

An exploratory study to identify rogue seasonality in a steel company's supply network using spectral principal component analysis

Nina F. Thornhill⁺ and Mohamed M. Naim*

⁺*Department of Electronic and Electrical Engineering, University College
London, Torrington Place, London WC1E 7JE*

*Corresponding author - *Logistics Systems Dynamics Group, Cardiff
Business School, Aberconway Building, Colum Drive, Cardiff, CF10 3EU*

Published in *European Journal of Operational Research*, 172(1): pp 146-162

An exploratory study to identify rogue seasonality using spectral principal component analysis

Abstract:

Variability in the information flows within a supply network requires production companies to either track the variations, hence leading to increased production on-costs, or to buffer themselves via the use of inventory which leads to stock holding costs.

Customer demands generate variability, often in the form of seasonal patterns, but must be satisfied. In contrast, “rogue seasonality”, i.e. unintended variability, may be generated by a company’s own internal processes such as inventory and production control systems. Importantly, rogue seasonality may propagate through a supply network. Thus there is a motivation for automated detection of network-wide rogue seasonality and for the diagnosis of its root cause.

In this article, a data-driven technique known as spectral principal component analysis is used to detect and characterise cyclical disturbances in a supply network that indicate seasonality. All the information and material flows participating in each disturbance are detected, and the distribution of each disturbance enables a hypothesis to be reached about its root cause. The technique is applied to a supply network consisting of four autonomous business units in the steel industry. Two main cyclical disturbances were detected and diagnosed. One was found to be rogue seasonality and the other was externally induced by the pattern of customer orders.

Key words: Uncertainty, variability, time series, frequency spectrum, multivariate analysis, spectral principal component analysis

1 Introduction

The role of information in the supply network: Recent innovations in information and communication technologies (ICT) have enabled the efficient and timely collection of state data from the supply network. Performance data measurement, the translating of the data into management information and the consequential management decisions based on the information determine the extent of effective supply network operability.

Data in itself is not information. It is only information when it reduces uncertainty in the supply network. An immediate consequence of ICT-enabled data collection is the quantity of data that has to be analysed in order for management to take appropriate actions. This paper tests the ability of a novel technique to

- extract meaningful signatures from time series data that characterise the dynamical behaviour of a complex supply network.
- differentiate between exogenous induced dynamics, such as customer orders, and rogue seasonality induced by production scheduling systems.

Analysis of a supply network: Products may pass through a number of nodes, with associated value add and non-value add activities, from raw materials to finished end product before they reach the final customer. Activities associated with such a system typically include manufacturing and distribution, which have to manage both information flows from, and the flow of material to, the end users. Historically, such a system has been described as a “supply chain” or “value chain” and may be more formally defined as

“a system whose constituent parts include material suppliers, production facilities, distribution services and customers linked together by the feedforward flow of materials and the feedback flow of information” (Stevens, 1989)

There are a number of different units of analysis that may be considered when analysing supply systems (Harland, 1996). At the lowest level we may analyse individual activities that constitute a work process which in themselves aggregate to a business process (Watson, 1994). Further aggregation yields an entire business unit, dyadic relationship, supply chain and ultimately a whole supply network as indicated in Figure 1 (based on Harland, 1996). The technique in this paper addresses the latter and enables the monitoring of operations in complex supply networks.

*** Figure 1 near here ***

Supply network performance: The long term goal of any system is to exhibit the characteristics of a homeostat (Ashby, 1966) and in a similar fashion a supply network aims to achieve stability and to reach equilibrium with its environment. At the most fundamental level a supply network’s emergent property (that is, the property of the system as a whole and not the properties of its individual parts) is aimed at matching supply with demand. That is, it should maximise value in terms of satisfying final customer needs in terms of quality, delivery time, availability and total costs.

A major inhibitor to supply network value delivery is uncertainty (Davis, 1993, Mason-Jones and Towill, 1998). Uncertainty in the supply network, which may be described as variance and/or perturbations, in the information and material flows, yields increased total logistics costs. Production companies in the supply network may either aim to track the variations, hence leading to increased on-costs, or buffer themselves against such variations via the use of

inventory, hence risking stock holding and obsolescence costs. Importantly, uncertainty may propagate through a supply network (Davis, 1993).

Davis (1993) has categorised uncertainty into three principal areas. There is the uncertainty associated with a company's own manufacturing processes such as equipment breakdown, operator absenteeism or material yield losses. There is then supplier uncertainty that includes late deliveries, incomplete orders or poor quality products. Finally, there is the uncertainty due to the customers which may materialise in the form of changing customer schedules or a change in delivery specifications.

Naim et al (2002) have developed and applied a systematic methodology for measuring uncertainty and developed heuristics for managers to eliminate or reduce that uncertainty. Ideally, uncertainty should be designed out of the supply network by means of analytical studies that supplement rather than merely replace experience. Considerable research has been undertaken on supply network design through the media of mathematical (e.g. Disney & Towill, 2002, Dejonckheere et al., 2003), statistical (e.g. Chen et al., 2000) and computer simulation (e.g. Berry et al., 1998) studies.

Supply network design should be seen as part of an overall business process re-engineering methodology. An example of such a methodology is that by Parnaby (1991) who proposes a five-stage approach. This paper is interested in the fifth stage.

- Market Analysis – that determines the market qualifier and order winning criteria by which the performance of the supply network system in delivering value to the customers is judged
- Business Process Analysis – including process mapping and input-output analysis
- Steady State Design – including ensuring that the system is balanced and synchronised
- Dynamic Design – which may utilise mathematical and computer simulation models

- Information & Control Design & System Integration – requires the development of data collection systems and appropriate analysis techniques in order to make effective decisions on system control.

Of particular interest in this paper is the ability to detect a specific type of uncertainty known as “rogue seasonality”; that is, cyclic disturbances with periods of a few months that should not be there. These are seasonal demand patterns that are induced by the internal processes themselves and not by any external influences. Forrester (1961) highlighted internal decision structures, such as inventory control policies, as the primary cause of such “rogue seasonality”. More recently the term “bullwhip” has been termed to describe this phenomenon (Lee et al, 1997).

Layout of the paper: The next section of the paper introduces a case study of a supply network in the steel industry. It motivates the need for analysis based upon the power spectra of the variables in the data set. Section 3 introduces the methods of spectral principal component analysis, and section 4 tests its utility for the detection of network-wide cyclical disturbances. The paper ends with a critique of the technique and suggestion for further research.

2. Case study

The steel sector application: The data collection for the application of the spectral principal component analysis was undertaken as part of an action based research project involving a number of industry sectors. The research was aimed at understanding and reducing uncertainty in supply networks and involved a number of data collection techniques as typified by Naim et al. (2002), which may be summarised as opinion, empirical, archival and analytical. This paper is particularly focused at the analytical research data that was collected and which is used to test the applicability of the technique described.

The particular supply network is in the steel sector and is shown in Figure 2. The supply network consists of four autonomous business units although they are part of the same holding group. The steel works manufactures a wide range of products that are used as raw materials by the three mills. The mills then produce an even wider range of products which are sold to customers who include stock holders as well as end users.

*** Figure 2 near here ***

The analytical data used in this paper was collected from the companies' management information systems. The data was made available in spreadsheet form and was checked for consistency using both analytical techniques and opinion based methods. The data includes such variables as flow rates of customer orders, production output, despatch and receipts, and inventory levels such as finished goods, raw materials and order books. The time series are available monthly and cover a period of six years.

Network description: The variables to be analysed are listed in Table 1 and are referred to as "tags" in the following discussions. Their meaning should mostly be self-evident from the descriptions in Table 1. The order book (tags 29-31) are the orders accepted by a company but that are yet to be delivered to a customer. Occasionally, when product is available from stock they will be despatched immediately.

The data available encompass each of the sources of uncertainty defined by Davis (1993) covering a number of nodes within a given supply network. In capital-intensive companies with particular technological constraints in their manufacturing process, such as the steel sector as well as others such as the chemical industries, level asset utilisation is the norm and varying production schedules are to be avoided. Such companies then face a dichotomy as increasing inventory holding to buffer against uncertainty simply exposes them to obsolescence risks.

