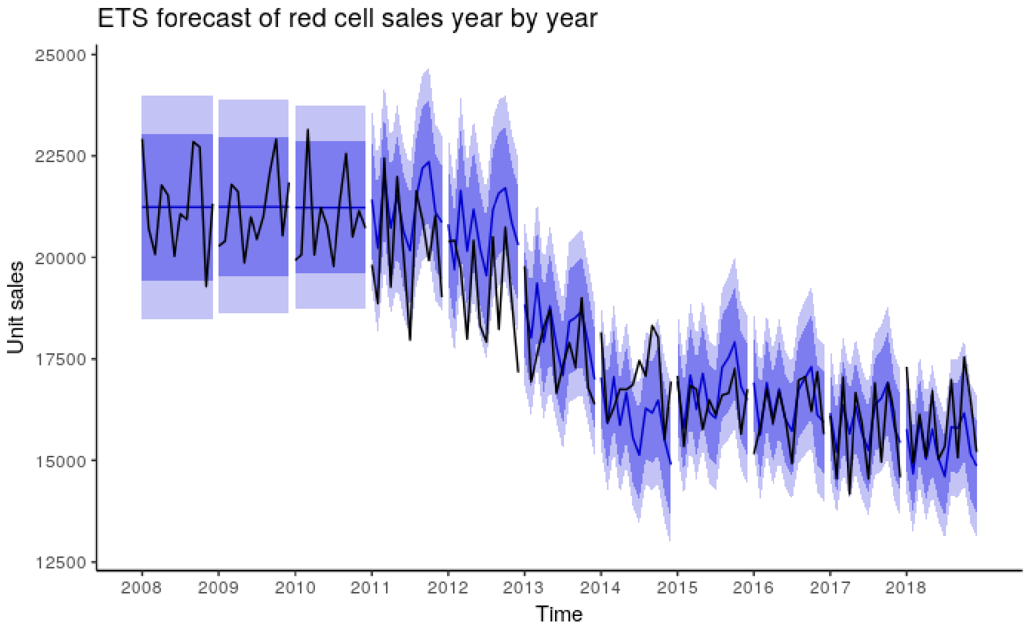


Figure 3.2: Result for running `ets()` piecewise (12 month forecast batches) on the entire red cell sales history. Original series marked in black, forecast and its 85% and 90% confidence intervals (CI) in shades of blue. At the beginning of the series, the modeler does not have enough data to find the seasonal component of the series, so it chooses an ETS(N,N) model for the first three years.

the seasonally adjusted series into the `ets()` modeler. For illustration purposes, Figure 3.3 presents a decomposed red cell sales history.
The way the STL modeler works means that it often agrees with the ETS modeler. The STL modeler should only ever disagree with the ETS modeler when it has a hard time finding the seasonal component or when there is not enough data to find it. This relationship is apparent comparing the entire forecasting histories: the STL modeler can take advantage of the seasonality in the series earlier than the ETS modeler (Figure 3.4).

3.3 Metrics

The most widely used performance metrics in the literature were MAPE and RMSE. We will settle for using mainly the same metrics throughout this thesis, but we also wanted to consider a metric that was more suitable for an operation critical context.

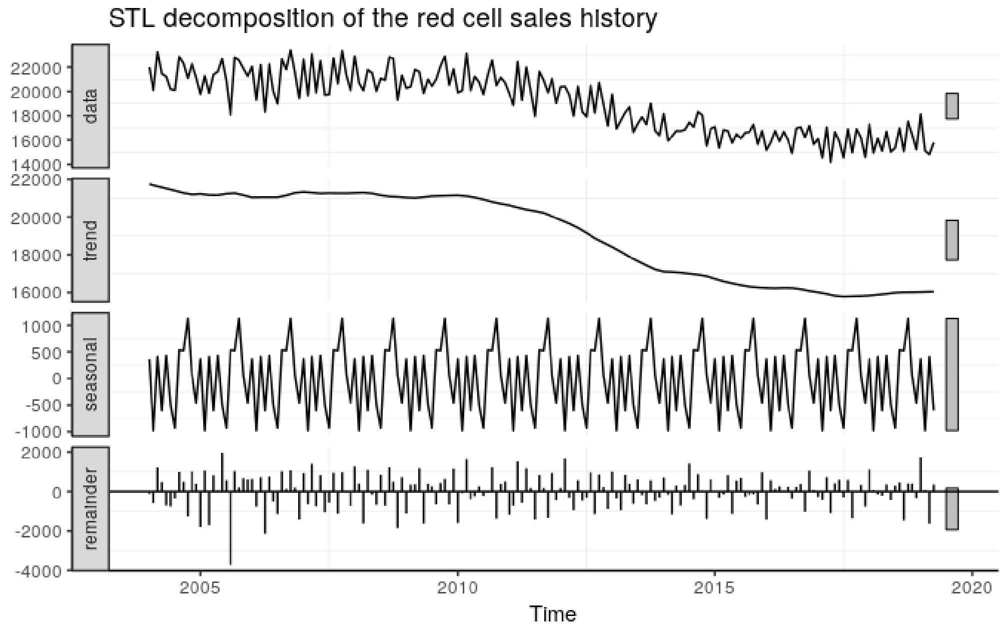


Figure 3.3: An STL decomposition. Original series up top, then trend, seasonal and remainder terms descending. The grey bar measures the "effect size" of the component: the more considerable the bar, the smaller the significance when summing back up to the original series.

3.3.1 Mean absolute percentage error

MAPE is defined as

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \cdot 100, \tag{3.1}$$

where $\hat{y}$ represents the forecast, y represents the actual observation, and n is the number of forecasts made. Percentage errors are unit-free, which means they can be used between different data sets and magnitudes. MAPE suffers from instability caused by division by zero: when y_i is zero, MAPE results in *undefined* or *infinity* depending on the implementation. Additionally, if y is a continuous variable, values close to zero result in extreme error values. Finally, MAPE requires that the data exist on a ratio scale (units have a meaningful zero; data does not reach beyond zero). However, we avoid all of these problems: our series can only have discrete, positive integer values (units of blood), and the demand is virtually never zero.

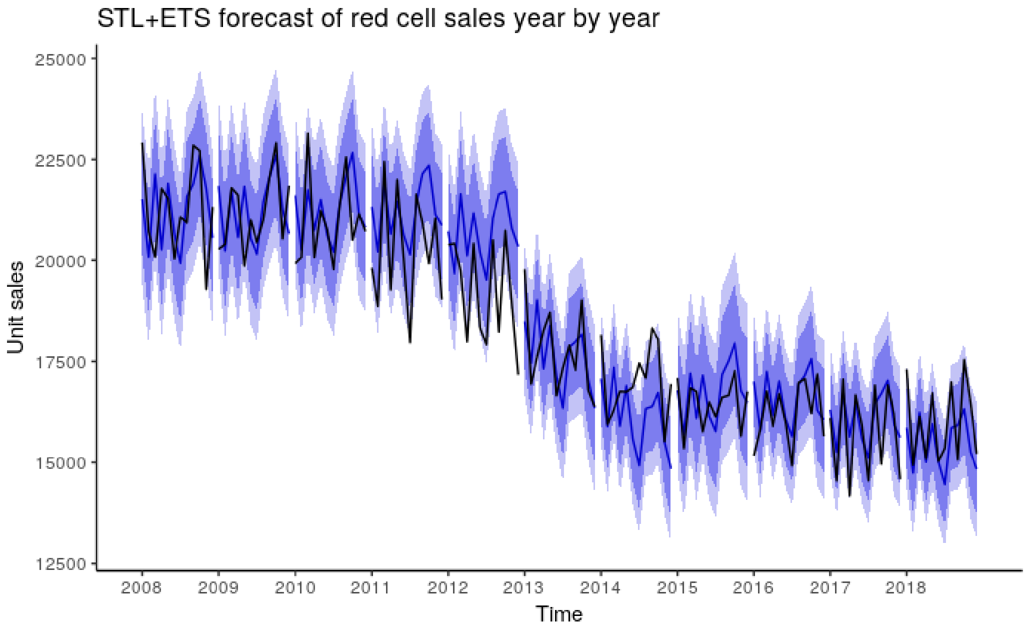


Figure 3.4: Result for running `stl()` piecewise (12 month batches) on the entire red cell sales history. Original series marked in black, forecast and its 85% and 90% confidence intervals (CI) in shades of blue. This modeler "finds" the seasonality much earlier than the ETS modeler.

3.3.2 RMSE

RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}. \tag{3.2}$$

RMSE is robust near zero and does not assume a ratio scale for the data. It is a scale-dependent error metric, meaning that the errors are on the same scale as data. However, comparing errors between different scales or units becomes very difficult with RMSE, and as such, the forecast errors from the total RBC series will not be comparable with the RMSE for the O+ demand series, for example.

3.3.3 Operation critical MAPE

Operation critical MAPE, or cMAPE for short, is a crude ad hoc metric for selecting models that overestimate rather than underestimate the blood demand by much discouraging underestimating demand. We accomplish this by doubling the error every time the demand is underestimated:

$$\text{cMAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{e^*}{y_i}\cdot 100, \quad e^* = \begin{cases} 2e & \text{if } e = |\hat{y}_i - y_i| < 0 \\ e & \text{if } e = |\hat{y}_i - y_i| \geq 0. \end{cases} \tag{3.3}$$

3.4 Evaluation of old forecasts

To properly evaluate the accuracy of the current forecasting models, we also evaluate the performance of a selection of the simplest known forecasting models. Assuming that the demand series is not merely a random walk and that it exhibits complex seasonality, these simple models should not outperform the ETS or the STL modeler.

3.4.1 Naïve model

A naïve forecast expects the next observation in the series to be the same as the previous one:

$$\hat{y}_{i+1} = y_i. \tag{3.4}$$

Naïve models work the best when the modeled series approximates a random walk: in the presence of true randomness, the best forecast is the previous observation. The `forecast` package contains two functions for naïve modeling: `naive(y, h)` and `rwf(y, h)`.

