Visualisation and statistical modelling techniques for the management of inventory stock levels

Winston L. Sweatman[1]   James McGree[2]

C. Jacobien Carstens[3]   Kylie J. Foster[4]   Shen Liu[5]

Nicholas Tierney[6]   Eloise Tredenick[7]   Ayham Zaitouny[8]

Abstract

This paper describes the investigations conducted in a Mathematics-in-Industry Study Group project from the Australian meeting at Queensland University of Technology in 2015. This concerned the management of stock levels of raw materials used to construct aortic stents. The approaches used included network visualisation, classification and regression trees, and time series modelling. This work will be of general interest to those who are managing stock levels in a highly volatile context. The methods applied show that there is potential

value in taking a statistical approach to understand and make decisions within such volatility. The work provides a basis for developing more advanced statistical approaches for specific inventory problems.

Contents

1 Introduction

Cook Medical was established in 1963, and is the world's largest family owned medical device manufacturing company, with annual sales greater than two billion. One of its primary products is the aortic stent. These are used for endovascular aneurysm repair. Each month Cook Medical manufactures about 500 custom-made stents, 300 customised stents and 900 standard stents. The manufacturing process of each finished product is composed of multiple sub-assemblies which are ultimately composed of a set of raw materials. There are about 1000 different raw materials and hundreds of sub-assemblies involved in manufacturing a single finished product. Unfortunately, managing the stock levels of the raw materials is particularly difficult for a number of reasons. First, as these stents are being placed in individuals, the inherent variability between individuals means that each stent needs to be customised in a variety of different ways. Consequently, the production of every single stent involves a different composition and number of raw materials. Secondly, the demand for stents varies over time, with urgent orders for patients with serious, life-threatening conditions affording further variability in demand for raw materials. Thirdly, each raw material has a certain storage or 'shelf-life' which, when exceeded, means the raw material is discarded. So one cannot simply order excessive amounts of raw materials as this may result in massive losses in expenditure. Thus, the management of inventory levels for Cook Medical is a non-trivial task. Indeed, one must not forget the serious implications of not meeting demand, particularly for urgent orders.

In this study, we investigate some techniques for visualising the complex production process of a number of finished products to try to understand the data at hand, and also to gain insight into potential bottlenecks and/or important raw materials in the production processes. Further, we use data mining techniques to discover what raw materials are causing delays in production as this will provide useful information for stock management. Finally, we develop a time series model to understand the variability in raw materials (for example, how the demand changes over the calendar year)

and then use this model to make predictions about demand. Then, once a prediction is made say for the next month, the current raw material levels can be inspected and new materials are ordered, if required.

2   Data provide by Cook Medical

Cook Medical provided data on the production of stents for the five year period of January 2010 to December 2014. The data included information about which raw materials (and the corresponding quantities of each) are needed in manufacturing each stent. From the data, we are also able to determine the timing and quantity of orders on a monthly basis, and also the time taken to manufacture each stent.

As these data are readily available to Cook Medical, it is of interest to determine what useful information could be extracted to assist in the efficient management of stock levels of raw materials. The proposed visualisation and statistical techniques that could be implemented are formally introduced in the next section. Further work is needed to implement these techniques in an automated and structured manner at Cook Medical. However, through the application of these techniques to the data we have been provided, we demonstrate the usefulness of these methods.

3   Methodology

We now describe the methodology used to address the aims of this research. Corresponding to the number of aims, we split this section into three parts.

Algorithm 1: Code that generates assembly graphs for each product

```
# Load the R-package and data
library('igraph')
source('load_data.R')
source('functions.R')


# Find the indices of all products
idx = which(data_2010_2014$WO. != "")


# Set the plotting layout of igraph
igraph.options(plot.layout=layout.reingold.tilford)


# Create and plot the assembly graph of product no. i
product = data_2010_2014[idx[i]:(idx[i+1]-1),]
g <- create_assembly_graph(product)
my_colors<-get_colors(get.vertex.attribute(g, 'name'))
plot_graph_detail(g, col_v = my_colors)
```

3.1 Visualisation of the manufacturing process

The production process is visualised using networks. Finished products are linked on a tree with their sub-assemblies' subsequent raw materials. By including several finished products on the same diagram it is clear how different products are interdependent through incorporating common sub-assemblies or raw materials.

For the visualisation of the assembly networks we use the 'Reingold–Tilford' layout [11]. For the visualisation of the monthly product networks we use the 'Fruchterman–Reingold' layout [5].

The R-package 'igraph' [4] is used. Algorithm 1 lists the code that generates assembly graphs for each product from the data provided by Cook Medical. Similarly, Algorithm 2 lists the code that generates a monthly network.

Algorithm 2: Code that generates a monthly network

```
# Load the R-package and data
library('igraph')
source('load_data.R')
source('functions.R')

# Create monthly network
g = create_monthly_network("jan", 2014)

# Set the graph layout
igraph.options(plot.layout=layout.fruchterman.reingold)

# Plot monthly product network
my_colors<-get_colors(get.vertex.attribute(g, 'name'))
plot_graph_detail(g, col_v = my_colors)
```

3.2 Classification and regression trees

Classification and regression trees (CARTs) are a widely used data mining technique to predict the response of a variable of interest [1, 6]. Classification trees predict a categorical response variable whereas regression trees are used when the response is continuous. The response is modelled via a tree-like structure based on identified thresholds or levels of important explanatory variables. The depth of the tree, or number of branches, is determined by various goodness of fit measures designed to trade off accuracy of estimation and parsimony. Cross-validation is used to explore how well the model can predict new data [1, 6, 3]. CART models handle missing data by using surrogate splits: when a value for a variable is missing and that variable needs to be used for a split, an alternative variable with a similar splitting property determines the direction of the split [3, 1].

The tree-like structure for the model allows potentially complex interactions

---

Algorithm 3: MATLAB code to produce the forecast

```
# Call the R-package
library(rpart)

# Find the tree
fit <- rpart(long ~ x1+x2+x3,method="class",data=Cook)

printcp(fit) # Show the results
plotcp(fit) # Show results of cross-validation
summary(fit) # Show splits

# plot tree
plot(fit,uniform=TRUE,main="Classification tree for long orders")
```

---

to be found but also facilitates straightforward interpretation of statistical results. This is one of the main reasons why such methods have been widely used in applied research. Other reasons for using CARTs include the ability to handle numerical and categorical data, the requirement for few statistical assumptions to hold, and the ability to perform well for large data sets. The strength of the CART analysis is its simplicity in building a single tree that is readily interpretable. This strength is balanced by a weakness in being less able to predict linear relationships and being sensitive to small variations in data, potentially leading to an oversimplification of the real model [2].

For this study, the response of interest is whether a particular order takes a longer time to manufacture than normal. As this is a binary variable, classification trees are considered for this analysis. The trees were estimated in the R-package called 'Recursive partitioning and regression trees' or 'rpart' by Therneau, Atkinson and Ripley. Algorithm 3 lists example code to generate a simple classification tree.