Case study data: The available data comprised of monthly time series data for 72 months. Figure 3 shows the time series and the power spectra of mean-centred trends. Mean centring removes the average value of the trend that would otherwise appear in the zero frequency channel, thus allowing the analysis to focus upon the deviations from the mean. The horizontal axis for the spectra is a normalised frequency axis such that 1.0 represents the sampling frequency. Since these were monthly samples a spectral peak at, say, 0.25 on the frequency axis corresponds to a cycle in the time series with a period of $1/0.25$, or 4 months. The spectra stop at 0.5 on the frequency axis because the Nyquist sampling theorem requires a sinusoidal signal to be sampled at least twice per cycle (Chatfield, 1996).

The following points can be observed:

- The variability in some of the time series such as tags 16, 32 and 33 appears visually to be random, while others such as 11, 12 and 13 have a more structured repeating pattern. Peaks in the spectra show that in many cases the variability indeed has a structure;
- There are prominent spectral peaks at about 0.17, 0.25 and 0.34 on the frequency axis corresponding with oscillatory features in some time series with period of about 6, 4 and 3 months. These are the features of interest in this paper because they correspond to the seasonality whose origin is to be determined;
- Some tags show low frequency spectral features at and below 0.08 on the frequency axis. These correspond to features with periods of more than one year. These trends correspond to long term effects such as gradual changes in demand because of long term economic cycle and will not be considered further in this paper because the focus here is on seasonality. However, it is noted that the mathematical processing steps will need to distinguish these long term cycles from the shorter cycles of interest;

- The data sets were not all complete. Missing data appear as zero values e.g. in tags 38-43. The main effect of a series of zeros is to distort the low frequency part of the spectrum.

The challenge is to automatically detect the presence of the cyclical disturbance in the tags where it can be seen visually, to discover other tags where it is also present and to characterise the spectral features of the disturbance. The cause of the disturbance should be determined following that analysis. The technique tested in this article for those purposes is spectral principal component analysis.

*** Figure 3 near here ***

3. Spectral Principal Component Analysis

Introduction: A multivariable dynamic system such as a supply network generally has more variables than distinct modes of behaviour. Therefore some of the variables have behaviour in common and may be correlated. Principal component analysis (PCA) finds the underlying modes of behaviour by means of a singular value decomposition of the data matrix. Descriptions of principal component analysis may be found from many sources, for example in Chatfield and Collins, (1980), Wold *et.al.*, (1987).

Near infrared spectra (NIR) and infrared spectra are routinely analysed by PCA (i.e.spectral PCA) for estimation of analyte concentrations in unknown samples such as trace amount of contaminants in a bioprocess stream (Karstang and Henrikson, 1992; Riley *et.al.*, 1997; Yeung *et.al.*, 1999). Seasholtz (1999) described the business impact of spectral PCA using NIR and other spectra at Dow Chemical. In those cases the instruments themselves created the spectra, but the spectra may also be calculated numerically as is done in this paper.

Spectral PCA performs a PCA analysis on the power spectra rather than on the time series and has been successfully applied in other industry applications. Spectral PCA was described by Belchamber and Collins (1993) for the classification of acoustic signals. Their patent gave examples of the classification of the acoustic spectra in 24 frequency channels from a pump and an industrial blender. Wu *et.al.* (1999) used a method known as “eigenfaces” in the recognition of sounds from car engines. Spectral PCA of dynamic data in a chemical plant was described by Thornhill *et. al.* (2002) where plant-wide cyclical disturbances were detected. The findings were presented in a compact form to the plant control engineers who were then able to determine the root cause. Spectral PCA also finds the underlying modes of behaviour in a data set, but characterises the modes by means of the power spectra rather than by their time-domain patterns. SPCA has an advantage over PCA analysis in the time domain because the power spectrum is invariant to time delays or phase shifts caused by network dynamics. For example, the production time series might be varying in a cyclical pattern with a period T_p such as $\cos(2\pi t/T_p)$, while the despatches have the same cyclic pattern but varying as $\sin(2\pi t/T_p)$ because there is a time delay in the system of one quarter of a cycle. Since sin and cos are orthogonal functions, analysis of the time series analysis would detect zero correlation whereas these oscillations are readily identified as the same in a spectral analysis because they both have a spectral peak at the same frequency.

Several other benefits have been identified in Thornhill *et. al.*, (2002), in particular that the effects of missing values are less serious than in time domain analysis because the effects are typically spread thinly across a range of frequencies.

Spectral PCA – formulation: The techniques presented in Thornhill *et. al.* (2002) are briefly reviewed here. The rows of the data matrix, \mathbf{X} , are the single-sided power spectra $P(f)$ of the

signals representing the signal power present in the frequency channels shown on the horizontal axis of the right hand plot in Figure 3.

$$\begin{array}{c}
 n \text{ frequency channels} \quad \rightarrow \\
 \mathbf{X} = \left(\begin{array}{ccc} P_1(f_1) & \dots & P_1(f_n) \\ \dots & \dots & \dots \\ P_m(f_1) & \dots & P_m(f_n) \end{array} \right) \begin{array}{l} m \text{ supply network} \\ \text{variables} \end{array} \quad \dots 1 \\
 \downarrow
 \end{array}$$

The PCA decomposition reconstructs the \mathbf{X} matrix as a sum over m orthonormal basis functions \mathbf{w}'_1 to \mathbf{w}'_m which are spectrum-like functions having n frequency channels arranged as a row vector:

$$\mathbf{X} = \begin{pmatrix} t_{1,1} \\ \dots \\ t_{m,1} \end{pmatrix} \mathbf{w}'_1 + \begin{pmatrix} t_{1,2} \\ \dots \\ t_{m,2} \end{pmatrix} \mathbf{w}'_2 + \dots + \begin{pmatrix} t_{1,m} \\ \dots \\ t_{m,m} \end{pmatrix} \mathbf{w}'_m \quad \dots 2$$

The above expression may be written as $\mathbf{X} = \mathbf{T}\mathbf{W}'$ where the i 'th column of \mathbf{T} is $(t_{1,i} \dots t_{m,i})'$ and the rows of \mathbf{W}' are \mathbf{w}'_1 to \mathbf{w}'_m . The orthonormality of the rows of \mathbf{W}' means that $\mathbf{T} = \mathbf{X}\mathbf{W}$. Singular value decomposition $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}'$ provides a means for computation of the requisite vectors with $\mathbf{T} = \mathbf{U}\mathbf{D}$ and $\mathbf{W}' = \mathbf{V}'$. Matrix \mathbf{D} is diagonal and its elements are the positive square roots of the eigenvalues of $\mathbf{X}'\mathbf{X}$.

A description of the majority of the variation in \mathbf{X} can often be achieved by truncating the PCA description. The following is a three-principal component (PC) model in which the variation of \mathbf{X} that is not captured by the first three principal components appears in an error matrix \mathbf{E} :

$$\mathbf{X} = \begin{pmatrix} t_{1,1} \\ \dots \\ t_{m,1} \end{pmatrix} \mathbf{w}'_1 + \begin{pmatrix} t_{1,2} \\ \dots \\ t_{m,2} \end{pmatrix} \mathbf{w}'_2 + \begin{pmatrix} t_{1,3} \\ \dots \\ t_{m,3} \end{pmatrix} \mathbf{w}'_3 + \mathbf{E} \quad \dots 3$$

The issue of the correct number of principal components is discussed by Chatfield and Collins (1980), Valle *et.al.*, (1999) and elsewhere. The supply network case study needed three PCs, the decision to truncate being made when the eigenvalue associated with the next principal component represented 5% or less of the sum of all the eigenvalues.

Each spectrum in \mathbf{X} may be represented graphically. For instance, when three \mathbf{w}' – vectors are in use the i 'th spectrum maps to a point having the co-ordinates $t_{i,1}$, $t_{i,2}$ and $t_{i,3}$ in a three-dimensional space called a scores plot. Similar spectra have similar t -coordinates. Therefore such groups form clusters. When two \mathbf{w}' – vectors are used, the scores plot is two dimensional.

The rows of the error matrix, \mathbf{E} , are the unmodelled part of the spectra in \mathbf{X} . Therefore the magnitudes of the rows of \mathbf{E} should be inspected. The following vector of squared prediction errors (SPE) can be used in which each term is the sum of squares of the elements of a row of \mathbf{E} .

$$\mathbf{SPE} = \begin{pmatrix} SPE_1 \\ SPE_2 \\ \dots \\ SPE_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N (E_{1,i})^2 \\ \sum_{i=1}^N (E_{2,i})^2 \\ \dots \\ \sum_{i=1}^N (E_{m,1})^2 \end{pmatrix} \quad \dots 4$$

A common reason why one particular spectrum may have a large SPE_i value (i.e. it is not approximated well by a reduced number of principal components) is that it has spectral features at frequencies where no other spectrum shows any features.