3.4.2 Naïve model with a drift parameter

It is possible to allow for drift in the random walk by adding the average change in the series to the previous observation:

$$\hat{y}_{i+h} = y_i + h(\frac{y_i - y_1}{i - 1}). \tag{3.5}$$

Naïve with drift can be called with `rwf(y, h, drift = TRUE)`.

3.4.3 Seasonal naïve model

The naïve model can also be adapted to incorporate seasonality:

$$\hat{y}_{i+h} = y_{i+h-m(k+1)}, \tag{3.6}$$

where m is the seasonal period (for example, 12 months), and k counts the number of full seasons in the forecast period. In practice, seasonal naïve

Model performance			
Method	cMAPE	MAPE	RMSE
ETS	7.636	4.826	1066.584
STL	8.238	5.204	1102.804
RWF	11.942	11.337	1891.135
NAIVE	12.296	11.760	1970.817
SNAIVE	**6.785**	**4.141**	**755.046**
MEANF	20.899	20.899	3427.415

Table 3.1: Comparison of accuracy metrics (cMAPE, MAPE, RMSE) between different models. Top down, `ets()`, `stl()`, `rwf()`, `naive()`, `snaive()`, `meanf()`. Best performing model in bold.

rewinds series time by the length of the season and gives that value as the forecast. This model works best when the series has strong seasonality with minimal variance. The seasonal naïve model can be called with `snaive(y, h)`.

3.4.4 Averaging model

Another take on the ideas behind naïve models is the averaging method, where the forecast is achieved by averaging all previous observations:

$$\hat{y}_{i+1} = \frac{y_1 + \cdots + y_i}{n}. \tag{3.7}$$

The averaging forecast is called with `meanf(y, h)`.

3.4.5 Results

The current forecasting practice is to use `stl()` and `ets()` to forecast the next 12 months of demand for all of the red cell products combined, so we test them in the same setup. History from 2004 to 2018 was used in creating a 12-month forecast. Model performance was recorded using cMAPE, MAPE, and RMSE. Results are presented in Table 3.1.

The clear winner seems to be the seasonal naïve with the smallest errors in all categories. The ETS and STL models come quite close with naïve, naïve with drift, and averaging method being far behind in performance. Interpreting the RMSEs and cMAPEs, the current models misestimate around 1100 units of blood for 12 months, with somewhat equal distribution between over- and underestimating demand.

Chapter 4

Studying the demand series

4.1 Data

The literature recognizes two kinds of time series that can be seen as demand data: data on actual transfusions performed in hospitals and data on sold/shipped blood units from collection centers. The FRCBS has full access to the latter and partial access to the former through the Ketju program [Rautonen, 2007]. Using the transfusion data is preferable to sales data, as it is the most direct measure of blood demand. The most beneficial scenario would be one where the transfusions are logged immediately as orders for the blood service, and the blood service could then scale collections accurately in real-time. Currently, we do not have such a system in place, so we will have to resort to demand forecasting. Unfortunately, only a part of all of the shipped blood goes to Ketju hospitals, and because hospitals differ in patient blood management schemes [Laurén et al., 2019], storage management procedures and ordering schedules, it is difficult to create any meaningful forecasts for the operational level using this data.

As such, we are left with blood sales data. To facilitate automatic forecasting, we pull the sales data directly from a business intelligence platform used by the FRCBS. Blood product sales data is automatically updated daily on the platform, which removes the problem with manually amending text documents, and we can perform our modeling with finer temporal resolution.

These daily sales data contains product codes for all sold products, which we filter to obtain the series for all types of RBCs and PLT. We decided not to implement forecasts for FFP products as their ordering and shipping policy is entirely different from RBC and PLT products.

Finally, we considered accessing national morbidity and traffic accident statistics and weather data as exogenous regressors for some of our models for a while. However, we did not have enough time to explore this data, so the use of exogenous regressors was ultimately abandoned. All of the modeling done in this thesis can be considered autoregressive.

4.2 Demand behaviour

Having access to the daily sales data lets us aggregate it as a weekly time series. The RBC series (all types combined) then reveals a striking transformation: the entire nature of the series appears to change around 2017.

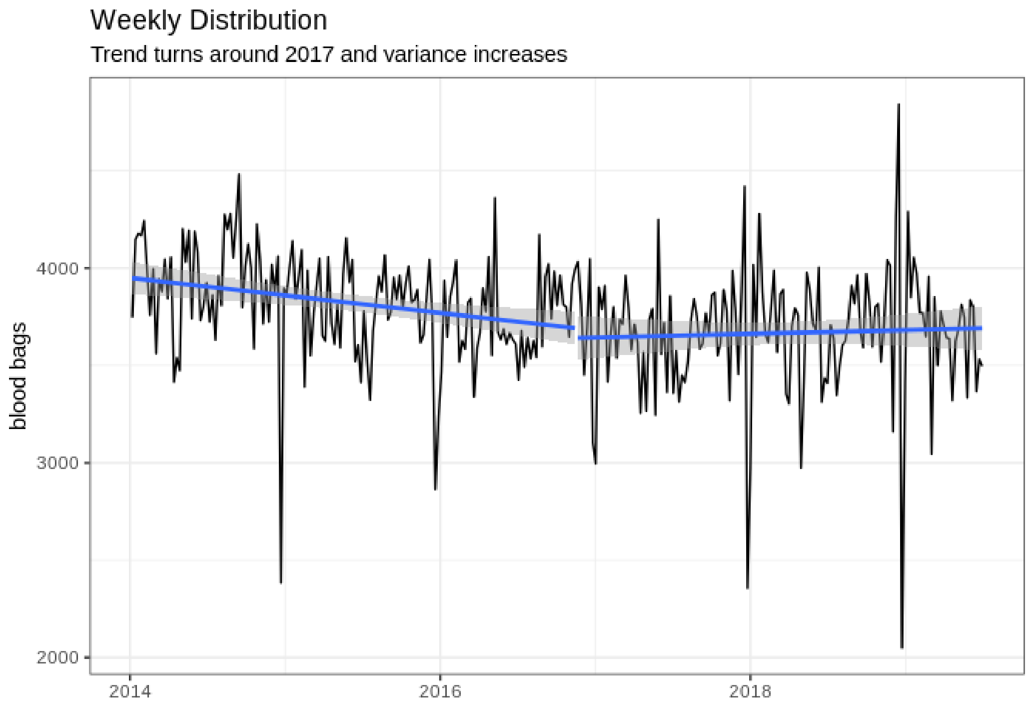


Figure 4.1: Weekly series of sold RBC turns from a decreasing "random walk" into a slightly increasing signal with stronger seasonality and a more significant variance.

This finding primes a critical question for a data science practitioner: how does one consider the possibility that the best model for the current state of the series might not be the best in a year or two? To drive this home, Figures 4.2 and 4.3 show how an STLF modeler (covered in Section 5.4.4.1) arrives at entirely different models based on the length of history shown to it.

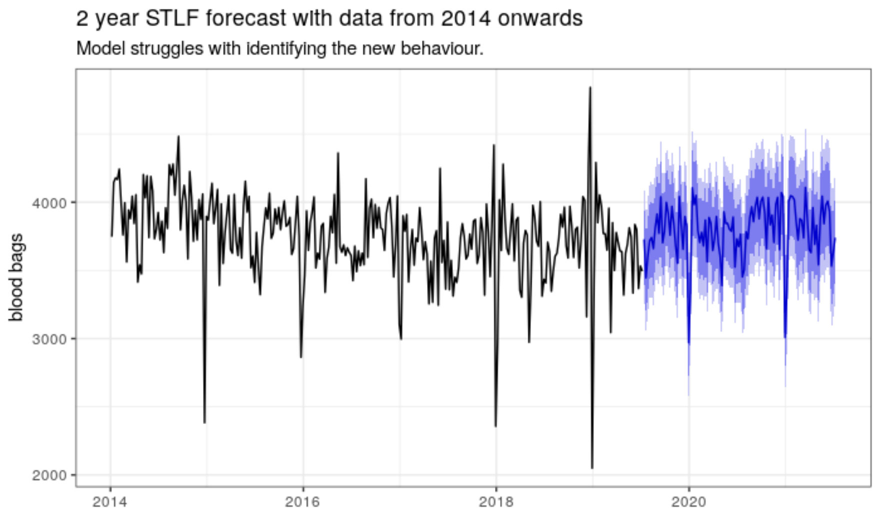


Figure 4.2: 2 year STLF forecast. Modeler was given data from 2014 to 2019.

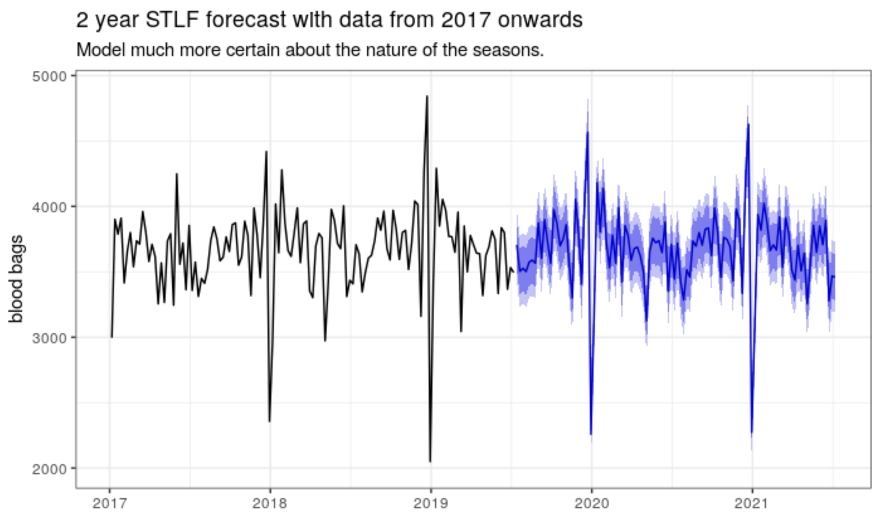


Figure 4.3: 2 year STLF forecast. Modeler was given data from 2017 to 2019.

