JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 383–399

Wavelet–PLS regression models for both exploratory data analysis and process monitoring

Pekka Teppola* and Pentti Minkkinen

Chemometrics Group, Laboratory of Inorganic and Analytical Chemistry, Lappeenranta University of Technology, PO Box 20, FIN-53851 Lappeenranta, Finland

SUMMARY

Two novel approaches are presented which take into account the collinearity among variables and the different phenomena occurring at different scales. This is achieved by combining partial least squares (PLS) and multiresolution analysis (MRA). In this work the two novel approaches are interconnected. First, a standard exploratory PLS model is scrutinized with MRA. In this way, different events at different scales and latent variables are recognized. In this case, especially periodic seasonal fluctuations and long-term drifting introduce problems. These low-frequency variations mask and interfere with the detection of small and moderate-level transient phenomena. As a result, the confidence limits become too wide. This relatively common problem caused by autocorrelated measurements can be avoided by detrending. In practice, this is realized by using fixed-size moving windows and by detrending these windows. Based on the MRA of the standard model, the second PLS model for process monitoring is constructed based on the filtered measurements. This filtering is done by removing the low-frequency scales representing low-frequency components, such as seasonal fluctuations and other long-term variations, prior to standard PLS modeling. For these particular data the results are shown to be superior compared to a conventional PLS model based on the non-filtered measurements. Often, model updating is necessary owing to non-stationary characteristics of the process and variables. As a big advantage, this new approach seems to remove any further need for model updating, at least in this particular case. This is because the presented approach removes low-frequency fluctuations and results in a more stationary filtered data set that is more suitable for monitoring. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: paper and pulp industry; activated sludge wastewater treatment plant; chemometrics; exploratory data analysis; process monitoring; partial least squares; multiresolution analysis; wavelets

INTRODUCTION

Recent years have given rise to new approaches to monitor complex industrial processes. The improvement in technology to produce and collect data has led to a dimensional explosion of the amount of data to be analyzed. Often it is simply not enough to monitor only a few quality-related

* Correspondence to: P. Teppola, BASF AG, Scientific Computing, ZDP/C–C13, D–67056 Ludwigshafen(Rh), Germany.
E-mail: pekka.teppola@basf-ag.de
Contract/grant sponsor: Graduate School of Chemical Engineering.
Contract/grant sponsor: Academy of Finland.

Received 1 September 1999
Accepted 4 March 2000

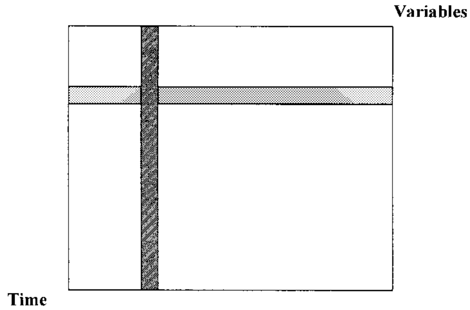

Figure 1. Two different modes of multivariate multiscale data analysis: the vertical mode is related to time series information and the horizontal mode to measured variables. Real processes are both multiscale and multivariate in nature. The vertical mode investigates events occurring at different scales, while the horizontal mode captures and models the variable interactions by exploiting both collinearity and redundancy of measured information.

variables. A collection of multivariate techniques has been developed to treat various kinds of processes, such as continuous and batch processes. Instead of just applying univariate monitoring charts, researchers have started applying multivariate methods that take into account the underlying intercorrelation structures among the variables. These methods include principal component analysis (PCA), principal component regression (PCR), partial least squares (PLS) and canonical variates (CV) to mention only a few [1–4]. These methods belong to the group of inferential or latent variable (LV) methods. However, even more sophisticated inferential methods have been developed that take into account the special characteristics of different processes. For instance, multiway models [5–8] have been applied to batch processes, multiblock models [9–11] to processes where one can form logical blocks of data, and dynamic models [4,12,13] to autocorrelated continuous processes. In this work we are interested in scrutinizing the process dynamics and then executing multivariate monitoring based on the acknowledgement of the underlying process dynamics. Traditionally this has led to time series modeling [14], which can be a very complex task to carry out successfully. This can be even more problematic if the process changes its characteristics over time. For this reason it is necessary to either update models or otherwise treat the undesired disturbing process dynamics. In general, each process has its own features that must be recognized and acknowledged. These features are often multivariate and multiscale in nature (see Figure 1). Figure 2 illustrates the multiscale nature of the data by using an artificial signal.