Data pre-processing and filtering of spectra: It is usual in time-domain PCA to mean centre the data sets and to scale to unit variance. In spectral PCA the following data pre-processing may be applied:

- Mean centre the time series and remove linear trends before calculation of the spectra;

- Filter the spectra, if required;
- Scale the spectra to the same total power such that:

$$\sum_{i=1}^N P(f_i) = \text{constant}$$

where $P(f_i)$ is the spectral power in the i 'th channel;

The first of these steps removes steady offsets or linear ramps from the time domain data which would otherwise appear in the $f = 0$ channel of the spectrum. The spectrum may be filtered by replacing the values in selected frequency channels with zeros in order to highlight subtle effects, for instance when a dominant cycle masks the presence of a small-amplitude cycle at a different frequency. It can also remove non-stationary low frequency trends such as those caused in a supply network by the economic cycle. The scaling step is equivalent to scaling to constant variance in the time domain because the signal variance is proportional to total power.

Practical computations: Once the spectra have been constructed with the above pre-processing steps then spectral PCA can proceed using an industry standard package such as PLS_Toolbox 3.0 (Eigenvector Research, Manson, WA, USA), SIMCA-P (Umetrics AB, Umeå, Sweden), or Unscrambler (CAMO ASA, Oslo, Norway). The option to mean-centre the \mathbf{X} matrix should be disabled because in spectral PCA the mean centring is applied to the time series data before the Fourier transformation.

Alternatively the codes can be developed from first principles in a general purpose package such as MATLAB (The Mathworks, Natick, USA). That approach was taken here because the integrated and flexible library and graphing functions of MATLAB allowed full insights into the performance of the spectral PCA.

4. Results and discussion

4.1 Spectral PCA plots

A three PC model: A three PC model captured 85% of the variability in the spectra of the supply network data. The fourth PC offered only an additional 5%, the fifth added 3%. Calculation of the *SPE* for the three PC model showed that no tag had an *SPE* of more than 1.5% and only three were above 1%. Therefore a three PC model gives an accurate approximation to the spectra of the 43 tags.

Score plots: Figure 4 shows a three-dimensional score plot whose purpose is to visualise clusters of tags having similar spectra. Each spectrum in the data set (i.e. each row of the \mathbf{X} matrix) is represented by one spot. For the i 'th spot, the coordinates on the three axes of the score plot are the weightings $t_{i,1}$, $t_{i,2}$ and $t_{i,3}$ from equation 3 which when applied to the \mathbf{w}' -vectors approximately reconstruct the i 'th spectrum in a three PC model. For instance, the spectra in Figure 3 for tags 24 and 26 are very similar. In PCA analysis this means that the numerical values of $t_{24,1}$ and $t_{26,1}$ in equation 3 are similar. Likewise $t_{24,2} \approx t_{26,2}$ and $t_{24,3} \approx t_{26,3}$. The spots for tags 24 and 26 therefore map to the same corner of the score plot, as can be seen in Figure 4. Several distinct clusters may be observed and are highlighted in Figure 4 by different symbols. The same symbols have also been placed beside the relevant tag in Table 1. Figure 5 shows three two-dimensional projections of the information from the score plot using the same symbols for the tags. The plots show that the identified clusters always appear together.

A time domain principal component analysis using mean centred time series data scaled to unit variance accounted for only 56% of the data variability using three PCs. Therefore spectral PCA had superior performance to time domain PCA for the analysis of historical data from the supply network, reflecting similar findings for other applications (Thornhill *et. al.*, 2002). Some distinct clusters were observed in the time-domain PCA score plot but those clusters captured irrelevant

features. For instance, tags 38-42 formed a cluster because they each have a similar episode of missing data. Missing values is typical of industry data even though new ICT applications mitigates the risk somewhat.

*** Figure 4 and 5 near here ***

4.2 Characterisation of PCA clusters

This subsection describes the features captured by the clusters which emerged during spectral principal component analysis of the full spectra.

Tags with open triangle symbols: Tags 1, 2, 4, 6, 10, 12, 15, 16, 18, 24, and 26 from the bar and section mills form a cluster. They have two or more of the distinctive spectral features noted earlier at 0.17, 0.25 and 0.34 on the frequency axis corresponding with cyclical features in the time series with periods of about 6, 4 and 3 months. Therefore one objective of the analysis has been met because tags with cyclical disturbances have been detected. The origin of the cycles will be discussed in section 4.3.

Tags with open circle symbols: Tags 14, 28, 35, 38-40, 42 and 43 are mainly stock tags and are characterised by very low frequency features in the spectrum. Therefore spectral PCA has identified the tags with long-term non-stationary trends which reflect long-term changes in stocking strategies.

Tags with open square symbols: Tags 29, 30 and 31 are the order book variables and appear as a cluster because they have a distinct peak at 0.041 on the frequency axis (a period of about 24 months). This indicates long term shifts in customer demand due to economic factors external to the supply network.

Tags with black diamond symbols: Tags 3, 5, 7-9, 11, 13, 17, 19-23, 25, 27, 32-34, 36, 37 and 41 are characterised by multiple features over the whole frequency range and thus are random in the

time domain. There is a danger, however, that the low frequency behaviour is dominating the interesting and relevant behaviour in the high frequency range above 0.1 on the frequency axis. That is to say, there may be cyclical disturbances present in these tags that are hidden by long term deviations. These are discussed in greater detail in the next section.

4.3 High-frequency analysis

Analysis of the high-frequency spectrum: An additional analysis was conducted without the low frequency features. The filtering procedure mentioned in the data preparation sub-section was used to remove spectral features with periods of eight months or longer (up to 0.125 on the frequency axis) in order to focus on cyclical disturbances with periods of 6, 4 and 3 months (0.17, 0.25 and 0.34 on the frequency axis). Figure 6 shows the filtered spectra. They have been rescaled to a unit maximum amplitude so that details can be more easily viewed.

Model and score plots: A three PC model captured 80% of the spectral variability in the filtered spectra and Figure 7 shows the score plot with clusters of tags highlighted. The clusters represented by the large symbols show one or more of the spectral features associated with the cyclical disturbances. For instance, at one extreme the spectra of tags 21 and 23 contain a strong component at 0.25 on the frequency axis (a 4 month cycle), but no other components. At the other extreme tags 24 and 26 have distinct peaks at 0.17 and 0.34 on the frequency axis (6 month and 3 month cycles) but no evidence of a 4 month cycle. The tags with the triangular symbols are dominated by 3 and 6 month cycles, those with the circle symbols are dominated by 4 month cycles and those with square symbols have 3, 4 and 6 month cycles.

Tag 20 shows an example of the benefit of analysis of the filtered spectra. Previously it was categorised as having multiple features across the frequency range but now the interference from

low frequency deviations have been removed it is clear that tag 20 has three distinct spectral peaks and is participating in the 6, 4 and 3 months cyclic behaviour.

Again, the tags with small black diamond symbols in Figure 7 have spectral content across the whole range and are thus random in the time domain.

*** Figures 6 and 7 ***

4.4 Discussions and insights

Origin of 3, 4 and 6 month cycles: The origins of the 3, 4 and 6 month seasonal behaviour can be inferred from an analysis of which tags share those cycles. Three and six month cycles are present in the bar and section mills (black triangles in Figure 7 and Table 1). The bar and section mills operate an order book and therefore make products for stock that are not necessarily synchronously with order arrivals. The spectra of the shifts worked tags (2 and 16) show the 3 and 6 month features especially strongly (0.17 and 0.34 on the frequency axis). The same cycles also appear in production (tags 1 and 15) and despatches of supplies of FG billets to the bar and section mills.

The four month cycle (black circles in Figure 7 and Table 1) are present in the orders of all three mills (i.e. tags 9, 23 and 37). Since the orders are independent inputs to the network it is concluded that the four month cycle is an externally induced disturbance. The same cycle is also present in the despatches.

The conclusion is that the three and six month cyclical disturbances are rogue seasonalities caused by the bar and section mills' production planning that propagate to material requests from the steel works. The three and six month cycles also have an impact on customers because the despatches show three and six month disturbances (tags 6 and 20).