The found literature on blood demand forecasts so far does not comment on the issue of adaptability, although some authors do remark on the need to review the forecasts periodically [Wilding et al., 2006]. However, the data makes it evident that if one aims to deploy a semi-autonomous forecasting script without constant supervision by an analyst, the system needs to be able to choose from a more extensive library of models than just, for example, a family of ETS functions. This adaptability is the main idea behind our final implementation.

Finally, we present some tangential findings uncovered while doing exploratory analysis on alternative data sets for the sake of completeness.

4.2.1 The need for monthly adjustments

The current forecast pipeline contains very little data preprocessing. After missing values are removed from the data, it is transformed directly into a time series object. However, it is advised and customary to adjust for possible variability caused by calendar months, especially if the data is a monthly series to remove the correlation with the number of days in any given month [Hyndman and Athanasopoulos, 2018]. It is easy to show how this creates artifacts into our series by plotting a seasonal subseries plot (Figure 4.4).

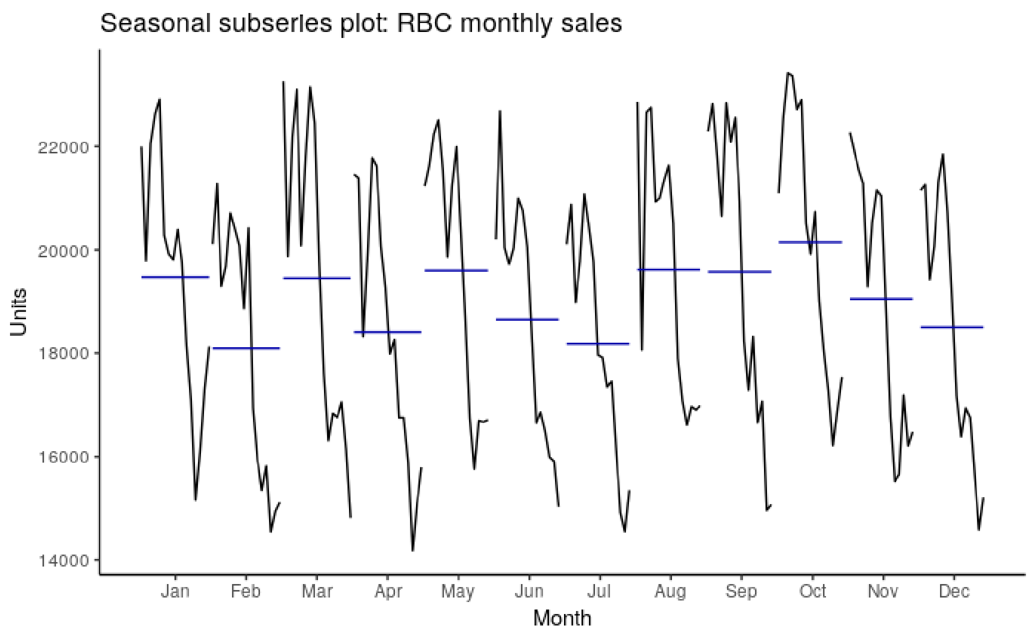


Figure 4.4: Red cells monthly series as a subseries plot. Each of the subseries depicts values for the indicated month throughout the entire history. The blue line indicates the mean, clearly going lockstep with month duration in some parts.

We can immediately remove some of this noise by adjusting for days in a month (Figure 4.5). The perfect adjustment counts only operational days, excluding, for example, all holidays. Such a function is not readily available for the Finnish calendar, so we have to rely on the number of market days in Zürich, the closest approximation offered by the `bizdays()` function in the `forecast` package.

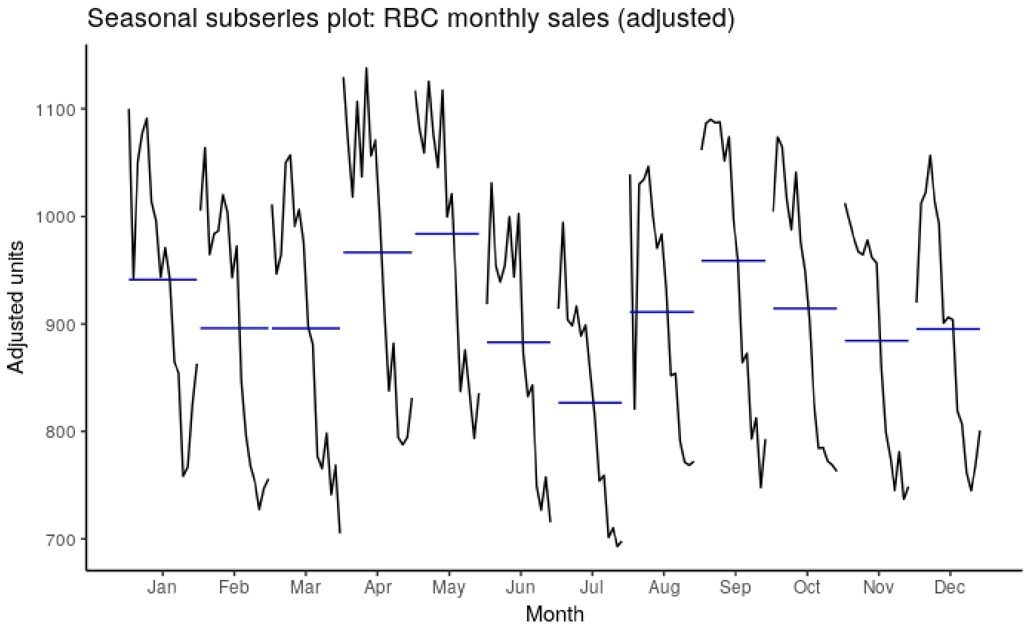


Figure 4.5: Red cells monthly series as a subseries plot. Each of the subseries depicts values for the indicated month throughout the entire history. The blue line indicates the mean, now revealing more real variation between months.

This small additional preprocessing step already improved the current forecasts by a significant margin, as shown in Table 4.1. The RMSE here is not comparable with the previous RMSE due to the adjustment, but cMAPE and MAPE indicate that the adjustment improved all models' performance, except for seasonal naïve. The systemic monthly variance may have given seasonal naive an unsubstantiated edge over other models. With one additional preprocessing step, we can improve model performance by 9-50% across the board.

4.2.2 Transfusions by weekday, expiration cycles

Transfusion data from Ketju hospitals reveal a couple of curiosities. Figure 4.6 reveals that transfusions are given mainly on business days, indicating possibly that fewer operations are performed and fewer treatments are given on weekends.
It is also interesting how blood freshness is managed in hospitals. The Ketju data contains information about the freshness of the transfused blood in the form of "days till expiration." Plotting these in the form of a histogram (by the hospital or by blood type), a pattern of use is revealed (Figures 4.7 and

Model performance after adjustment			
Method	cMAPE	MAPE (Δ%)	RMSE
ETS	7.657	4.400 (-8.8)	44.436
STL	6.857	**4.050** (-22.2)	41.640
RWF	10.477	5.938 (-47.6)	60.322
NAIVE	10.020	5.917 (-49.7)	57.375
SNAIVE	**6.785**	4.141 (+0.0)	**36.951**
MEANF	17.557	17.557 (-16.0)	143.651

Table 4.1: Comparison of accuracy metrics (cMAPE, MAPE, RMSE) between different models after adjusting for the approximate number of business days in a given month. Top down, `ets()`, `stl()`, `rwf()`, `naive()`, `snaive()`, `meanf()`. Best performing model in bold.

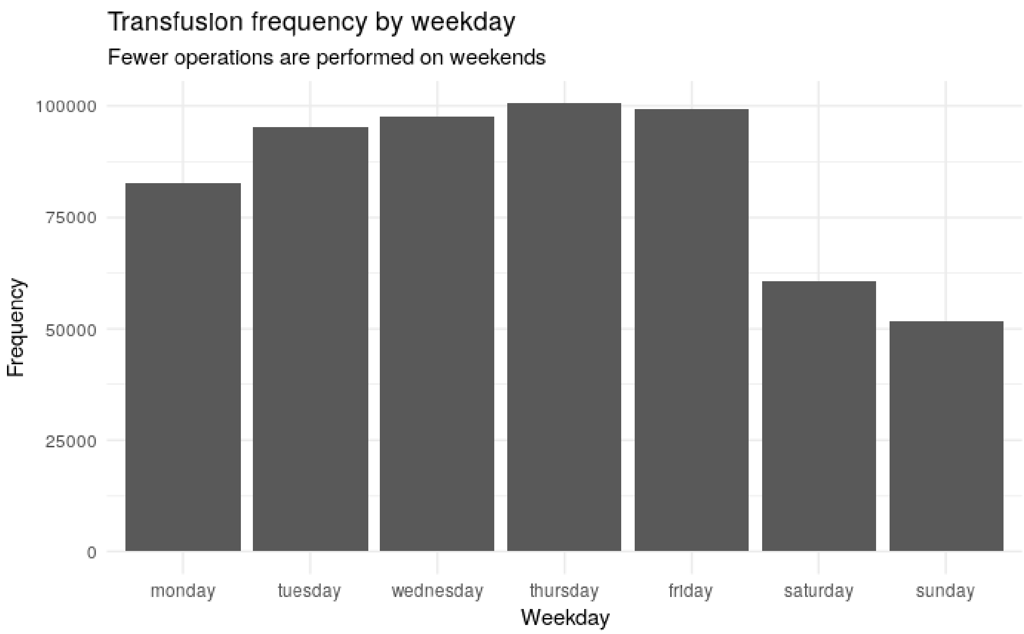


Figure 4.6: Transfusion data from Ketju hospitals.