Figure 1: Monthly usage of two raw materials
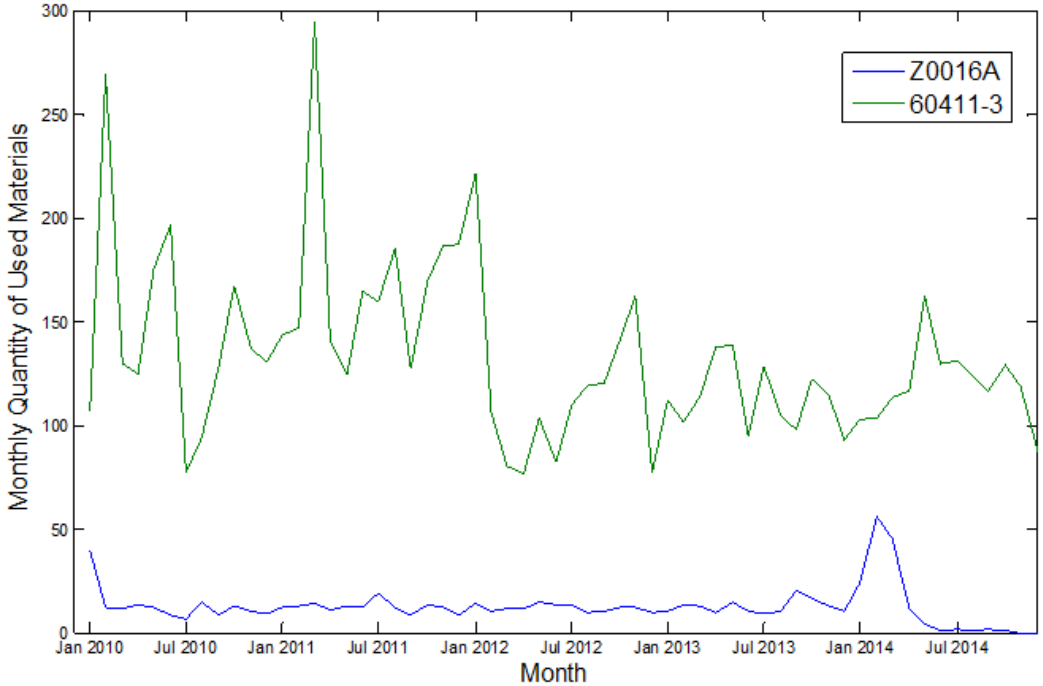
3.3   Time series modelling

It is of critical importance to monitor the usage of raw materials when managing inventory levels for finished products. The usage of each raw material exhibits serial correlation through time which can be accounted for by time series models. For this approach, we demonstrate how time series models were applied to the raw material usage data, and how we obtain forecasts of the future usage.

We consider the monthly usage data of raw materials Z0016A and 60411-3 that were recorded from January 2010 to December 2014. Figure 1 displays the time plots of the observed sample data. We split the sample into training and test sets, with the former covering the period from January 2010 to

December 2013 and the latter covering January to December 2014.

The autoregressive integrated moving average (ARIMA) model is fitted to the training set. Using the backshift notation an ARIMA(p,d,q) model is

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d y_t = \mu + (1 + \theta_1 B + \cdots + \theta_q B^q) e_t$$

where p, d and q stand for the autoregressive order, the differencing order and the moving average order, respectively, and B is the time lag (backshift) operator. The error term e_t is assumed to be independently and identically distributed with zero mean and finite, fixed variance. The p, d and q values are determined by the the Bayesian Information Criterion (BIC).

Once the unknown parameters are estimated, the fitted model is used to produce one-step-ahead forecasts of the usage data, over the twelve month period of the test set. To evaluate the associated uncertainty, standard residual bootstrapping obtains forecast replicates at each forecasting horizon, which are used to approximate the forecast density function conditional on the observed values. Inference is conducted based on the approximated forecast density functions. For instance, the 95% prediction intervals of the raw material usage is constructed by taking the 2.5th and 97.5th quantiles as the lower and upper limits, respectively. Algorithm 4 lists MATLAB code to produce the forecast where:

- `var_ic` is the model selection function;

- `lsm` represents the least square estimator for the selected model;

- `lsmb` denotes the estimator for the backward autoregressive model [7, 8];

- `biascor` is the function that applies bias-correction to parameter estimates [9];

- the `resamp` function carries out bootstrap sampling of residuals;

- the `ysb` function generates pseudo time series using the estimated parameter values, on the basis of the backward autoregressive model [9];

Algorithm 4: MATLAB code to produce the forecast

```
boot=500; boot1=500;
horizon=12; h=3;
h1=1; h2=2; h3=3;
output1 = zeros(boot+2,horizon);
output1(1,:) = x(end-horizon+1:end,1)';
output2 = zeros(boot+2,horizon);
output2(1,:) = x(end-horizon+1:end,2)';
for i = 1:horizon
    xx = x(1:end-horizon-1+i,:);
    [n,k] = size(xx); maxlag = 10; x1 = ones(n,1);
    [aic1, hq1, bic1] = var_ic(xx,maxlag,x1);
    p1 = bic1; [b1,ef1,sigma11,zmat11] = lsm(xx,p1);
    [bb1,eb1] = lsmb(xx,p1);
    [bs1,bsb1,biasf1] = biascor(xx,b1,bb1,ef1,eb1,p1,boot1);
    fbh11=zeros(boot,k); fbh21=zeros(boot,k);
    fbh31=zeros(boot,k);
    index = (1:k)';
    p = p1;
    d1 = zeros(k,1);
    b1 = biasf1(:,1);
    b2 = biasf1(:,2:k*p+1);
    for j = 1:p
        d = b2(:,index)*xx(end-j+1,:)';
        index = index+k;
        d1 = d1+d;
    end
    d1 = d1+b1;
    output1(2,i) = d1(1);
    output2(2,i) = d1(2);
    for j = 1:boot;
        es1 = resamp(eb1,p1);
        xb1 = ysb(xx,bsb1,es1,p1);
        [bss1,efs1,sigus1,zmats1] = lsm(xb1,p1);
        bs11 = adjust(bss1,biasf1,p1);
        fb1 = foreb(xx,bs11,ef1,p1,h);
        fb11 = fb1(h1,:); fb21 =fb1(h2,:); fb31=fb1(h3,:);
        fbh11(j,:) = fb11; fbh21(j,:) =fb21; fbh31(j,:)=fb31;
    end
    output1(3:502,i) = fbh11(:,1);
    output2(3:502,i) = fbh11(:,2);
end
```

- the `adjust` function applies bias-adjustment on the estimated parameter values [8];

- the `foreb` function produces forecast replicates, which are employed to approximate the distribution function at the `h`-step-ahead forecast horizon [9].

3.4   Final comments on methodology

The data provided by the Cook Medical representatives presents an excellent opportunity for applying and developing methodologies for mathematical/ statistical modelling techniques. The main challenge for this study is the complexity of the dataset, and hence suitable methods need to be developed to extract useful information in an efficient way. The study group concentrated on approaches to data visualisation, data mining and time series analysis. These techniques can be extended to achieve better performance. For example, to obtain more accurate forecasts, flexible covariance structures between raw materials are worth investigating. There are opportunities to further continue the research.

4   Results

We now present the results of the three approaches discussed in Section 3.

4.1   Visualisation of the manufacturing process

Cook Medical produces both standard and customised grafts and graft delivery systems. An assembled product consists of both a graft and a graft delivery system. The production varies in complexity: some products are made out of just a few raw materials whereas others are made out of a combination of

raw materials and room stocks. Here, *raw material* refers to any material or part that is bought by Cook Medical. The term raw material is thus not to be taken literally: for example, Cook Medical buys 'stockings' of different sizes, and each different size is classified as a different raw material. *Room stocks* are complex assembled parts that themselves are constructed out of raw materials.

We decided to represent the assembly of a product as a network. A network consists of nodes and connections between nodes. In our network representation, nodes correspond to either the assembled product, a room stock or a raw material. Connections point from either the assembled product or from a room stock, to their constituent parts. We refer to such a network as the assembly network of a product. We visualised assembly networks to get an insight into the diversity of products and the complexity of the assembly process. We discuss assembly networks in more details in Subsection 4.1.1.