There are no three and six month cycles evident in the rod mill. The rod mill is a make-to-order factory. The production and shifts-worked variables in the rod mill (tags 32, 33 and 37) all have the four month cycle with a spectral peak at 0.25 on the frequency axis. It is concluded that production in the rod mill is responsive to customer orders because production follows the pattern of customer orders. The material receipts from the steel works to the rod mill (tag 39) also shows a four month cycle.

Time-domain insights: The shifts-worked data (tags 2 and 16 in the bar and section mills and tag 33 in the rod mill) and the total orders (tags 9, 23 and 37) were studied further in order to validate the findings of spectral PCA. Figure 8 shows monthly averages for years 1 to 5. The sixth year was not included in the averages because the data set for the rod mill was partly missing and therefore the comparison could not be made.

The top panels (row (a)) in Figure 8 are for the rod mill. The orders show the four month cycle that was highlighted by spectral analysis because there are peaks in February-March, June and October and valleys in January, April-May, and August-September. The shifts worked (upper left panel) shows a similar pattern in which the months with the three lowest average shifts worked are April, August and December. The timing of peaks and valleys in the orders match well with the peaks and valleys in the shifts worked to within ± 1 month. That is why spectral PCA placed tags 37 (orders) and 33 (shifts worked) in the same cluster in Figure 7. A difference is that the average of shifts worked in December is relatively low compared with the December orders while the average shifts worked in November is relatively high, possibly in anticipation of the Christmas holiday period.

The middle panels of Figure 8 (row (b)) are for the section mill. The pattern of monthly average orders (tag 23) is similar to that of the rod mill orders (tag 37) and is also characterised by a four

month cycle. That is why spectral PCA placed tag 23 close to 37 in the same cluster. The pattern of shifts worked in the section mill is not similar to the pattern of orders (middle left panel in Figure 8). The average shifts worked show two peaks six months apart, one in Feb-April and the other in Sept-Nov. Spectral PCA also found that the shifts worked tag (16) had different characteristics than the orders (tag 23) and placed it in a different cluster in Figure 8.

The bottom panels in Fig 8 (row (c)) are for the bar mill. The pattern of shifts worked (Tag 2) is similar to that in the section mill (Tag 16) but it does not match the pattern of orders. These observations reflect the findings of spectral PCA which placed tags 2 and 16 in the same cluster and placed the orders (tag 9) fairly close to tags 37 and 23 in a different cluster. The bar mill orders are higher in March, July and October, similar to the pattern in the section and rod mills.

4.5 Practical uses for spectral PCA

Figure 9 shows a process chart outlining the procedure by which the spectral PCA may be applied. The first phase is purely technical, as illustrated by this exploratory study. This technical phase is automated and requires little, if any, input from management. With the advent of radio frequency identification (RFID), coupled with information and communication technologies (ICT) there is the potential to capture considerable volumes of real time data from the supply network. The automated phase enables the detection of rogue seasonality information quickly, consistently and rigorously thus allowing management more time to spend on interpretation than on numerical analysis.

The second phase requires interpretation of the results and relies on the tacit knowledge and experience of management. Table 2 provides the criteria by which to determine whether or not rogue seasonality is present in any of the internal variables in the supply network. If such rogue seasonality is identified it will then be incumbent on management to identify why it is occurring

and take action to reduce or eliminate it. For example, from the steel company case we identified 3, 4 and 6 monthly cycles of which only the 4 month cycle was present in the end customer demand. It is therefore necessary for management to determine the origin of the 3 and 6 monthly cycles, which can be classified as rogue, as these create unwanted variability, leading to uncertainty and increased total costs in the network (Davis, 1993). It is outside the scope of this paper to detail typical management actions but a first step is identifying cause and effect relationships via an appropriate business diagnostic technique such as a Quick Scan (Naim et al., 2002).

*** Figure 9 near here ***

*** Table 2 near here ***

The case study presented in this article used historical monthly data for six years of operation. It performed an historical audit of the operation of a supply network over that period and would therefore be of value to a consultant or manager addressing the performance issues of a supply network for the first time. It offers the following benefits:

- Automatically highlighting significant seasonal characteristics and showing which parts of the business share them;
- Reduction of large data sets that are hard to analyse into a smaller number of information-rich quantities in the form of the scores for each tag;
- Focusing of attention on relevant features and clusters of tags, thus shortening the time taken to reach conclusions. Managers can thus spend more time on gaining valuable insights into the behaviour of the network.

It is an attractive idea to use shorter historical episodes in order to quantify the immediate effects of new strategies on the supply network. Similar results were found when analysing a two year

episode. For instance spectral PCA explained 78% of the variability in months 25 to 48 using 3 PCs and also placed tags 23, 33, and 37 in one cluster and 16 and 2 in another. However, no result was found when using just one year of data.

The key parameters in the selection of data are sampling interval and the number of cycles of data captured in the time series. The technique is scalable and can be applied to rapidly changing environments such as high technology items or fast moving consumer goods provided the sampling interval is fast enough to capture dynamic effects and short period cycles. The Nyquist sampling theorem poses a constraint for short period cycles because it requires at least two samples per period. For example, with monthly samples it is not possible to detect seasonality with cycles shorter than two months. To detect weekly variations in high-tech goods such as microprocessors then the supply chain should be monitored at least two times per week. There are also constraints with very long period cycles. In the case of the steel industry where 7-10 year business cycles are present then at least ten years of data are needed to detect a ten year cycle rather than the six years available in this study.

The resolution of the spectral peaks becomes poor when the time series is short because the number of frequency channels is fewer. Analysis of shorter periods would require the use of weekly or daily rather than monthly data. Moreover, numerical considerations related to the finite data length cause leakage between frequency channels such that a sine wave belonging in one frequency channel may in fact influence adjacent frequency channels. A window function such as the Hanning window may be applied to the time series before calculating the spectrum in order to reduce spectral leakage. Alternatively, the problem can be avoided altogether if an integer number of complete cycles is used in the analysis (see Press *et. al.*, 1999).

Table 3 summarises the strengths and weaknesses of the spectral PCA and highlights the limitations of the exploratory study. This forms the basis for future research as outlined in the conclusions section.

*** Table 3 near here ***

5. Conclusions

The exploratory study reported in this article suggests that principal component analysis of power spectra is a useful tool for identification of those variables in a supply network that have dynamic features. Spectral PCA captured more of the variability in a data set than a conventional multivariate analysis based upon the time series, and it also provided a spectral signature to characterise the behaviour. It was insensitive to missing data in the historical record and produced a set of descriptive plots to aid communication of the results.

Spectral PCA has the capability to detect and characterise network-wide cyclical disturbances in information and material flows. A consultant or manager can then consider the groups of variables participating in each disturbance to reach a conclusion about the root cause. The technique was applied to data from a supply network in the steel sector comprising four autonomous business units. Two cyclical disturbances indicating seasonality were detected and characterised. Their distribution within the network showed that one was externally induced by a cyclic pattern of customer orders, while the other was found to be rogue seasonality induced through the internal production planning systems.

As this was an exploratory study we have highlighted both the strengths and weaknesses of the technique. The strengths highlight the significant potential for the technique in a “real time” data collection and analysis system. The weaknesses suggest the way ahead for further research such as;

- capturing the management deductive processes in interpreting the outputs from the tool, say via knowledge based systems or expert systems
- testing the tool through simulation where it is possible generate data with different sampling periods and inducing different disturbances other than cyclical disturbances
- comparing the spectral PCA with independent component analysis (ICA), which has the potential for providing greater clarity in establishing clusters
- developing the tool for short time series analysis (Grindlay, 2001).

6. Acknowledgements

The industrial data set and supporting material was provided by a steel company partner on EPSRC research grant GR/F 56475 & DTLR grant PPAD 9/64/127. Nina Thornhill gratefully acknowledges the financial support of the Royal Academy of Engineering (Foresight Award).

7. References

- Ashby, W. R. (1966) *Design for a brain: the origin of adaptive behaviour*, Chapman and Hall, London.
- Belchamber, R.M., and Collins, M.P. (1993). *Method for monitoring acoustic emissions*, European Patent Office, Publication N^o 0 317 322 B1.
- Berry, D., Evans, G. N. and Naim, M. M. (1998) Pipeline information survey: a UK perspective, *International Journal of Management Science, OMEGA*, Feb., Vol. 26, No. 1, pp 115-132.
- Chatfield, C., and Collins, A.J. (1980). *Introduction to multivariate analysis*, Chapman and Hall, London, UK.
- Chatfield, C., (1996), *The Analysis of Time Series: An Introduction*, 5th edition, *CRC Press*.