4.8).
It would seem that hospitals manage their blood inventory in such a manner that results in units 21, 14, and 7 days from expiration being administered the most. This phenomenon suggests the use of a weekly rearranging scheme.

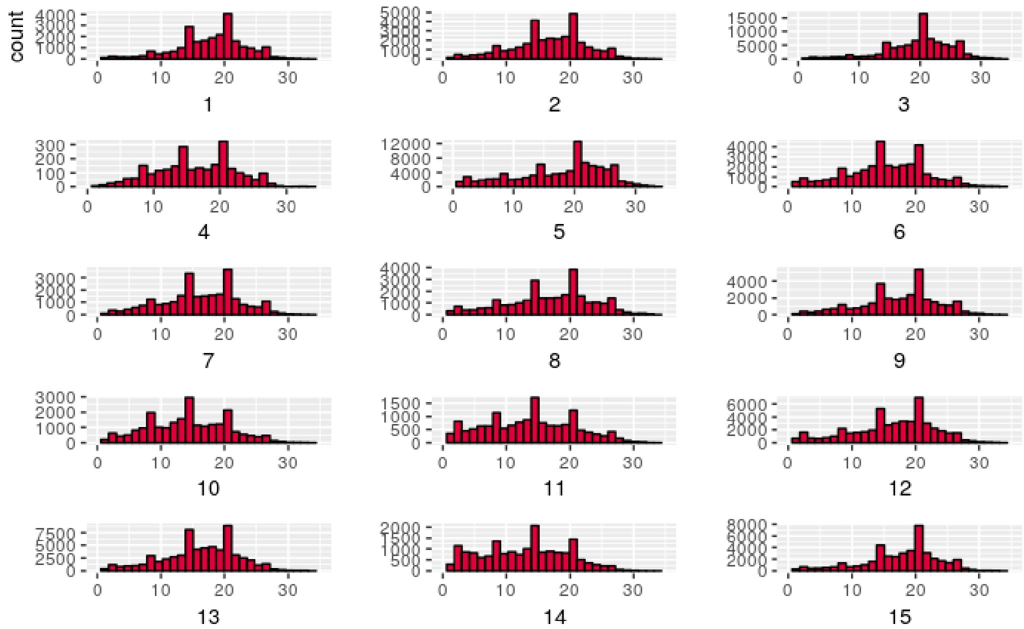


Figure 4.7: Blood unit expiration data from Ketju hospitals by the hospital. Hospital names changed arbitrarily to numbers to as it may be sensitive data. X-axis indicates the number of days the administered unit would have expired in.

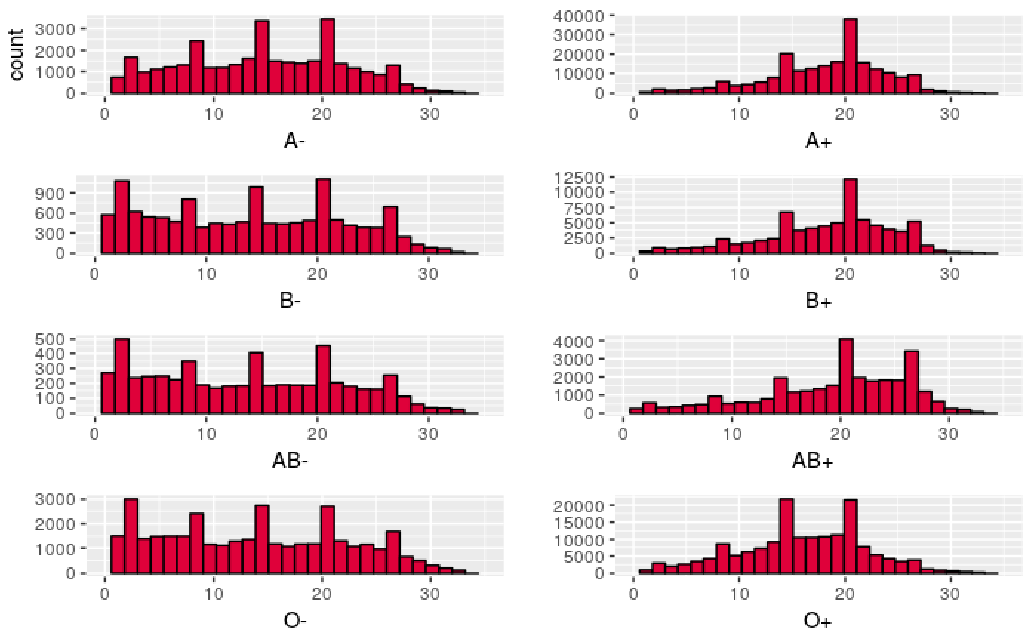


Figure 4.8: Blood unit expiration data from Ketju hospitals by blood type. X-axis indicates the number of days the administered unit would have expired in.

Chapter 5

Implementation

The existing literature on forecasting operational blood demand reveals that there is not yet a singular model or approach that works the best for this kind of demand data, as the factors driving demand in different countries might be entirely different (as is the case with Tanyavutti and Tanlamai [2018]). Furthermore, it is crucial to note that the nature (trend, seasonality, total variance) of the blood demand series might change with a surprising rate [Benjamin and Whitaker, 2011], unexpectedly [Ali et al., 2010] and in ways that are difficult to anticipate [Wilding et al., 2006, Sasongko et al., 2019]. These are all pressing reasons for avoiding selecting a singular best performing model and opting for a semi-autonomously adapting forecasting system instead.

Designing an adaptive (univariate, autoregressive) forecasting system requires several choices from the analyst:

1. **Should we be able to allow for optional data processing steps?** For example, it could be that during a period of low signal-to-noise ratio, some low pass filtering might improve the forecast accuracy by a significant margin.

2. **By what metric should we choose our forecasting models?** If we aim to forecast on an operational and tactical level, the metrics should be easily interpretable and lead to a useful model selection.

3. **What is the library of models we want the modeler to choose from?** This is mostly a question of available computing resources because we can not have the run times for any given forecast to get out of hand. However, properly justifying a model choice may be difficult when the pretext is that we want to prepare for unforeseen changes in the demand signal.

4. **How long is our training period? How long is our validation period? How far ahead are we aiming to forecast?** The recommended size of the validation set usually varies between 10% and 25%, but it should cover at least the forecasting period.

5.1 Code

All development was performed on Ubuntu 18 OS. All scripts during development were written with R version 3.6 [R Core Team, 2020] using Rstudio development environment version 1.2 [RStudio Team, 2020]. Session info (`sessionInfo()`), containing also loaded packages and their versions is included in Appendix A. All development code will be publicly available at `https://github.com/FRCBS/production_forecasts` at some point.

5.2 Data preprocessing

Data selection and product filtering are explained in Section 4.1. To further process the data for forecasting, we replace all missing values with zeroes to avoid issues with `NaNs`. We do this instead of imputing with averages because we estimated that a missing value in the business intelligence data is much more likely to be an unreported zero value than an actual missing data point. We aggregate all of our daily series into weekly and monthly variants to forecast on both an operational and a tactical level. After aggregation, no zero-valued weeks or months are left in any of our series, ensuring that we can safely use percentage error metrics, such as MAPE. Finally, we adjust the monthly series by workdays, as described in Section 4.2.1.

5.2.1 Additional preprocessing options

To account for worsening signal quality (increasing signal-to-noise ratio $\frac{S}{N}$), we can allow for the application of low pass filters on our series. Figure 5.1 shows the effect of low-pass filters with cutoff frequency settings of 0.25 and 0.10. For smoothing we use the package `istmr` and the function `smooth.fft(x, f)` within.

5.3 Final choice of metric

Gardner [1979] remarked the disadvantages of deploying a modeler that would require constant or at least regular expert supervision. An autonomous

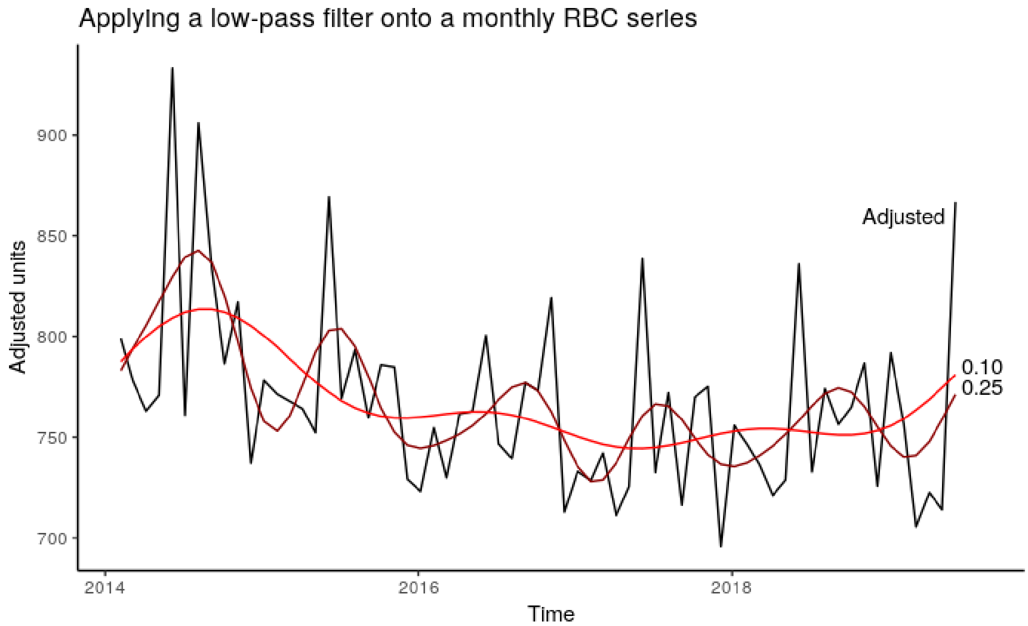


Figure 5.1: Low-pass filtering the monthly RBC sales series. The lower the cutoff frequency, the smoother the resulting series.

and adaptive modeler solves this particular issue, but only if the forecast target audience trusts its judgment. In the context of blood demand forecasts, good judgment can be interpreted, for example, as the preference to overestimate demand at the cost of most accuracy metrics. Pereira [2004] used *coverage rate* as a metric for model goodness and Fortsch and Khapalova [2016] used mean percent error (MPE) to find if the model was biased to a specific direction, although deciding that 0 bias was the optimum. In this thesis, we have used the *operation critical MAPE*, which conveys the same information as MPE, if in a less interpretable form. The upside is that it can be used as-is as a model selection metric, unlike MPE.