After our initial exploration of the product assembly data, we wanted to get an insight into the overall use of materials in a certain time period. To do so we first simplified the assembly networks. We removed the room stock nodes, while retaining all of the raw materials and the quantity needed to assemble the room stock. We then combined all simplified assembly networks of products produced in a single month into one large network. In this network the number of incoming edges of a raw material corresponds to the number of products in which it was used. The quantity of the product that was needed during the month is extracted from the network. We discuss the monthly network in more detail in Section 4.1.2.

4.1.1   Assembly networks

As discussed above, an assembly network consists of nodes corresponding to either an assembled product, a room stock or a raw material. Figure 2 shows two examples of assembly networks. The complexity of these products is vastly different. In each case, the purple 'root' nodes correspond to the

Figure 2: The assembly networks of two custom products produced in September 2012. The purple nodes correspond to the completed product, green nodes to raw materials and blue nodes to room stocks.

Product AC909108 (6/09/2012)   Product AC909151 (26/09/2012)
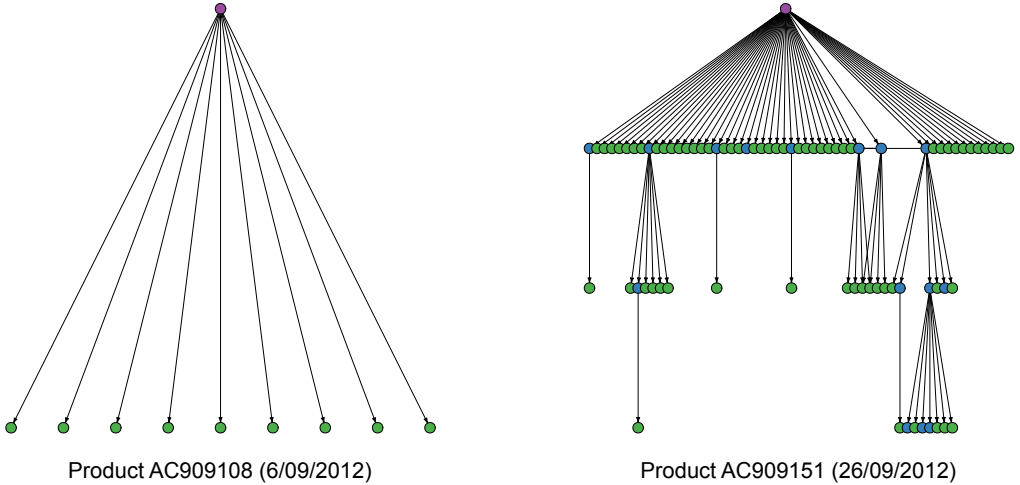
assembled products. There are connections from the assembled product to all raw materials (green) and room stocks (blue) required to produce it. Room stocks in turn are connected to their constituent raw materials and room stocks.

This includes information about which raw materials (and the corresponding quantities of each) were needed in manufacturing each stent. From the data, we were also able to determine the timing and quantity of orders on a monthly basis, and also the time taken to manufacture each stent.

Cook Medical provided data in the form of Excel sheets, that include information on the stents produced and their components. Items listed have identifiers in the form of component numbers and lot numbers. Room stock numbers begin with 'RS'. Assembly networks were constructed from these Excel sheets using code written in the R programming language. A few things stood out while inspecting the assembly networks.

First, many room stocks that were used in the assembly of products, are not in the list of room stock assemblies. In the assembly networks, these room stocks are identified as the blue nodes that have no outgoing edges. By further inspecting the data, we found that overall there are 22384 lot numbers starting with 'RS' that are not in the list of room stock assemblies. These 22348 lot numbers correspond to 237 unique component numbers. For the investigation, these missing room stocks are treated as if they were raw materials.

Secondly, we found that there are several room stocks that only consist of one raw material, see for instance the assembly network of product AC909151 in Figure 2. It seems that treating these 'room stocks' as raw materials would simplify the inventory administration, since at the moment there are essentially two names for a single inventory item. Specifically there is the room stock name and the raw material name, both corresponding to the same raw material.

Finally, there is a lot of variety in the complexity of the assembled products. This really stood out while visualising the assembly networks, as illustrated in Figure 2.

4.1.2   Monthly material use networks

We visualise and analyse the inventory problem by creating a network of all of the finished products produced and all of the raw materials used in January 2014. This process is repeated for each month in the data set.

In order to analyse the large number of materials required to make multiple finished products we simplified their assembly networks. This simplification removes all room stock nodes, while retaining all of the raw materials required to build these room stocks as well as the quantities of raw materials needed. We refer to this process as the *flattening* of the assembly network.

A simple example of the flattening process is illustrated in Figure 3. The

Figure 3: A simple example illustrating the flattening process. On the left is the original network. In the centre we eliminate room stock nodes. Finally, on the right, we eliminate multiple edges.
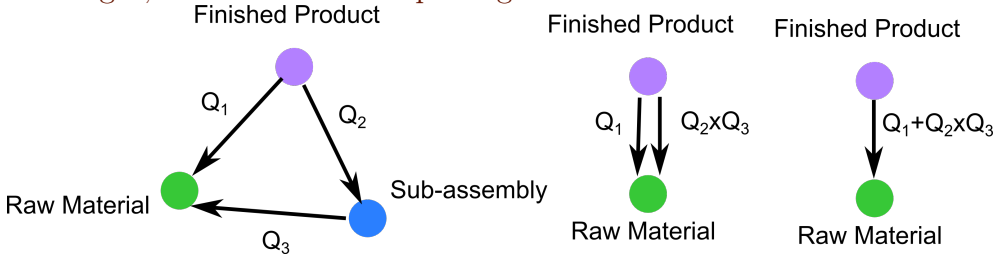
original network is on the left of the figure and is a triangle linking a finished product, a raw material and a sub-assembly (room stock). The finished product is assembled using quantity Q_1 of the raw material and quantity Q_2 of the sub-assembly. The sub-assembly itself is assembled using quantity Q_3 of the raw material. The first stage of the flattening process eliminates the sub-assembly (room stock node) to obtain the central network of the figure. Then we combine the multiple edges. The amount of raw material required to make a finished product depends on the amount of raw material directly used and the amount needed to make any required room stocks. The latter quantity is equal to the sum of amounts of different room stocks required each multiplied by the amount of raw material it requires. Therefore, in the example of Figure 3, the total quantity of the raw material required for the finished product is $Q_1 + Q_2 \times Q_3$. The resulting flattened network is always a star-shaped network. The central node corresponds to the finished product. All edges point from this central node to the raw materials needed to produce the finished product. Edges have an associated weight, corresponding to the quantity of raw material needed.

We combined all flattened assembly networks of products produced in January 2014 into one network. This network contains a node for each product finished in January 2014 and a node for each unique raw material used to make these products. Connections point from finished products to the