Chen, F., Drezner, Z., Ryan, J.K. and Simchi-Levi, D. (2000) Quantifying the Bullwhip effect in a simple supply chain: The impact of forecasting, lead-times and information, *Management Science*, Vol. 46, No. 3, , pp 436-443.

Davis, T., (1993) Effective Supply Chain Management, *Sloan Management Review*, Summer, pp 35-46.

Dejonckheere, J., Disney, S. M., Lambrecht, M. R. and Towill, D. R., (2003) Measuring and avoiding the bullwhip effect: a control theoretic approach, *European Journal of Operational Research*, Vol. 147, pp 567-590.

Disney, S.M. and Towill, D.R., (2002) A discrete transfer function model to determine the dynamic stability of Vendor Managed Inventory supply chains, *International Journal of Production Research*, Vol. 40, No. 1, pp179-204

Grindlay, J., (2001) On an application of a generalisation of the discrete Fourier transform to short time series, *Canadian Journal of Physics*, Vol. 79, No. 5, pp 857-868.

Harland, C.M., (1996) Supply Chain Management: Relationships, Chains and Networks, *British Journal of Management*, Vol 7 (Special Issue) S63-S80,.

Karstang, T.V., and Henriksen, A. (1992). Infrared spectroscopy and multivariate calibration used in quantitative analysis of additives in high density polyethylene, *Chemometrics and Intelligent Laboratory Systems*,, 14, 331-339.

Mason-Jones, R. and Towill, D.R., (1998) Shrinking the supply chain Uncertainty Circle. *IOM Control*, September, 17-22.

- Naim, M. M., Childerhouse, P., Disney, S. M. and Towill, D.R., (2002) A supply chain diagnostic methodology: determining the vector of change, *Computers & Industrial Engineering: An International Journal*, Vol. 43, No. 1-2, pp 135-157
- Parnaby, J. (1991) Designing effective organisations, *International Journal of Technology Management*, Vol 6 No 1/2 pp 15-31.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1999), *Numerical Recipes in C*, Cambridge University Press.
- Riley M.R., Rhiel M., Zhou X., Arnold M.A., and Murhammer D.W. (1997). Simultaneous measurement of glucose and glutamine in insect cell culture media by near infrared spectroscopy. *Biotechnology and Bioengineering*. 55: 11-15.
- Seasholtz, M.B., 1999, Making money with chemometrics, *Chemometrics and Intelligent Laboratory Systems*, 45, 55-64.
- Stevens, G., (1989) Integrating the Supply Chain, *International Journal of Physical Distribution and Materials Management*, Vol 19, No 8, pp 3-8,
- Thornhill, N.F., Shah, S.L., Huang, B., and Vishnubhotla, A., (2002) Spectral principal components analysis of dynamic process data, *Control Engineering Practice*, 10, pp 833-846.
- Wold, S., Esbensen, K., and Geladi, P. (1987) Principal Component Analysis, *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.
- Wu, H.D., Siegel, M., Khosla, P., (1999) Vehicle sound signature recognition by frequency vector principal component analysis, *IEEE Transactions on Instrumentation and Measurement*, 48 1005-1009.

Valle, S., Li, W.H., and Qin, S.J. (1999) Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods, *Industrial and Engineering Chemistry Research*, 38, 4389-4401.

Watson, G.H., (1994) *Business Systems Engineering*, John Wiley & Sons Inc, New York

Yeung, K. S. Y., Hoare, M., Thornhill, N.F., Williams, T., Vaghjiani, J.D. (1999). Near infra red spectroscopy for bioprocess monitoring and control, *Biotechnology and Bioengineering.*, 63, 684-693.

Tag No	Description			Tag No	Description		
Bar mill				Section mill (cont'd)			
1	Production	←	←	24	Receipts from FG	←	←
2	Shifts worked	←	←	25	Receipts, other	◆	◆
3	Output per shift	◆	◆	26	Receipts, total	←	←
4	Despatches, home	←	■	27	Billet stocks, total	◆	◆
5	Despatches, export	◆	◆	28	Total stocks	○	◆
6	Despatches, total	←	■	Bar/section mill			
7	Home orders	◆	■	29	Order book, bar mill	□	◆
8	Export orders	◆	◆	30	Order book, section mill	□	◆
9	Total orders	◆	●	31	Order book, total	□	◆
10	Receipts from FG	←	←	Rod mill			
11	Receipts, other	◆	●	32	Production	◆	●
12	Receipts, total	←	←	33	Shifts worked	◆	●
13	Billet stocks, total	◆	◆	34	Output per shift	◆	◆
14	Total stocks	○	◆	35	Billet stocks, total	○	◆
Section mill				36	Total stocks	◆	◆
15	Production	←	←	37	Total orders	◆	●
16	Shifts worked	←	←	FG Billets			
17	Output per shift	◆	◆	38	Total stocks	○	◆
18	Despatches, home	←	■	39	Despatches, rod	○	●
19	Despatches, export	◆	●	40	Despatches, bar/section	○	■
20	Despatches, total	◆	■	41	Despatches, other	◆	◆
21	Home orders	◆	●	42	Despatches, total	○	■
22	Export orders	◆	◆	43	Production	○	■
23	Total orders	◆	●				

Spectral PCA using full spectra (left hand columns)

- ← Tags with two or more distinctive spectral features
- Tags with long term non-stationery trends
- Tags with a two-year cycle
- ◆ Tags with no distinctive spectral features

High frequency spectral PCA (right hand columns)

- Tags with 4 month cycle
- Tags with 3, 4 and 6 month cycles
- ← Tags with 3 and 6 month cycles
- ◆ Tags with no distinctive spectral features

Table 1. Results of spectral PCA analysis of the steel supply network data

Notes for typesetting

- ← is Monotype Sorts character 115 outline font – should be a white triangle
- ← is Monotype Sorts character 115 – should be a black triangle
- is WingDings character 161 – should be a white circle
- is WingDings character 108 – should be a black circle

- is WingDings character 168 – should be a white square
- is WingDings character 110 – should be a black square
- ◆ is WingDings character 119 – should be a small black diamond

Tag No	Description			Tag No	Description		
Bar mill				Section mill (cont'd)			
1	Production	△	▲	24	Receipts from FG	△	▲
2	Shifts worked	△	▲	25	Receipts, other	◆	◆
3	Output per shift	◆	◆	26	Receipts, total	△	▲
4	Despatches, home	△	■	27	Billet stocks, total	◆	◆
5	Despatches, export	◆	◆	28	Total stocks	○	◆
6	Despatches, total	△	■	Bar/section mill			
7	Home orders	◆	■	29	Order book, bar mill	□	◆
8	Export orders	◆	◆	30	Order book, section mill	□	◆
9	Total orders	◆	●	31	Order book, total	□	◆
10	Receipts from FG	△	▲	Rod mill			
11	Receipts, other	◆	●	32	Production	◆	●
12	Receipts, total	△	▲	33	Shifts worked	◆	●
13	Billet stocks, total	◆	◆	34	Output per shift	◆	◆
14	Total stocks	○	◆	35	Billet stocks, total	○	◆
Section mill				36	Total stocks	◆	◆
15	Production	△	▲	37	Total orders	◆	●
16	Shifts worked	△	▲	FG Billets			
17	Output per shift	◆	◆	38	Total stocks	○	◆
18	Despatches, home	△	■	39	Despatches, rod	○	●
19	Despatches, export	◆	●	40	Despatches, bar/section	○	■
20	Despatches, total	◆	■	41	Despatches, other	◆	◆
21	Home orders	◆	●	42	Despatches, total	○	■
22	Export orders	◆	◆	43	Production	○	■
23	Total orders	◆	●				

Spectral PCA using full spectra (left hand columns)

△ Tags with two or more distinctive spectral features

○ Tags with long term non-stationery trends

□ Tags with a two-year cycle

◆ Tags with no distinctive spectral features

High frequency spectral PCA (right hand columns)

● Tags with 4 month cycle

■ Tags with 3, 4 and 6 month cycles

▲ Tags with 3 and 6 month cycles

◆ Tags with no distinctive spectral features

Table 1. Results of spectral PCA analysis of the steel supply network data

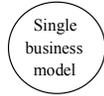
Exogenous demand (Tags 7-9, 21-23, 37)	Internal variables (remaining tags)	Rogue seasonality present?
Shows no seasonality	Shows no seasonality	No
Shows seasonality	Shows no seasonality	No
Shows seasonality	Shows seasonality – same characteristics as for exogenous demand	No
Shows no seasonality	Shows seasonality	Yes
Shows seasonality	Shows seasonality – different characteristics compared to exogenous demand	Yes

Table 2: Look up table for identifying whether rogue seasonality is present or not – see Figure 9.