Currently, this development project is exploratory, and the forecasting system developed will not be used operationally in the near foreseeable future, but as a complement to budget planning. This shifts the weight from operational criticality to metric legibility, meaning that models should be selected strictly by forecast accuracy. Thus, our final modeler will select models based on MAPE and not cMAPE. RMSE is another strictly accuracy focused metric, but as the final implementation will be used to forecast ten different blood products (RBC, ±A, ±B, ±O, ±AB, PLT) on two different time scales (monthly and weekly), RMSE does not allow us to compare inter-series accuracies.

5.4 Model selection

Now that we have established the appropriate model selection metric and the data for our forecasting system, we can run performance analysis on a handful of the most widely used models. We will test out most of the models tested in the literature and see if they exhibit any properties that warrant the inclusion into the model library used by the modeler. We use monthly aggregate data for model selection, as monthly forecasting is our primary goal for our implementation. Our training set extends from February 2014 to June 2018, and our validation set comprises the following 12 months to June 2019.

5.4.1 Simple moving averages

A simple moving average model (SMA) of qth order is defined as:

$$\hat{y}_{i+1} = \frac{1}{q} \sum_{j=i-q}^{i} y_j. \tag{5.1}$$

SMA models are an extension on the averaging method, where we select a time window inside which we average all observations to generate a forecast. The idea is to smooth out the most erratic parts of the series to find its trend-cycle [Hyndman and Athanasopoulos, 2018, chapter 6.2].

5.4.2 Dynamic regression

Linear regression means fitting a line through the existing observations.

$$y_t = x_t b + a + \varepsilon_t, \tag{5.2}$$

where b is the slope of the line, a is the intercept (the level) and ε captures "everything outside a linear relationship" between the variables x and y, often taken as white noise. Forecasting with linear regression means that we follow that fitted line into the future:

$$\hat{y}_t = x_t b + a, \tag{5.3}$$

and we can generalize univariate linear regression into multivariate linear regression by adding dimensions:

$$\hat{y}_t = \sum_{i=1}^{n} x_{i,t} b_i + a. \tag{5.4}$$

However, we are not interested in evaluating the performance of simple linear regression, as it can only find linear relationships [Hyndman and Athanasopoulos, 2018, chapter 5.1], and we already know by having taken a look at our series that the relationship between demand and time is most certainly not linear. There are various ways to transform linear models into nonlinear models, as explained in Hyndman and Athanasopoulos [2018, chapter 5.8], but we will instead approach the problem via dynamic regression by assuming that the noise term ε contains enough auto-correlation to be relevant for forecasting. We can efficiently model the auto-correlation contained in ϵ using ARIMA modeling (Section 5.4.3 below). If we combine the linear regression model and the ARIMA error model, we can model nonlinear relationships between the dependent and independent variables. In practice, we run a linear regression modeler using `tslm()` from the `forecast` package and then run `auto.arima()` on its residuals.

5.4.3 ARIMAX

As stated in Section 2, ARIMA models are generalisations of ARMA models, which can be defined with two polynomials: the autoregressive polynomial and the moving-average polynomial. An AR-polynomial of order p is defined as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{5.5}$$

which can be interpreted as multivariate linear regression using p lagged values of the series as regressors. A q-order MA-polynomial is given by:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \tag{5.6}$$

which can be thought as a regression formula with moving-averages of the past forecast errors. A full ARMA model can then be written as:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t. \tag{5.7}$$

ARIMA models take an additional differencing step to ensure series stationarity. ARIMA models can be identified using the following notation: ARIMA(p, d, q), where p and q signify the orders of the AR and MA-polynomials, respectively, and d indicates the differencing degree. ARIMA models can be further expanded by introducing exogenous regressors:

$$y_i = \beta x_i + \sum_{j=1}^{p} \phi_j y_{i-j} + \varepsilon_i + \sum_{j=1}^{q} \theta_j \varepsilon_{i-j}. \tag{5.8}$$

As discussed in Section 4.1, we also considered using various other time series (weather, accident statistics) as exogenous regressors but ended up deciding against this on the grounds of schedule. However, we can bestow a sense of time to an ARIMA model by using a one-hot encoded monthly calendar as an exogenous variable, effectively creating an ARIMAX model. An example of a one-hot encoded monthly calendar is shown in the Figure 5.2. The `auto.arima()` function also has the capability to take in exogenous regressors, and thus it functions as our ARIMAX modeler.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 5.2: One year in one-hot encoded months. Each row and column represents a specific month starting from January. The rows extend to December while the columns extend only to November. This is because we want December to be captured by the intercept, so as to avoid the *dummy variable trap* detailed, for example, in [Hyndman and Athanasopoulos, 2018, chapter 5.4].

5.4.4 Complex decompositions

The existing literature does not contain any attempts to forecast blood demand with models designed to detect complex seasonalities, even though the blood demand signal is likely to have weekly, monthly and yearly seasonalities. The seasonality detecting models tried so far (STL, ETS) are designed for finding a unique seasonality in the series, which warrants exploration of models designed to detect multiple seasonalities.

5.4.4.1 STLF

One way to detect multiple seasonalities in a series is to detect them recursively one at a time. This is what the `stlf()` from the `forecast` package tries to accomplish. It calls the forecasting function using a fit derived with the `mstl()` function, which works by calling `stl()` recursively multiple times, stripping away seasonalities layer by layer [Hyndman and Athanasopoulos, 2018, chapter 11.1].

5.4.4.2 TBATS

TBATS (Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components) models build ARIMAX models by using *Fourier terms* derived using *dynamic harmonic regression* as exogenous regressors, as a representation for series seasonalities. As neither Fourier analysis nor dynamic harmonic regression is in this thesis's scope, we will sidestep here the theoretical background of the model by referring to De Livera et al. [2011]. Nevertheless, TBATS allows for multiple *changing* seasonalities in the modeled series, and we will thus analyze its performance alongside STLF.

5.4.5 Neural network autoregression

Fanoodi et al. [2019] found that artificial neural networks outperformed ARIMA models in all blood type categories, which is a good reason to try autoregressive neural networks for ourselves. NNAR models feed lagged observations of the series into the network input layer and try to find the optimal weights for forecasting the future [Hyndman and Athanasopoulos, 2018]. The `forecast` package offers the `nnetar()` function for automatic NN fitting.

5.4.6 Combination forecast

Bates and Granger [1969] showed already some decades ago that combining the forecasts from different models, for example, by averaging, often result in improved accuracy. We will thus include a combined mean forecast from all of the chosen models in our evaluations.

5.4.6.1 Results

Table 5.2 presents the combined results of the performance evaluation of naïve, naïve with drift, seasonal naïve, averaging, SMA, STL, ETS, STLF, TBATS, dynamic regression, ARIMAX, NNAR, and combination forecast models on the RBC demand series with three possible levels of data processing: as-is and low-pass filters with cutoffs at 0.25 and 0.10. The same validation process is applied to all our smaller blood product series, and Table 5.1 presents the best model for each with the associated error scores.

The ARIMAX and combination forecast (CF) models outperform all other models by a hefty margin when forecasting the RBC series. Other better contenders are a TBATS model, an ETS model, and somewhat surprisingly also the seasonal naïve model. When forecasting by blood type (series with smaller volume), simple moving-average models tend to outperform more complex models. ARIMAX and seasonal naïve models achieved the highest accuracy among the best performing models. B minus proved to be the most challenging series to forecast.

Furthermore, it appears that currently, the low pass filters provide virtually no improvement to any of the forecasts and that we can probably safely omit the most simple baseline models from the final pool of models since only seasonal naïve appears competitive. Our final pool of models thus comprises

1. a seasonal naïve modeler, `snaive()`
2. an STL modeler, `stl()`
3. an ETS modeler, `ets()`
4. Simple moving averages, `ma(order = {5, 7, 9, 12})`
5. a dynamic regression modeler, `tslm() + auto.arima()`
6. an ARIMAX modeler, `auto.arima(xreg = one-hot-calendar)`
7. an STLF modeler, `stlf()`
8. a TBATS modeler, `tbats()`
9. a NNAR modeler, `nnetar()`
10. a combination forecast model (by averaging)

Best model or modeler for each series				
Series	Model	cMAPE	MAPE	RMSE
RBC	ARIMAX	3.531	2.068	22.387
A-	12-MA	7.21	3.706	2.484
A+	ARIMAX	5.026	2.624	9.345
B-	STLF	11.587	6.765	2.015
B+	9-MA	8.09	4.649	5.536
AB-	12-MA	7.848	4.724	0.468
AB+	5-MA	5.024	4.151	1.701
O-	ARIMAX	4.347	3.542	2.751
O+	SNAIVE	3.638	2.133	6.888
PLT	STL	4.684	3.288	5.259

Table 5.1: A table of best models measured in cMAPE, MAPE and RMSE when forecasting the each of the product subseries.