Figure 4:   The combined network of ten simplified assembly networks.
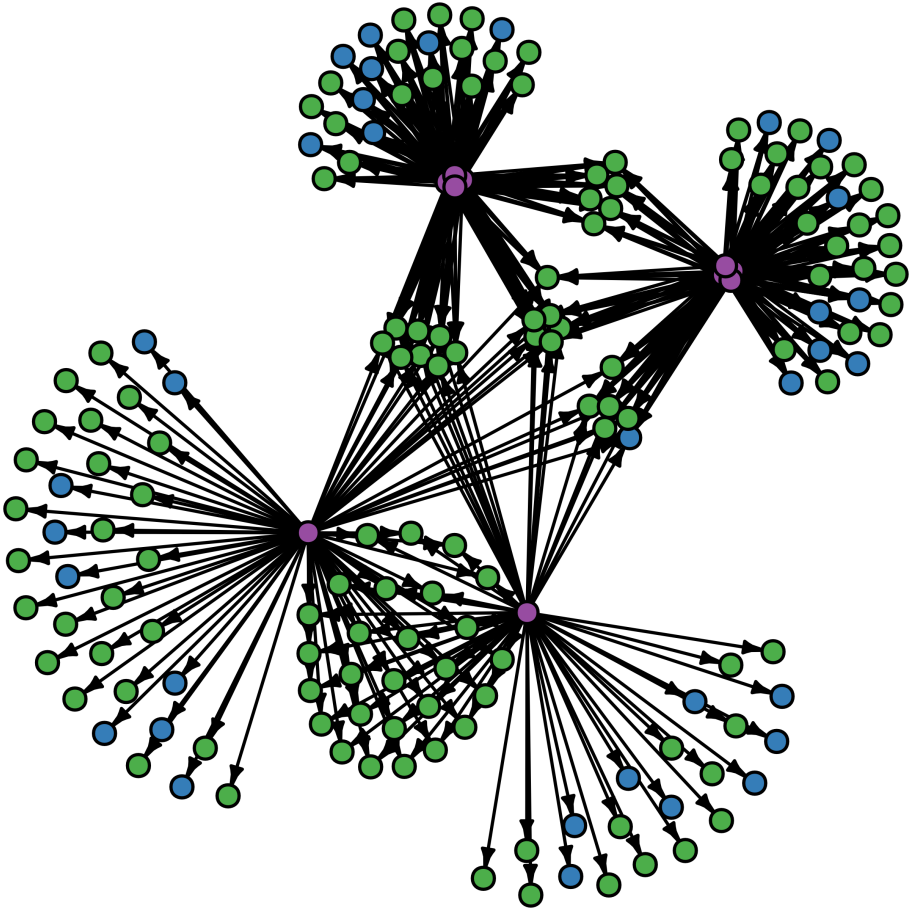
raw materials required to assemble them. To illustrate this idea, Figure 4 shows a network consisting of ten finished products. After inspecting the corresponding assembly networks, it turns out that among these ten finished products, there are two custom devices, and eight standard devices. Out of the eight standard devices, four are product 123999 and four are product 173881. This figure shows that these standard products have many raw materials in common.

Figure 5 shows the whole network for January 2014. Further analysis gains some valuable information.

First, raw materials corresponding to nodes with high in-degree (nodes with a large number of incoming connections) are raw materials that are required for many finished products. On the other hand, raw materials corresponding to nodes with low in-degree are raw materials that are only required by a few finished products.

Secondly, for each raw material we calculate the sum of incoming edge weights (quantities). This sum corresponds to the quantity of the raw material needed in January. Again we identify the raw materials that were used most and least, but this time in terms of quantity needed.

In January 2014, there were a total of 800 finished products. Figure 6 shows the 40 raw materials that were used in half or more than half of these finished products. Even though these materials are critical in the sense that they are required for most products, they may be very easy to acquire and thus not a real bottleneck in the production process. For example, A1729 corresponds to a silicon lubricant and U1180-M to glue. The silicon lubricant was the most used raw material, and was used in all but three of the finished products.

A file was produced that contains the details of raw materials that were only used in the production of a single finished product.

Figure 5:   The combined network of all products made in January 2014.
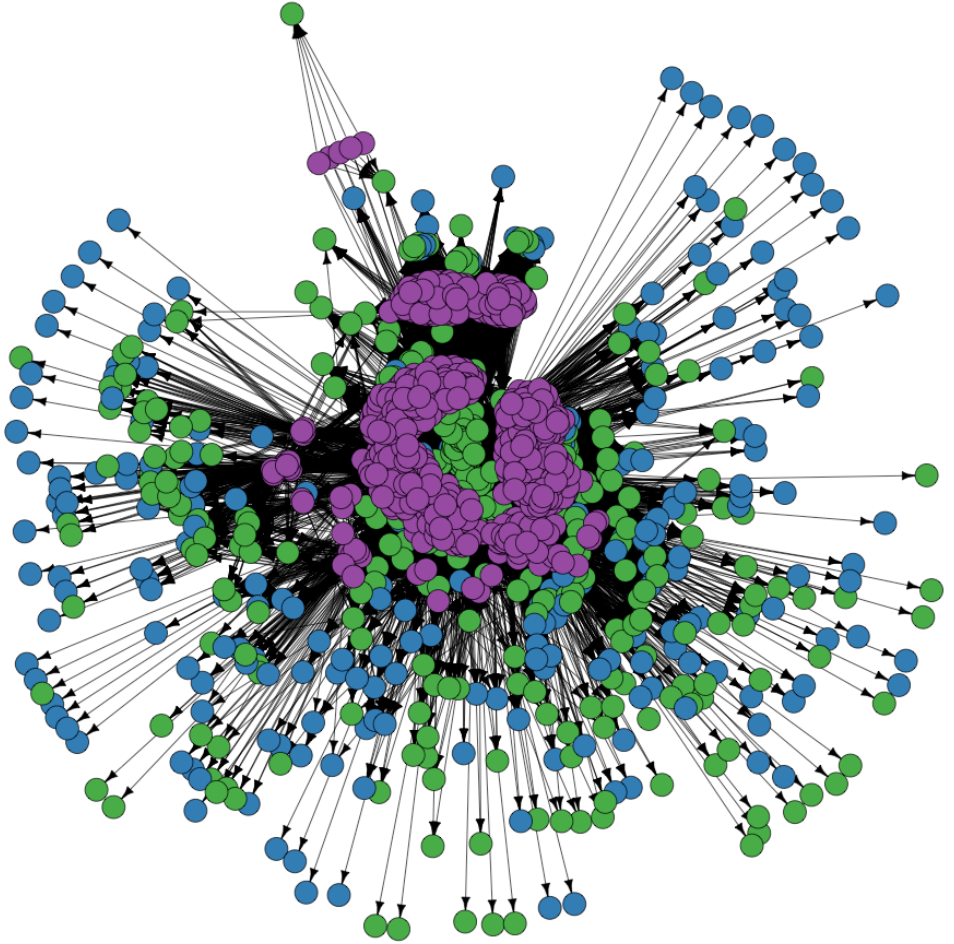
Figure 6:   The 40 raw materials that were required for at least half of all finished products in January 2014.
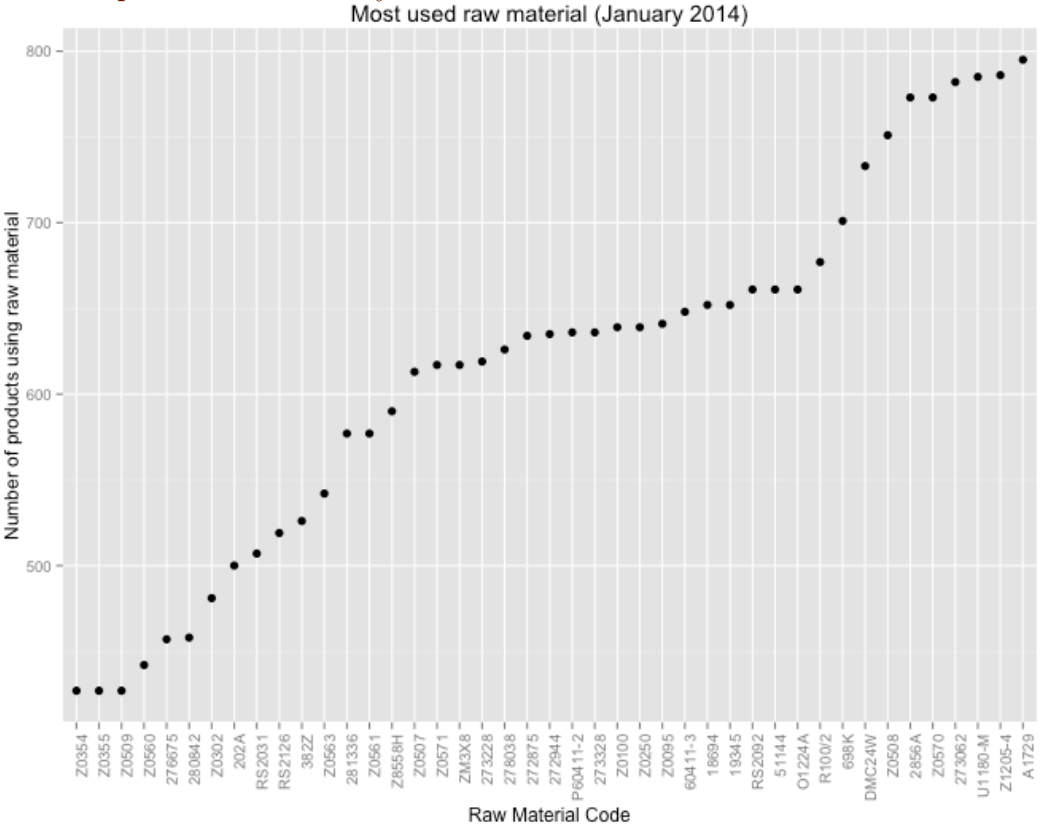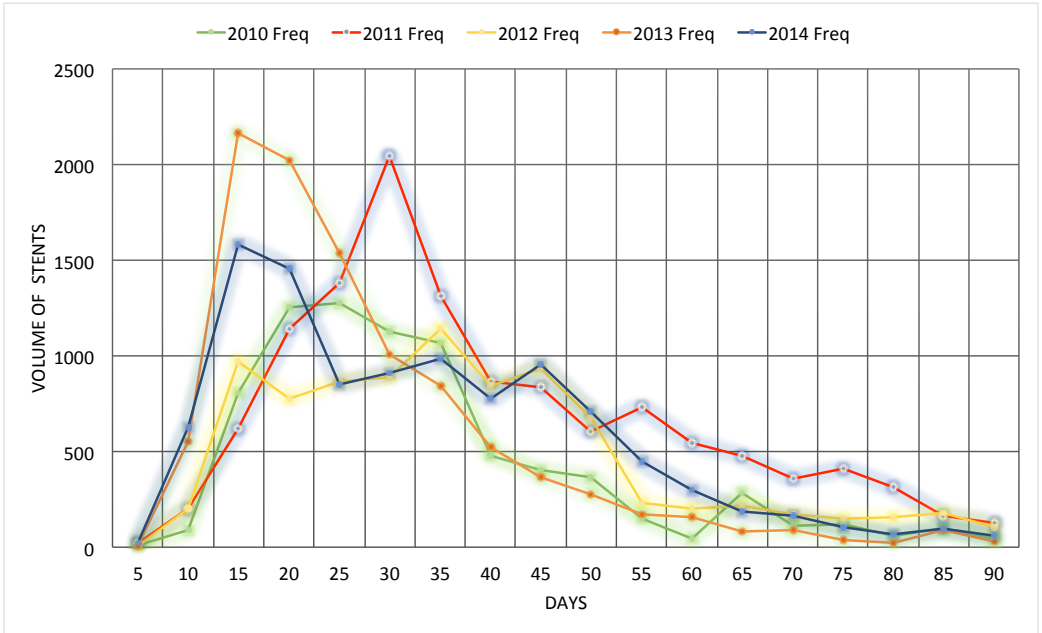
Figure 7: Production time of stents in days.