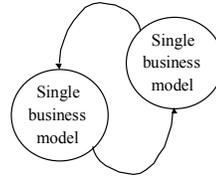
Strengths	Weaknesses
Automated data analysis procedure	A considerable amount of data is needed to get meaningful spectra and statistics – such data is not always available
Data analysis is quick	The management phase requires experience – can be subjective
Analysis yields unified statistics	The study only used a monthly sample period.
Technique allows more time on trying to gain insights than data analysis	The technique has been limited to identifying rogue seasonality and not other disturbances – e.g. pulses
Technique is proved in other applications	Spectral PCA has not yet been directly compared with other techniques
Technique is robust to missing data	
The study used real world data	

Table 3: Strengths and weakness of the spectral PCA as applied in the exploratory study

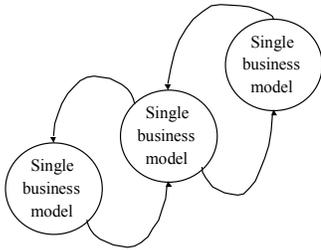
Single unit



Dyadic structure



Supply chain structure



Network structure

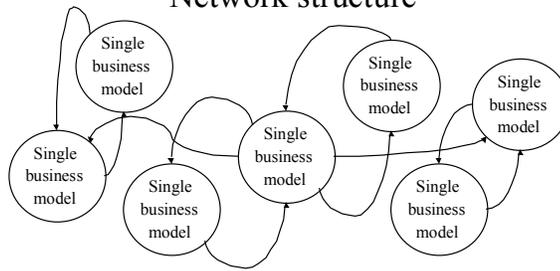


Figure 1: Supply system units of analysis (based on Harland, 1996)