Model performance			
Method	cMAPE	MAPE	RMSE
RWF	11.838	11.623	92.611
RWF .25	6.233	4.123	42.862
RWF .10	6.756	4.196	43.84
NAIVE	11.474	11.177	88.622
NAIVE .25	6.136	4.195	42.55
NAIVE .10	6.474	4.197	42.838
SNAIVE	4.066	2.327	22.707
SNAIVE .25	6.832	4.243	42.951
SNAIVE .10	6.834	4.338	43.464
MEANF	6.005	4.596	43.993
MEANF .25	6.003	4.585	43.940
MEANF .10	6.002	4.581	43.917
STL	4.836	2.575	26.660
STL .25	11.3	5.816	63.039
STL .10	6.942	4.182	44.844
ETS	3.719	2.268	23.960
ETS .25	6.095	5.002	46.653
ETS .10	6.762	4.196	43.87
5-MA	5.286	3.347	35.894
7-MA	5.286	3.236	35.853
9-MA	5.248	2.975	34.721
12-MA	4.582	2.745	31.746
ARIMAX	**3.531**	**2.068**	**22.387**
ARIMAX .25	6.681	3.877	44.206
ARIMAX .10	6.867	4.208	44.187
DynR	8.334	4.167	38.909
DynR .25	6.504	3.992	41.54
DynR .10	9.805	4.929	58.027
STLF	4.648	2.472	26.159
STLF .25	7.568	4.192	47.776
STLF .10	6.571	4.159	43.044
TBATS	3.896	2.333	23.302
TBATS .25	3.896	2.333	23.302
TBATS .10	3.896	2.333	23.302
NN	6.940	3.882	33.906
NN .25	4.910	2.828	26.026
NN .10	5.738	3.253	29.55
CF	4.785	2.685	29.678
CF .25	6.002	3.575	38.573
CF .10	6.378	3.767	40.416

Table 5.2: A table of model performance measured in cMAPE, MAPE and RMSE when forecasting the entire RBC series. Best performing models in bold.

5.5 Rolling window size

We chose the size of the rolling window for our adaptive modeler by examining the speed at which the most recent significant behavioral change happened in the series and how long the new behavior has persisted so far. Based on these metrics alone, a rolling window size of two years is likely to be the most suitable option, as the new behavior has persisted now for some two years. However, we decided to make the modeler a bit more conservative, to be more robust under short, fleeting changes in the behavior of the signal. A validation period of one year further assures that each time a model is selected, it has been *consistently* the best throughout an entire year.

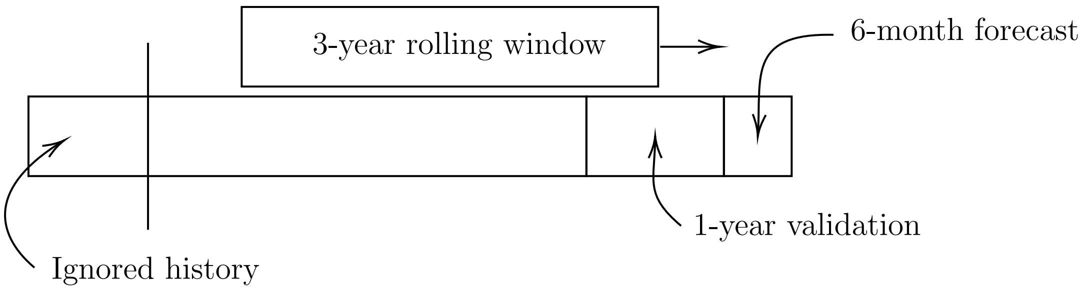


Figure 5.3: Diagram of the model selection process.

The purpose is to run the modeling script at pre-determined times each month or each week, depending on the desired forecasting resolution.

5.6 Reporting

The current forecasting implementation at FRCBS outputs a small collection of graphs as a Portable Document Format file and uploads it into the FRCBS intranet, where it is viewable by downloading it. When an employee needs to record the forecasts to file, they need to input the forecast values by reading the PDF file manually.

To provide more accessibility to the forecasts, we integrated the new modeler into an R Markdown script that is converted (”knit”) into a Hypertext Markup Language (HTML) file immediately displayable on the intranet. Using HTML, we can provide the users with means to interact with the reported forecasts, for example, by downloading a Comma Separated Values (CSV) file of a table of forecasts. HTML allows us also to present figures in tabs.

Our report begins with figures of 6-month forecasts for each of the product types (and RBC total). Each figure has a subtitle that tells the type of the

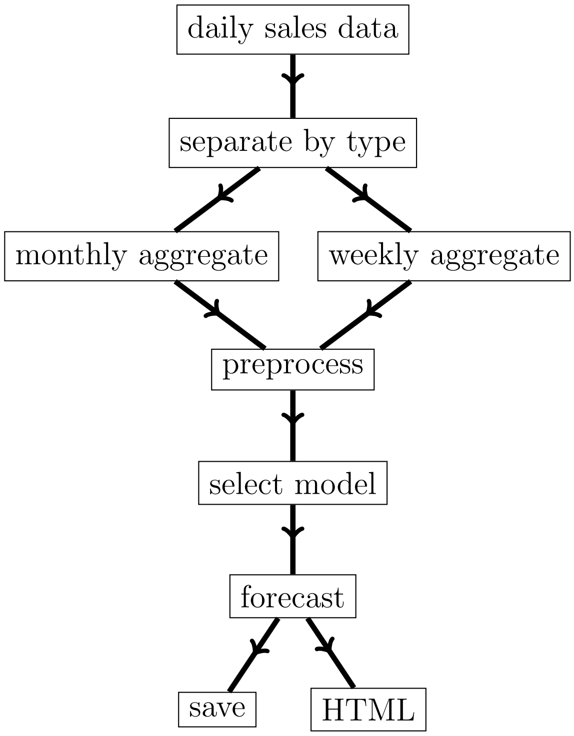


Figure 5.4: Diagram of the pipeline for the new modeler to be used at FRCBS.

model or modeler used for this six-month forecast. A development mock-up is shown in Figure 5.5. The shaded portions are the 80 and 95 percent confidence intervals for this forecast. Below each figure is highlighted the exact forecast for the next month. A tabled version follows the monthly forecast figures, easily saved by copying to the clipboard. The figures and tables cover the primary function for the modeler and the reporting, but they are followed by figures for four-week forecasts for exploratory purposes (Figure 5.6).

Finally, the report contains forecasting histories, both as figures and tables, for the monthly and weekly forecasts. Forecast histories are accumulated as new forecasts are saved each week. Histories are included to give insight into the behavior of the demand signals and the modeler and help users assess its reliability, hopefully facilitating trust in the system. Figure 5.7 shows simulated monthly forecast histories for the past 17 months and the corresponding history table detailing which models get selected the most and what are their respective historical accuracies.

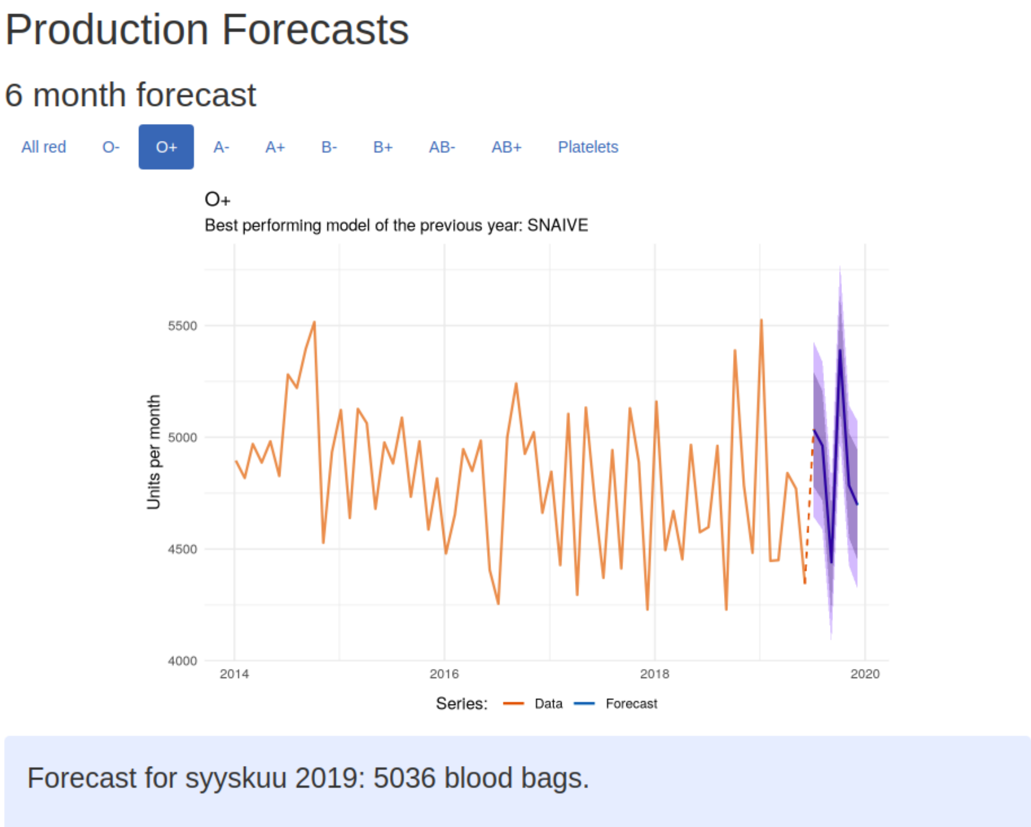


Figure 5.5: Screenshot of the generated HTML report. A 6-month forecast for the O+ type demand series is shown here, as indicated by the figure and the highlighted tab. The plot subtitle reports the best performing model or modeler of the previous year.