4.1.3 Production times

Here we compare the shifts in production time in days of all stents created in the last five years by year, showing a maximum of three months in production time. The stents taking longer than three months to produce were not included in the analysis. This is because such stents were typically outliers as they often had much longer production times than three months. Further, such lengthy production times were associated with non-urgent patients, and therefore the immediate availability of raw materials is not so vital.

Figure 7 shows that the total production time in 2011 had a higher volume peaking around the thirty day mark, indicating production overall was slower than in 2013 and 2014, which were the most efficient years as the peaks are seen earlier.

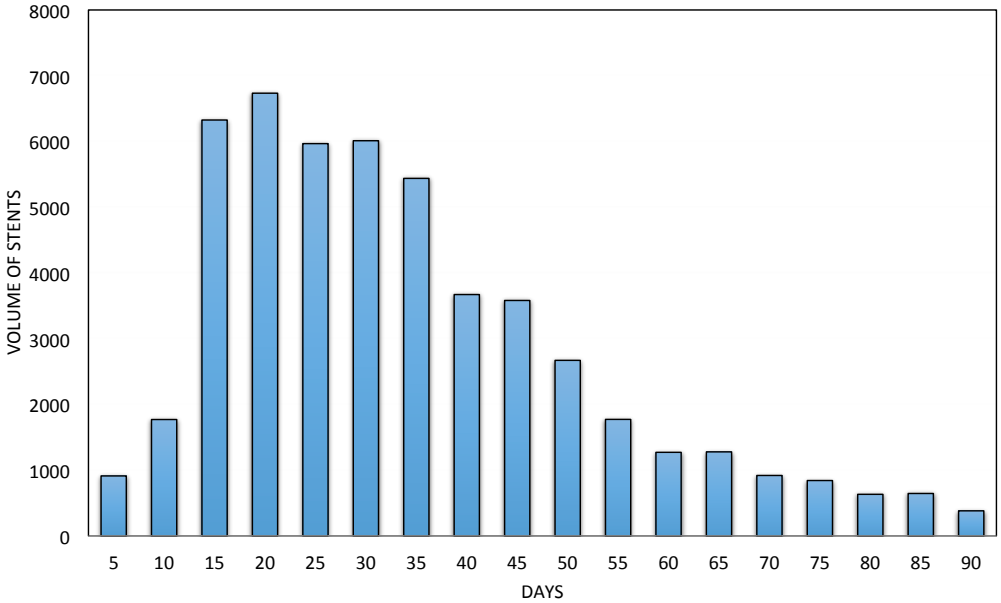Figure 8: Production time in days for all stents.

Figure 8 is production time over combined years with production time of up to three months. This figure is helpful to compare to 2014 (Figure 9) as this graph shows the production times for the last five years have been largely in the 15–55 day range.

Figure 9 is 2014 production time in days showing most stents were produced in under 55 days and a high volume was created in 15–20 days. Comparing this trend to the overall graph of all years in Figure 8 we see the production time in 2014 has improved as the production days are now in the 15–20 day range, instead of being spread over 15–55 days.

Figure 10 and Figure 11 show the mean and median days of production split up into different families or categories of stents. Figure 10 shows that families L-D, L-P, 0 and F have the longest production times in days. The most important family is A, as the stents are produced in the highest volume over