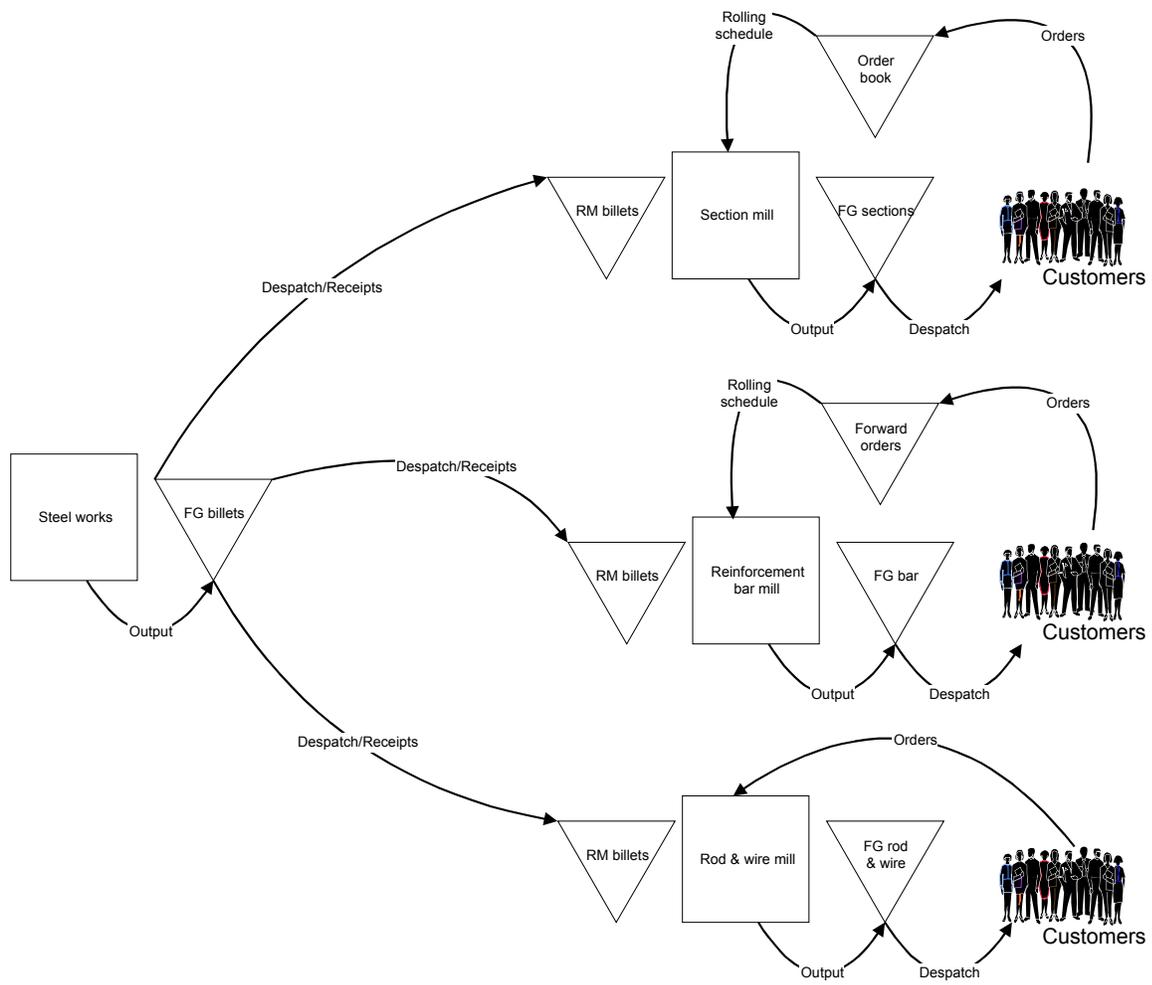


Figure 2: Supply network under study

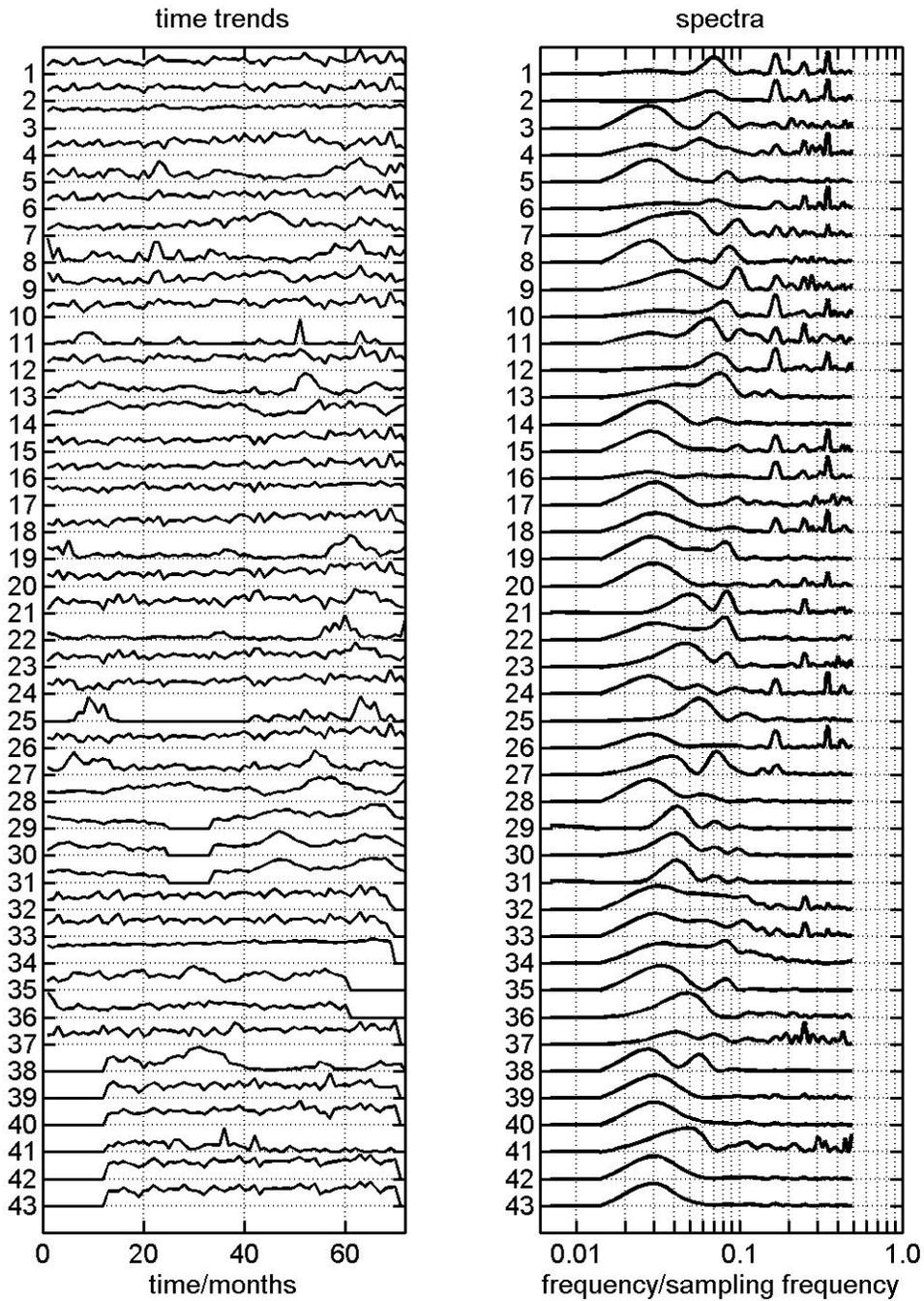


Figure 3. Time series and spectra of the supply network data. Spectra are scaled to the same maximum peak height and are plotted on a linear vertical axis.

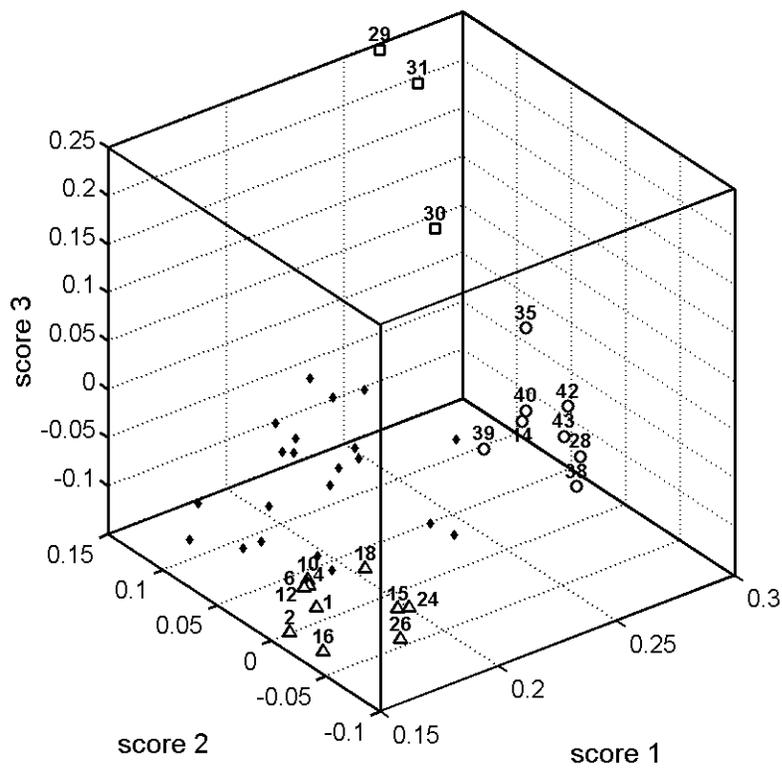


Figure 4. Three-PC score plot from spectral PCA analysis. The numerals indicate tag numbers.

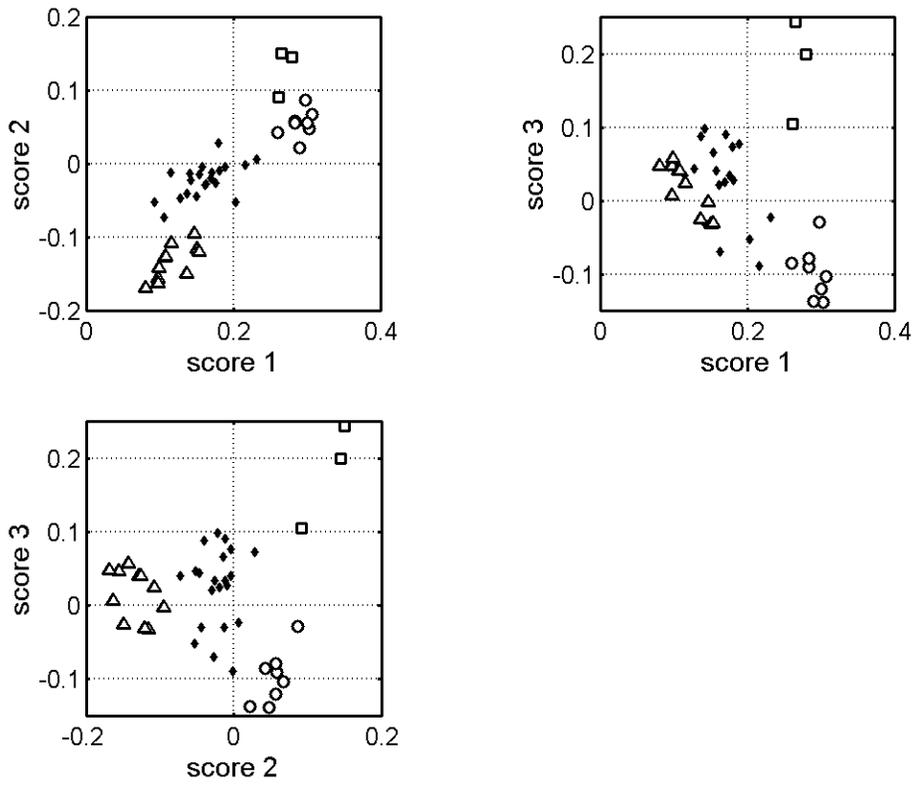


Figure 5: Two dimensional score plots. The meanings of the symbols are explained in section 4.1.

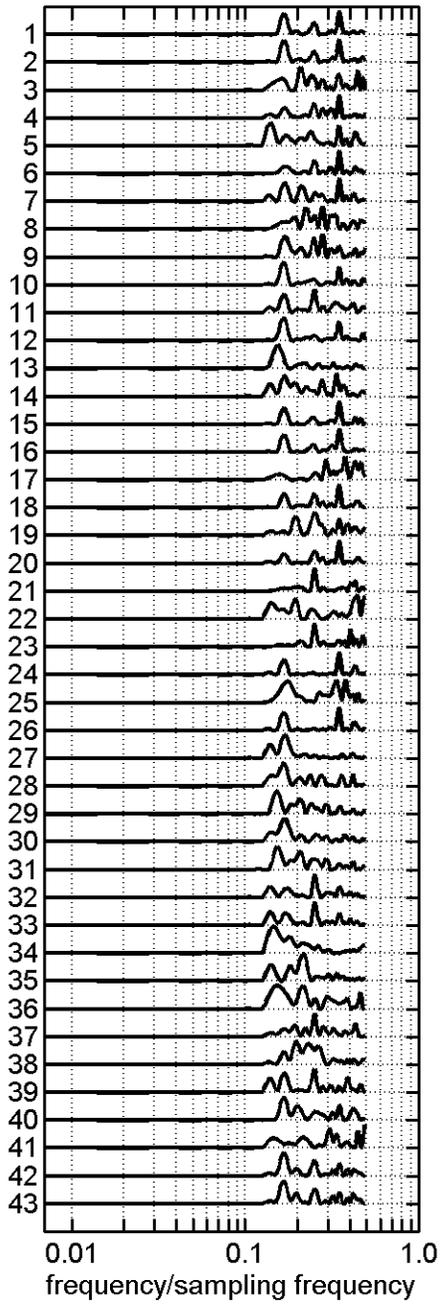


Figure 6: Filtered spectra with low frequency features removed. Spectra are scaled to the same maximum peak height and are plotted on a linear vertical axis.