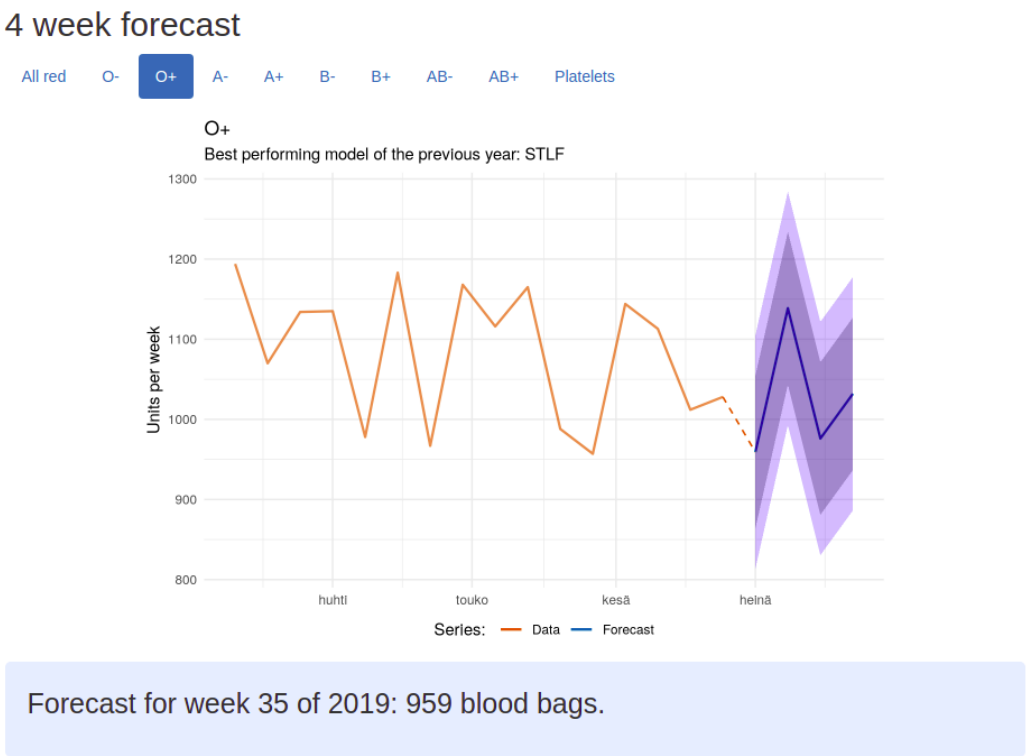
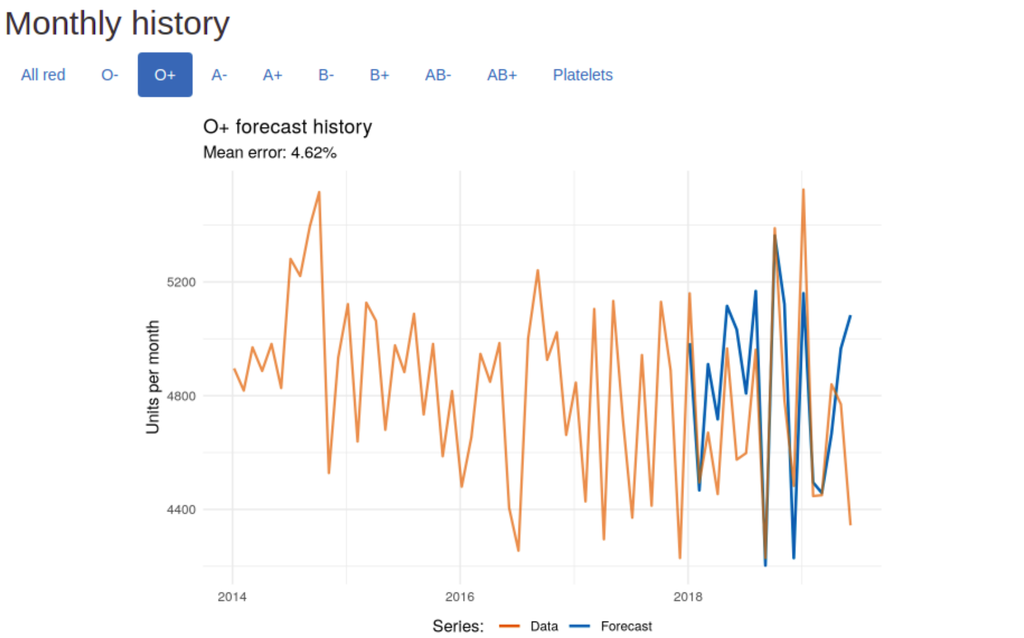


Figure 5.6: A 4-week forecast for the O+ type demand series. The previous year's best performing model is different between monthly and weekly series (seasonal naïve and stlf, respectfully).



Show 9 entries

Search:

	frequency	error
	All	All
SNAIVE	66.67	4.59
STL	33.33	4.68

Figure 5.7: Monthly forecast history for the O+ demand series, simulated using the adaptive modeler. The forecaster has opted for a seasonal naïve model $\frac{2}{3}$ of the times, and the remaining $\frac{1}{3}$ of the forecasts employ the `stl()` modeler. The MAPE for this forecast history is 4.62%, as indicated in the plot subtitle.

Chapter 6

Discussion

6.1 What we learned

This thesis aimed to evaluate the performance of the current demand forecasts at FRCBS and explore ways to improve them. We were able to improve the forecast accuracy of the original pipeline immediately by 22.2% by adjusting the data for monthly workdays. We reviewed the existing literature on blood demand forecasting and found that they are often incomparable in their results as the blood demand signal behavior varies from locale to locale. We also found that while there is a growing interest in the future trends of blood demand globally and nationally, few studies advise on the approaches to be taken when forecasting a subtly or suddenly changing series.

After taking a look at the data and the nature of the demand signal in Finland, we noticed a marked change in the demand signal's behavior. This observed change and the incommensurability of the results in the existing literature suggested that searching for one single best performing model for tactical or operational level forecasting is inadvisable, and we should instead focus on developing an adaptive forecasting system that would be able to choose the best model from a pool of models automatically. We decided to use mean absolute percentage error for its interpretability in and between all data scales. However, we also explored using an operation critical version of MAPE with the idea that we would prefer demand overestimation in our forecasts if they were ever used on an operational level.

Evaluation of model performance on monthly data revealed that whenever a sales series was large enough in magnitude, ARIMAX models with a one-hot encoded calendar of months as external regressors provided the best fit, indicating that there is real and predictable variability in blood demand between months in a year. As the magnitude of the series decrease, simple moving average models with different period lengths seem to work the best, possibly due to increased noise in the signal. The ARIMAX models offer a significant 50.1% improvement against the old forecasts.

6.2 Challenges, caveats, and future directions

It is necessary to keep in mind that while we were able to significantly reduce forecasting errors in the tactical level forecasts used at the FRCBS, the system can not be used standalone for operational level forecasts. The primary reason for this is that the FRCBS currently incorrectly estimates operational blood demand only by sub-1%, which is even more accurate than what our system achieves on a tactical level. This is partly because the FRCBS can mobilize donors effectively whenever the blood storage levels approach depletion. However, donor mobilization is costly, which might motivate the use of this kind of forecasting system as a complementary tool.

Another issue and target for future research is the size of the rolling window. Our choice of three years is inclined towards the conservative end, and we are probably sacrificing some accuracy because of it. A smaller rolling window might help some models detect patterns that are otherwise muddled by ongoing changes. It might be worth to check in the future, what kind of rolling windows find the best compromise between accuracy and robustness and possibly even identify the best window for each subseries.

On the practical side, we ran into problems when attempting to deploy our script onto the server where it was supposed to be running. First, the server refused to draw any of our plots in the report. We fixed this issue by forgoing the colorblind palette, with which the plots were initially drawn. Second, when the script runs the first time, some history is "simulated" and by running the forecast blind on past observations, and sometimes the server kills the script in the middle of this. The history generation is robust to abrupt kill commands, but continually restarting the script requires some supervision, which is a problem that needs to be addressed at some point.

Also, the current running time of the script is rather long. This is only expected when we select from 13 different models using a rolling window validation for ten different time series, first on a monthly level and then on a weekly level. This should not be a concern if the script can be run automatically and without interruptions once a week or once a month, but it might be worth checking if there are any possibilities for optimization, or if a smaller window would suffice for the weekly forecasts.

In terms of academic interest, the overall forecastability of the blood demand series is still very relevant. This thesis's findings emphasize the need to analyze the blood demand time series in the context of change. An analysis of the factors driving short-term and long-term demand in different countries could probably help blood supply chain operators to better understand the blood demand series, which is so critical to the modern medicine.

Bibliography

H. Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle*, pages 199–213. Springer New York, 1973.

T. Akita, J. Tanaka, M. Ohisa, A. Sugiyama, K. Nishida, S. Inoue, and T. Shirasaka. Predicting future blood supply and demand in Japan with a Markov model: application to the sex- and age-specific probability of blood donation. *Transfusion*, 56(11):2750–2759, 2016. ISSN 1537-2995. doi: 10.1111/trf.13780.

A. Ali, M.-K. Auvinen, and J. Rautonen. Blood donors and blood collection: The aging population poses a global challenge for blood services. *Transfusion*, 50(3):584–588, 2010. ISSN 1537-2995. doi: 10.1111/j.1537-2995.2009.02490.x.

American Red Cross Blood Services. Blood components. `https://www.redcrossblood.org/donate-blood/how-to-donate/types-of-blood-donations/blood-components.html`, 2020.

S. A. Anderson, M. Menis, K. O'Connell, and D. R. Burwen. Blood use by inpatient elderly population in the United States. *Transfusion*, 47(4):582–592, 2007. ISSN 1537-2995. doi: 10.1111/j.1537-2995.2007.01159.x.

S. K. Bar-Lev, O. Boxma, B. Mathijsen, and D. Perry. A Blood Bank Model with Perishable Blood and Demand Impatience. *Stoch. Syst.*, 7(2):237–262, 2017. ISSN 1946-5238. doi: 10.1287/stsy.2017.0001.

J. M. Bates and C. W. Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

C. Beguin, M.-C. Closon, P. Vandekerckhove, and P. Baele. Concentration of transfusion resources on a few pathologies and a few patients: analysis of the comprehensive in-hospital patient database. *Transfusion*, 47(2):217–227, 2007. ISSN 1537-2995. doi: 10.1111/j.1537-2995.2007.01092.x.