Figure 9: 2014 production days.
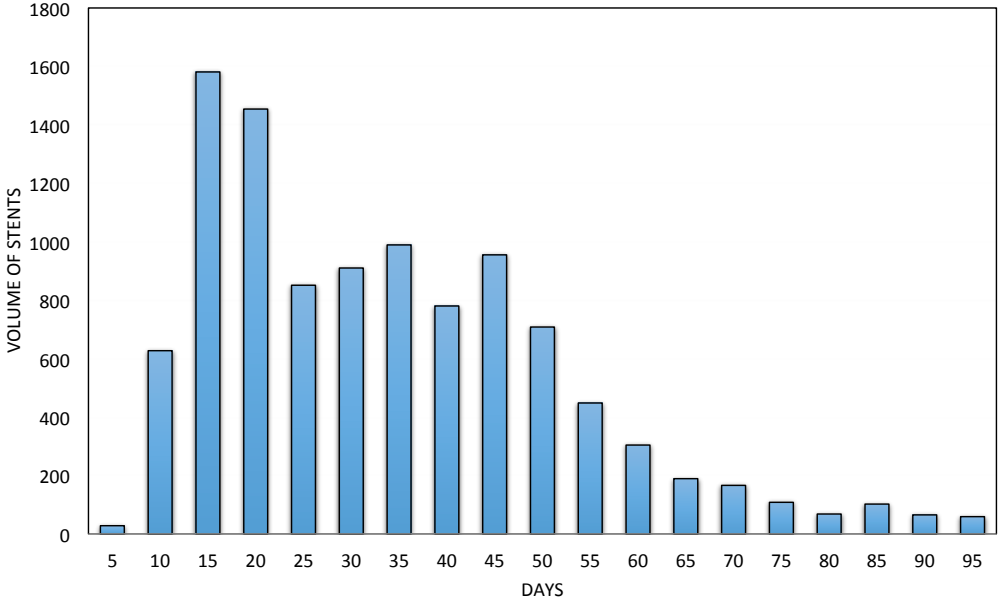
five years (volume of 14,153). Therefore having a small standard deviation is important for family A as shown in Figure 10 and 32 days of production is quite good, indicating that the ideal of a 3 week (21 day) production time is within reach. Also families B, C and K are important having total production volumes of 3395, 3909, and 89, respectively. These three families have low total production days which is ideal. The highest production days on average were from L-D, L-P and O families. However, these families are not important to be considered in the production times as L-D and L-P are only ordered by one physician and family O is a component (not a stent) produced at a very low quantity.

Figure 10: Mean days of production by family including standard deviation.
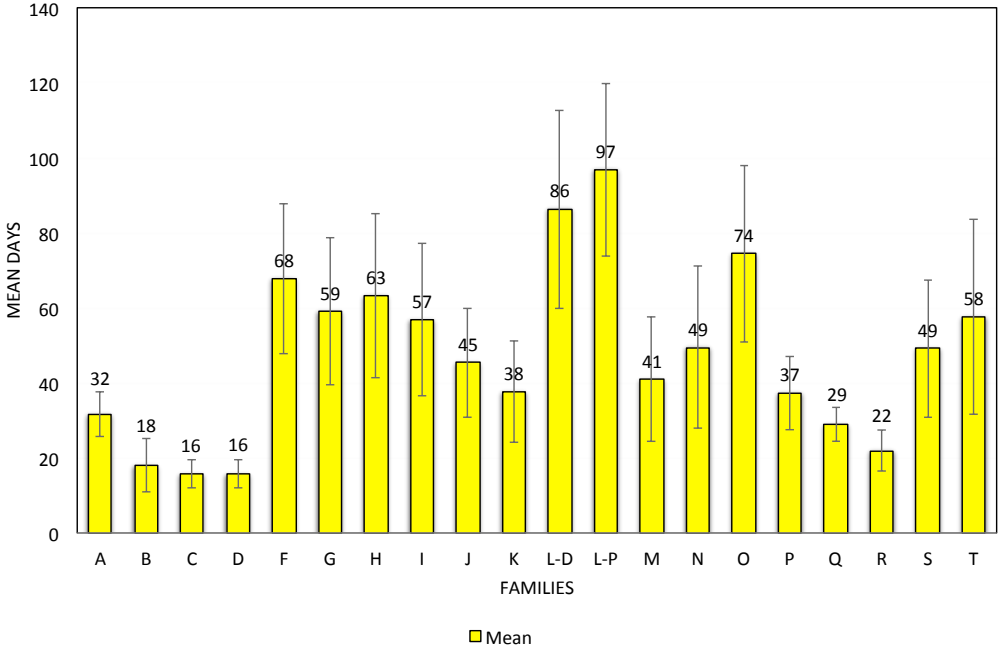
4.2   Classification and regression trees

The data are explored using classification and regression trees (CARTs) in order to determine which variables were associated with problematic orders. If such variables are identified, then one may be able to construct an early-warning system. As stated in Subsection 3.2, the response of interest was a binary indicator for whether the production time for a particular order took longer than expected; so TRUE refers to production times longer than expected, and FALSE refers to production times which are equal to or less than expected. This investigation is complicated by relevant variables in the decision tree being dominated by common trivial sub-assemblies and materials, for example stickers and tip protectors, which may mask more interesting effects. However,

Figure 11: Mean and median days of production by family.
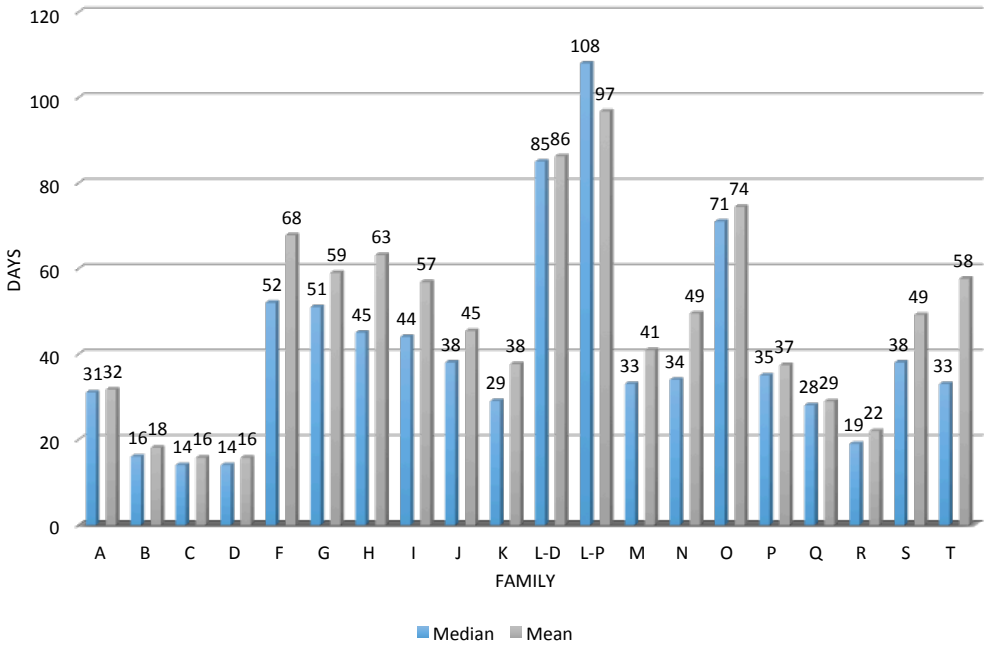
the results should still be useful for managing raw stock levels.

Figure 12 was produced when a CART was fitted to the data supplied by Cook Medical. This CART model is interpreted by following the binary logic of the tree—the root branch asks "Is the level of stock Z0621 less than 0.002?"—if yes, we go left, and if no we go right. The nodes at the bottom indicate whether the predominant response was TRUE or FALSE for that subset identified, with the numerator indicating how many of those responses in that subset were TRUE or FALSE. So when the level of stock Z0621 was less than 0.002, orders did not take longer than expected for 2270 out of 2364 observations in that subset. Following along the tree we use the same logic. All terminal nodes were FALSE bar the final right-most node. We describe a decision that finds a substantive portion of orders taking longer than normal to manufacture—this