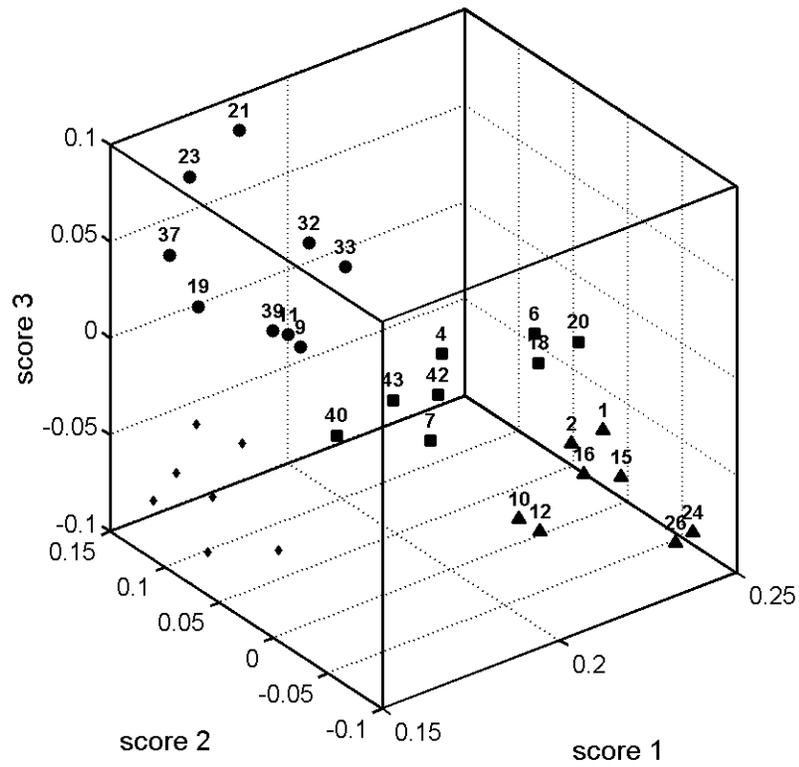


Figure 7: Three-PC score plot from analysis of filtered spectra with low frequency features removed. The numerals indicate tag numbers.

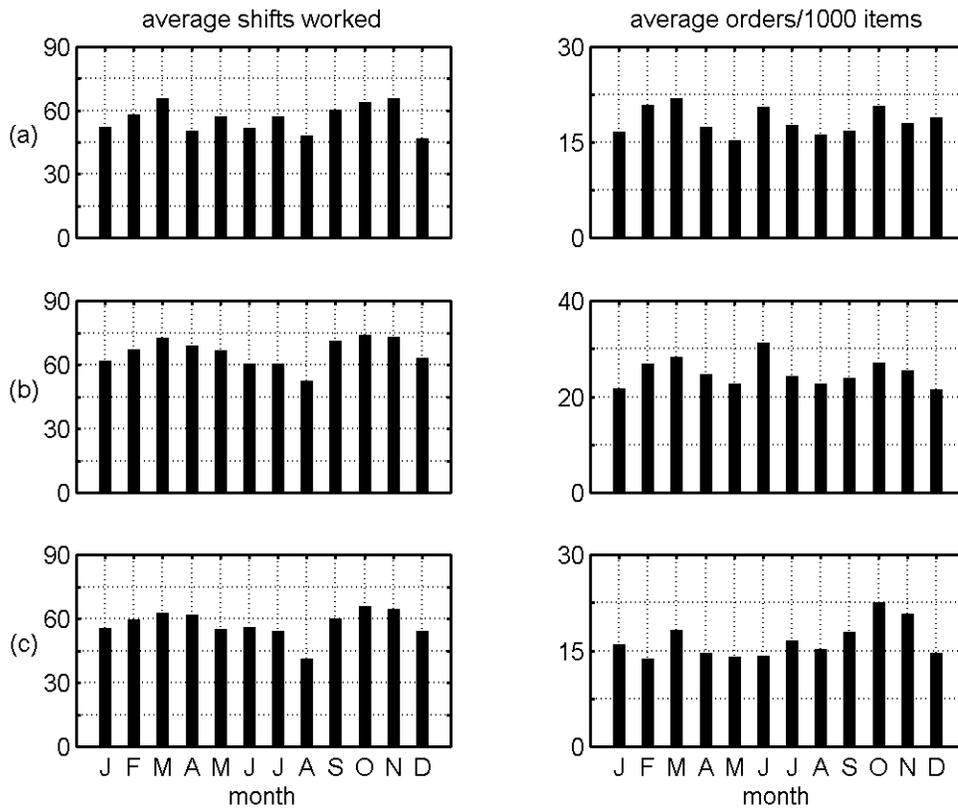


Figure 8: Five year averages of shifts worked and orders at the (a) rod, (b) section and (c) bar mills.

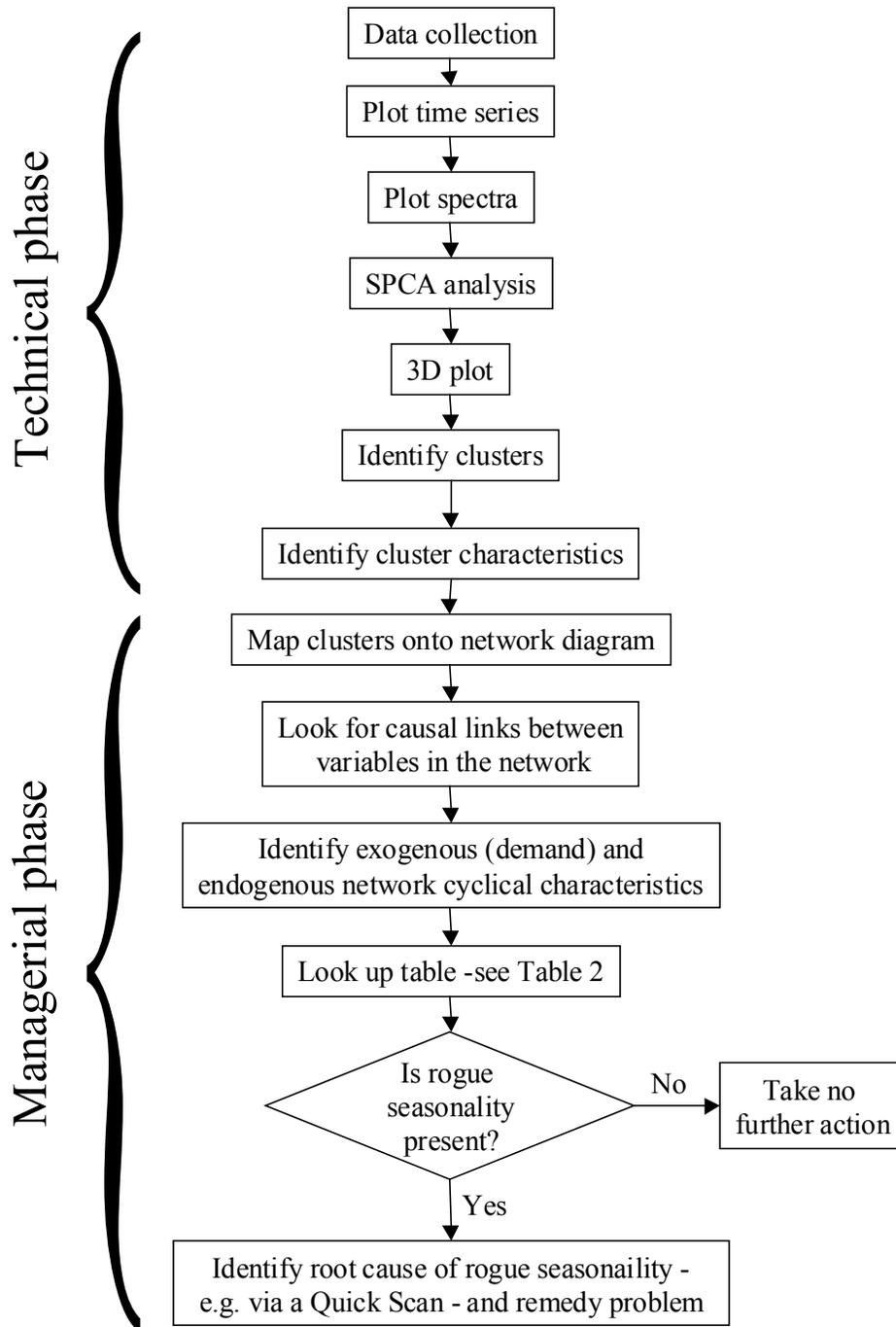


Figure 9: Process for application of the spectral PCA to identify rogue seasonality in a supply network