R. J. Benjamin and B. I. Whitaker. Boom or bust? Estimating blood demand and supply as the baby boomers age. *Transfusion*, 51(4):670–673, 2011. ISSN 1537-2995. doi: 10.1111/j.1537-2995.2011.03071.x.

B. A. Borkent-Raven, M. P. Janssen, and C. L. V. D. Poel. Demographic changes and predicting blood supply and demand in the Netherlands. *Transfusion*, 50(11):2455–2460, 2010. ISSN 1537-2995. doi: 10.1111/j.1537-2995.2010.02716.x.

C. Chatfield. *Time-series forecasting*. Chapman & Hall, 2000.

C. J. Currie, T. C. Patel, P. McEwan, and S. Dixon. Evaluation of the future supply and demand for blood products in the United Kingdom National Health Service. *Transfusion Medicine*, 14(1):19–24, 2004. ISSN 1365-3148. doi: 10.1111/j.0958-7578.2004.00475.x.

A. M. De Livera, R. J. Hyndman, and R. D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 106(496):1513–1527, 2011.

M. Dillon, F. Oliveira, and B. Abbasi. A two-stage stochastic programming model for inventory management in the blood supply chain. *International Journal of Production Economics*, 187:27–41, May 2017. ISSN 0925-5273. doi: 10.1016/j.ijpe.2017.02.006.

A. Drackley, K. B. Newbold, A. Paez, and N. Heddle. Forecasting Ontario's blood supply and demand. *Transfusion*, 52(2):366–374, 2012. ISSN 1537-2995. doi: 10.1111/j.1537-2995.2011.03280.x.

B. Fanoodi, B. Malmir, and F. F. Jahantigh. Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. *Computers in Biology and Medicine*, 113:103415, Oct. 2019. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2019.103415.

O. S. S. Filho, M. A. Carvalho, W. Cezarino, R. Silva, and G. Salviano. Demand Forecasting for Blood Components Distribution of a Blood Supply Chain. *IFAC Proceedings Volumes*, 46(24):565–571, Sept. 2013. ISSN 1474-6670. doi: 10.3182/20130911-3-BR-3021.00092.

Finnish Red Cross Blood Service. 2019 Annual Report. `https://vuosikertomus.veripalvelu.fi/media/vuosikertomukset/vp_eng_vk2019_netti.pdf`, 2020.

S. M. Fortsch and E. A. Khapalova. Reducing uncertainty in demand for blood. *Operations Research for Health Care*, 9:16–28, June 2016. ISSN 2211-6923. doi: 10.1016/j.orhc.2016.02.002.

G. M. Frankfurter, K. E. Kendall, and C. C. Pegels. Management Control of Blood Through a Short-Term Supply-Demand Forecast System. *Management Science*, 21(4):444–452, 1974. ISSN 0025-1909.

J. Gardner, Everette S. Box-Jenkins Vs Multiple Regression: Some Adventures in Forecasting the Demand for Blood Tests. *Interfaces*, 9(4):49–54, Aug. 1979. ISSN 00922102. doi: 10.1287/inte.9.4.49.

J. E. Hanke, A. G. Reitsch, and D. W. Wichern. *Business forecasting*, volume 9. Prentice Hall New Jersey, 2001.

W. He, D. Goodkind, and P. R. Kowal. *An Aging world: 2015.* U.S. Government Publishing Office, 2016.

C. C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *O.N.R. Memorandum No. 52*, 1957.

R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder. *Forecasting with exponential smoothing: the state space approach.* Springer Science & Business Media, 2008.

R. J. Hyndman and G. Athanasopoulos. Forecasting: principles and practice. `https://otexts.com/fpp2/`, 2018.

R. J. Hyndman and Y. Khandakar. Automatic Time Series Forecasting: The forecast Package for r. *J. Stat. Soft.*, 27(3), 2008. ISSN 1548-7660. doi: 10.18637/jss.v027.i03.

R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, Oct. 2006. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2006.03.001.

I. Kaastra and M. Boyd. Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10:215–236, 1996.

E. Laurén, L. Vikatmaa, E. Kerkelä, T. Kivipuro, T. Krusius, M. Syrjälä, J. Ihalainen, and V. Pettilä. Red blood cell transfusion in southern Finland from 2011 to 2016: a quality audit. *Transfusion Medicine*, 29(1):41–47, 2019. ISSN 1365-3148. doi: 10.1111/tme.12568.

S. Nahmias. *Perishable Inventory Systems.* Springer Science & Business Media, May 2011. ISBN 978-1-4419-7999-5.

R. Palo. *Epidemiology of blood component use in Finland.* PhD thesis, University of Helsinki, 2013.

C. C. Pegels. Exponential forecasting: some new variations. *Management Science*, pages 311–315, 1969.

A. Pereira. Performance of time-series methods in forecasting the demand for red blood cell transfusion. *Transfusion*, 44(5):739–746, 2004. ISSN 1537-2995. doi: 10.1111/j.1537-2995.2004.03363.x.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

J. Rautonen. Redesigning supply chain management together with the hospitals. *Transfusion*, 47(s2):197S–200S, 2007. ISSN 1537-2995. doi: 10.1111/j.1537-2995.2007.01385.x.

J. Roh, S. J. Choi, S. Kim, H. Min, and H. O. Kim. Blood Supply and Demand in Korea: What is in Store for the Future? *Yonsei Med J*, 61(5): 400–405, May 2020. ISSN 0513-5796. doi: 10.3349/ymj.2020.61.5.400.

RStudio Team. *RStudio: Integrated Development Environment for R.* RStudio, PBC., Boston, MA, 2020. URL `http://www.rstudio.com/`.

P. L. Sasongko, M. Rolink, K. v. d. Hurk, M. v. Kraaij, and M. Janssen. Past, present, and future: a qualitative and literature study identifying historical trends, drivers, and transformational factors for the future demand of blood supply in the Netherlands. *Transfusion*, 59(11):3413–3423, 2019. ISSN 1537-2995. doi: 10.1111/trf.15525.

R. W. Schafer. What is a savitzky-golay filter? [lecture notes]. *IEEE Signal Processing Magazine*, 28(4):111–117, 2011.

E. Seifried, H. Klueter, C. Weidmann, T. Staudenmaier, H. Schrezenmeier, R. Henschler, A. Greinacher, and M. M. Mueller. How much blood is needed? *Vox Sanguinis*, 100(1):10–21, 2011. ISSN 1423-0410. doi: 10.1111/j.1423-0410.2010.01446.x.

M. Sihvola. Mikä on muuttunut: Verivalmistehuolto 2000-luvulla. `https://www.veripalvelu.fi/Koulutusmateriaalit/Mik%C3%A4%20on%`

20muuttunut%20-%20verivalmistehuolto%20Suomessa%202000-luvulla_Marja%20Sihvola.pdf, 2016. Presentation.

A. Tanyavutti and U. Tanlamai. Arimax versus holt winter methods: The case of blood demand prediction in thailand. *International Journal of Environmental & Science Education*, 13(6):519–525, 2018.

T. Volken, A. Buser, D. Castelli, S. Fontana, B. M. Frey, I. Rüsges-Wolter, A. Sarraj, J. Sigle, J. Thierbach, T. Weingand, and B. M. Taleghani. Red blood cell use in Switzerland: trends and demographic challenges. *Blood Transfus*, 16(1):73–82, Jan. 2018. ISSN 1723-2007. doi: 10.2450/2016.0079-16.

A. W. Wells, P. J. Mounter, C. E. Chapman, D. Stainsby, and J. P. Wallis. Where does blood go? Prospective observational study of red cell transfusion in north England. *BMJ*, 325(7368):803–804, Oct. 2002. ISSN 0959-8138.

R. D. Wilding, R. S. J. Hick, and J. Chapman. Understanding the dynamics of the blood supply chain: is demand prediction possible? Blackwell Publishing Ltd, Sept. 2006.

P. R. Winters. Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342, 1960.

World Health Organization and International Federation of Red Cross and Red Crescent Societies. The Melbourne Declaration on 100% voluntary non-remunerated donation of blood and blood components. https://www.who.int/worldblooddonorday/Melbourne_Declaration_VNRBD_2009.pdf, 2009.

World Health Organization and International Federation of Red Cross and Red Crescent Societies. Towards 100% voluntary blood donation: a global framework for action. http://www.ncbi.nlm.nih.gov/books/NBK305667/, 2010.

Appendix A

First appendix

```
> sessionInfo()
>>
R version 3.6.3 (2020-02-29)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 18.04.4 LTS

Matrix products: default
BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.7.1
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.7.1

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] DT_0.8            R.utils_2.9.0     R.oo_1.22.0       R.methodsS3_1.7.1
 [5] data.table_1.12.2 numbers_0.7-1     lubridate_1.7.4   plyr_1.8.4
 [9] knitr_1.23        gridExtra_2.3     ggplot2_3.1.1     forecast_8.7

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.1        urca_1.3-0        pillar_1.4.1      compiler_3.6.3
 [5] tseries_0.10-46   tools_3.6.3       xts_0.11-2        digest_0.6.19
 [9] nlme_3.1-140      tibble_2.1.2      gtable_0.3.0      lattice_0.20-38
[13] pkgconfig_2.0.2   rlang_0.3.4       rstudioapi_0.10   curl_3.3
[17] parallel_3.6.3    xfun_0.7          stringr_1.4.0     withr_2.1.2
[21] dplyr_0.8.1       htmlwidgets_1.3   lmtest_0.9-37     grid_3.6.3
[25] nnet_7.3-12       tidyselect_0.2.5  glue_1.3.1        R6_2.4.0
[29] purrr_0.3.2       TTR_0.23-4        magrittr_1.5      htmltools_0.3.6
[33] scales_1.0.0      assertthat_0.2.1  quantmod_0.4-14   timeDate_3043.102
```

```
[37] colorspace_1.4-1  fracdiff_1.4-2    quadprog_1.5-7    stringi_1.4.3
[41] lazyeval_0.2.2    munsell_0.5.0     crayon_1.3.4      zoo_1.8-6
```