Figure 12: Example CARTs for all stents
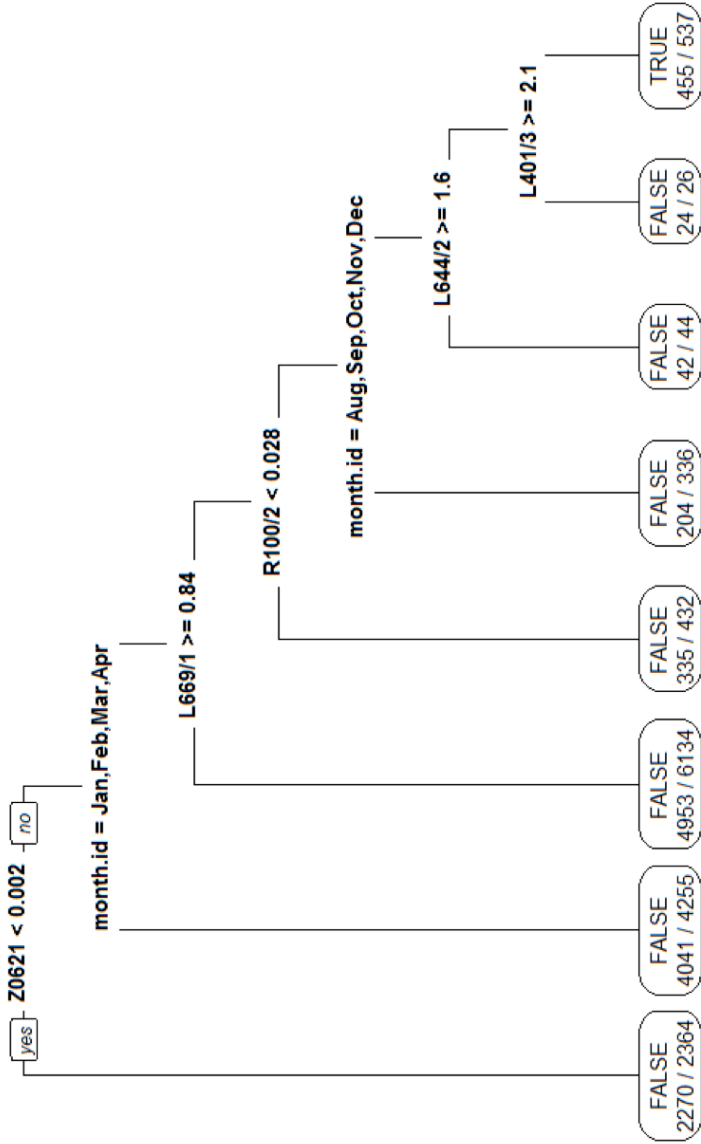
is quite a long decision rule, but it would be

> When Z0621 is less than 0.002, and the month is May, June, or July, and L6691/1 is less than 0.84, and R100/2 is greater than 0.028, and L644/2 is less than 1.6 and L401/3 is less than 2.1, then we find that orders take longer than expected to manufacture.

As the CART identifies those variables important in predicting the outcome it is also worthwhile to investigate further the relationships amongst these variables and the outcome. Figure 12 exhibits example CARTs for all stents.

4.3 Time series modelling

Figure 13 displays the obtained forecast replicates against the observed real value, from July to December 2014. The observed real values are well covered by the simulated forecast replicates, demonstrating desirable performance of the ARIMA model. In addition, most of the approximated forecast densities show a departure from a Gaussian distribution, indicating the appropriateness of the residual bootstrapping procedure. Had this not been considered, the uncertainty quantification would have been based on the normality assumption, and consequently the inference would be unreliable. With the reliability demonstrated, it is possible for forecasts of raw materials usage to be incorporated into the management of inventory levels. For instance, orders of those raw materials may be placed in advance so that they arrive on time.

The bootstrapping-based bias-correction technique was applied to achieve better forecasting accuracy. In practice, it is recommended to carry out bias-correction especially when the size of time series is not fairly large. As pointed out by Liu and Maharaj [8], the benefit from bias-correction tends to be substantial when the length of time series is relatively short, as the bias-corrected coefficients almost always lead to more accurate forecasts than those from non-bias-corrected parameter values. As the length of time series increases, the gain from bias-correction tends to reduce, but better results

Figure 13: Histogram of forecasts versus observed value (red line)
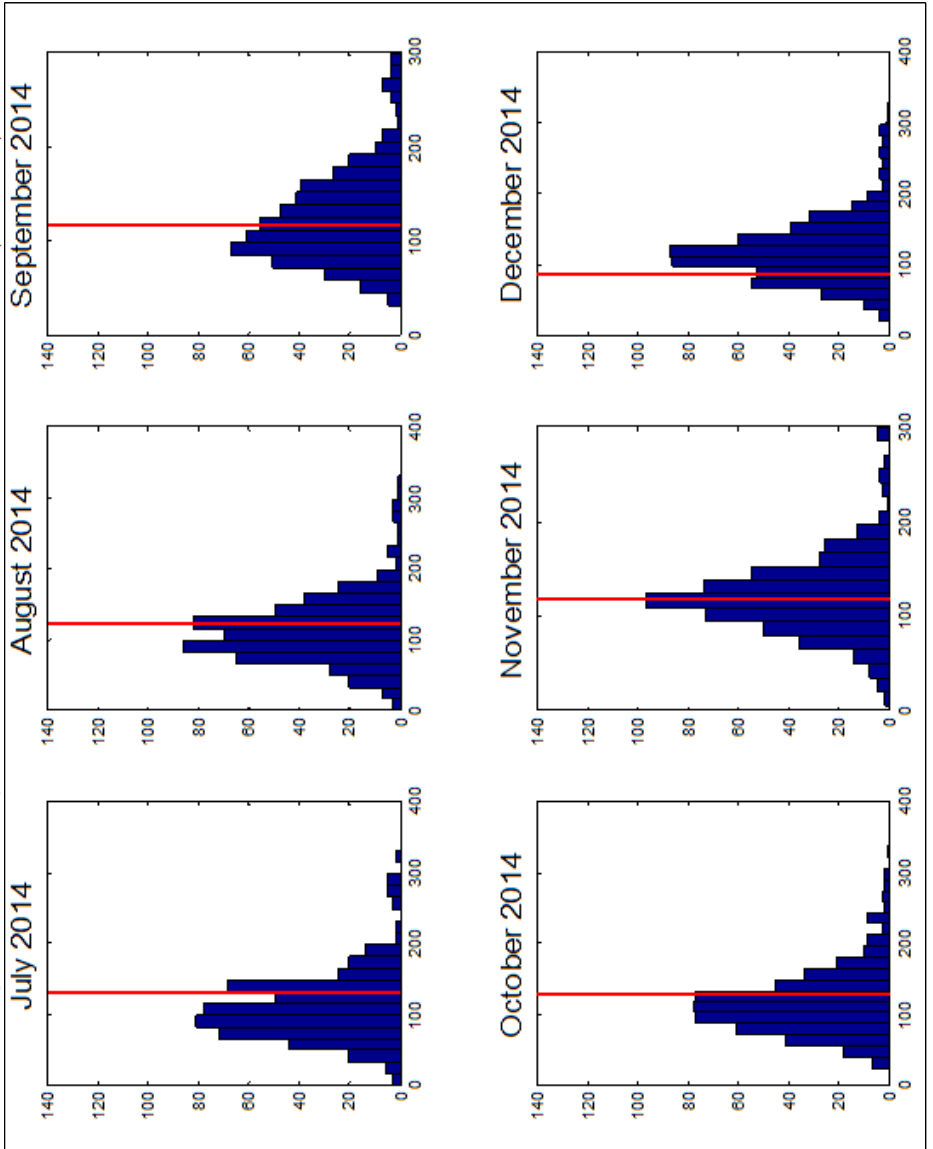
can be still achieved.

We also believe that the forecasting performance can be further improved by considering multivariate time series models. Since the usage of multiple raw materials or sub-assemblies is sometimes highly correlated, the inter-dependency structure may be utilised to enrich the information throughout the forecasting process. Furthermore, as there are numerous raw materials and sub-assemblies involved in manufacturing a single product, time series clustering may be useful to decide on groups of raw materials or sub-assemblies that would have similar usage in future. Such tasks [9, 10] are out of the scope of this paper.

5   Discussion and conclusions

The analysis described in Section 4.1 only shows a small part of the information that can be derived from the network representation of the monthly product network. We now describe some ideas that could be further developed.

1. Standard products with the same product code can be merged into one product, simplifying the monthly product network. If there are small differences between products that share a product code then this can be taken into account by adjusting the edge weights.

2. Raw materials that are of little interest (such as glue or lubricant) can be removed from the network to simplify it.

3. Raw materials can be assigned either a monetary value or a numerical value corresponding to the ease or difficulty of acquiring that raw material. When we multiply this by the quantity needed, this would give us a different indication of which raw materials are truly critical.

4. Comparing the monthly networks for a few consecutive months can give us insights into trends.

5. We can create a one-mode projection of the monthly network to find correlations between raw materials. In the one-mode projection only nodes corresponding to raw materials are present, they are connected by an edge if they occur in the same finished product. Weights indicate how many finished products contain both the materials.

6. Comparing assembly networks may reveal custom products that share similar features.

Figure 14 illustrates the first of these ideas. Four simplified assembly networks are shown on the left of the figure. The normal result of merging these networks is shown in the top right; raw materials are identified and edges from finished products to raw materials maintained. In contrast, on the bottom right, a simplified version of the combined network is shown. The four standard products are merged into one and the edge weights correspond to the total quantities of raw materials required.

The results in Section 4.2 seemed biased towards the frequent but largely unimportant raw materials in the production process. Hence, one could improve the results by only considering raw materials that have a shorter shelf-life and/or longer ordering time. Further, imbalance in the proportion of long orders compared to not so long orders meant that the not so long orders were better predicted than long orders. Thus, the results could potentially be improved by correcting for this. Approaches of interest here are replicating or resampling the under- and over-represented groups, respectively.

The results presented in Section 4.3 appear to be satisfactory, demonstrating the good performance of the ARIMA model. However, it was the monthly usage data that the ARIMA model was applied to, whereas its performance remains unknown if weekly or daily data are considered. In general, forecasting weekly or daily time series data involves modelling short-term time-varying patterns, where high-frequency time series models would be more appropriate. Moreover, if the obtained usage data exhibit non-linearity, then one needs to consider non-linear time series models, such as the logistic smooth transition autoregressive model or the exponential smooth transition autoregressive model.

Figure 14: A illustration of network merging. On the left are four simplified assembly networks. On the top right is the normal result of merging these networks. On the bottom right is a simplified version of the combined network.
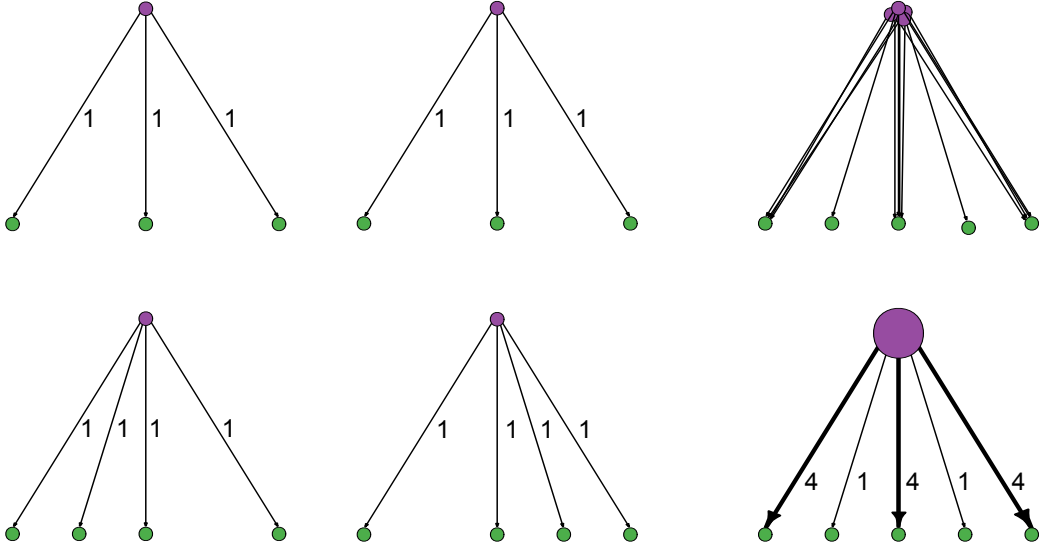
References

[1] Breiman, L. (1996) *Bagging Predictors.* Machine Learning, **24**, 2, 123–140. doi:10.1023/A:1018054314350 M135

[2] Breiman, L. (2001) *Statistical Modeling: The Two Cultures* Statistical Science, **16**, 3, 199–231. https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726 M136

[3] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984) *Classification and Regression Trees* Wadsworth, Belmont, Ca. ISBN-13: 978-0412048418 ISBN-10: 0412048418 M135

[4] Csardi, G., Nepusz, T. (2006) *The igraph software package for complex network research.* InterJournal: Complex Systems, **1695**, 5, 1–9. http://www.interjournal.org/manuscript_abstract.php?361100992 M134

[5] Fruchterman, T. M. J., Reingold, E. M. (1991) *Graph drawing by force-directed placement.* Software: Practice and Experience, **21**, 11, 1129–1164. doi:10.1002/spe.4380211102 M134

[6] Hastie, T., Tibshirani, R., Friedman, J. H. (2009) *The elements of statistical learning : Data mining, inference, and prediction, 2nd edition* New York: Springer Verlag. ISBN 978-0-387-84858-7 (eBook), ISBN 978-0-387-84857-0 (Hardcover) doi:10.1007/978-0-387-84858-7 M135

[7] Kim, J. H., Wong, K., Athanasopoulos, G., Liu, S. (2011) *Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals.* International Journal of Forecasting, **27**, 887–901. doi:10.1016/j.ijforecast.2010.02.014 M138

[8] Liu, S., Maharaj, E. A. (2013). *A hypothesis test using bias-adjusted AR estimators for classifying time series in small samples.* Computational Statistics and Data Analysis, 60, 32–49. doi:10.1016/j.csda.2012.11.014 M138, M140, M155

[9] Liu, S., Maharaj, E. A., Inder, B. (2014) *Polarization of forecast densities: A new approach to time series classification.* Computational

Statistics and Data Analysis, 70, 345–361.
doi:10.1016/j.csda.2013.10.008 M138, M140, M157

[10] Liu, S., McGree, J., Ge, Z., Xie, Y. (2015) *Computational and Statistical Methods for Analysing Big Data with Applications.* Elsevier, London. ISBN: 978-0-12-803732-4 M157

[11] Reingold, E. M., Tilford, J. S. (1981) *Tidier drawings of trees.* IEEE Transactions on Software Engineering, **7**, 2, 223–228.
doi:10.1109/TSE.1981.234519 M134

Author addresses

1. **Winston L. Sweatman**, Centre for Mathematics in Industry, Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand.
mailto:w.sweatman@massey.ac.nz

2. **James McGree**, School of Mathematical Sciences, Queensland University of Technology, Australia.
mailto:james.mcgree@qut.edu.au

3. **C. Jacobien Carstens**, RMIT, Australia.
mailto:corriejacobien.carstens@rmit.edu.au

4. **Kylie J. Foster**, University of South Australia, Australia.
mailto:kylie.foster@mymail.unisa.edu.au

5. **Shen Liu**, Taylor Fry Analytics and Actuarial Consulting, Australia.
mailto:mailto:shen.liu@taylorfry.com.au

6. **Nicholas Tierney**, Queensland University of Technology, Australia.
mailto:nicholas.tierney@gmail.com

7. **Eloise Tredenick**, Queensland University of Technology, Australia.
   mailto:e.tredenick@hdr.qut.edu.au

8. **Ayham Zaitouny**, University of West Australia, Australia.
   mailto:20720394@student.uwa.edu.au