The present paper describes a new approach to study the characteristic features of a particular activated sludge wastewater treatment plant. Recently we have made several attempts to treat the process dynamics of this particular plant, such as a dynamic PLS modeling [15], a model updating based on Kalman filtering [16], and a hidden process path modeling [17,18] that to some extent is similar to interpretative case-based reasoning (CBR) methodology [19], where new cases are compared to the casebase of prototypes of both the process faults and the normal operating conditions in different seasons. The first approach did not work well, for several reasons that are related to the plant itself and which are listed in the next paragraph. The second approach worked well in relation to prediction, but it was difficult to interpret the behavior of the process over time, since parameters were updated recursively. The third approach is still under investigation and looks very promising, because it emphasizes learning by experience and interpretation based on fuzzy and possibilistic classification of objects. However, it can be made adaptive without losing interpretability.

The present work was also motivated by several reasons arising from the plant itself. The particular plant is very complex in that it is a biochemical process that is, in addition, under the influence of slow

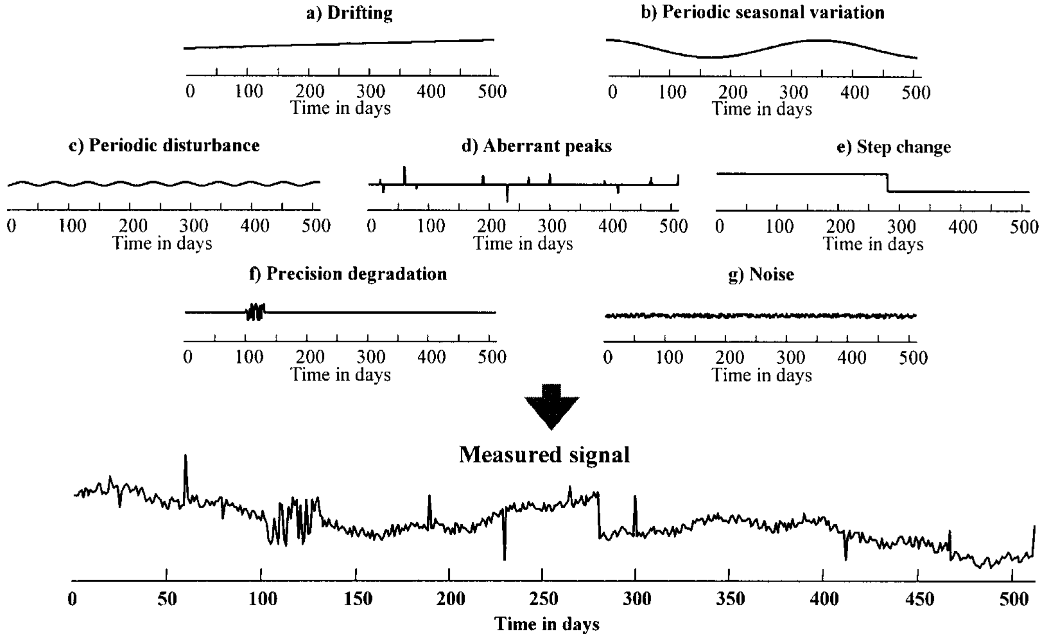

Figure 2. Multiscale nature of the measured signal. Measurement signals can have both deterministic and stochastic components. The former can often be localized in either/both the time or/and the frequency domain, while the latter behave like random variables and are distributed more evenly. The stochastic components can often be both scale- and time-dependent. For instance, noise occurs at high-frequency scales, while drifting and seasonal fluctuation occur at low-frequency scales. Different types of disturbances and faults have been exemplified above. These different events sum up to the measured signal. Note that the signal is artificial.

variations, such as seasonal fluctuations and other long-term dynamics. Moreover, the feed to the process depends on both the preceding process units and the previous process stages, which are controlled on a very different basis from those related to the wastewater treatment plant itself. Hence the feed cannot be controlled directly and thus can have large variations, which must be compensated by the operators in the wastewater treatment plant. In addition, the effects of these compensating actions usually take place very slowly. Furthermore, the process can undergo instrumental changes and even fundamental changes in the way it is operated over the years. Bearing in mind these issues, the particular plant is far from the stationary processes that are relatively easy to model and monitor, because in such processes only the multivariate nature of the data must be taken care of. However, last but not least, these reasons result in an excessive amount of work related to model updating that can be too demanding for the process operators.

Traditionally in process monitoring the majority of the models used are based on the i.i.d. principle [4,13], i.e. the measurement vectors are independent and identically distributed or, in other words, not autocorrelated over time. As aforementioned, this is not the case in this particular plant, which gives rise to problems related to sensitivity and robustness. Sensitivity can have various meanings, but here it is meant that some disturbances can mask others and even deteriorate all the monitoring efforts. For example, small but significant faults can be masked by larger seasonal fluctuations (see Figure 3). By using the term robustness, it is meant that the model is no longer representative over time owing to changes that have taken place in the process. From the modeling point of view these changes can take place in the mean values and/or the covariance structure [20]. If this happens, either the model must

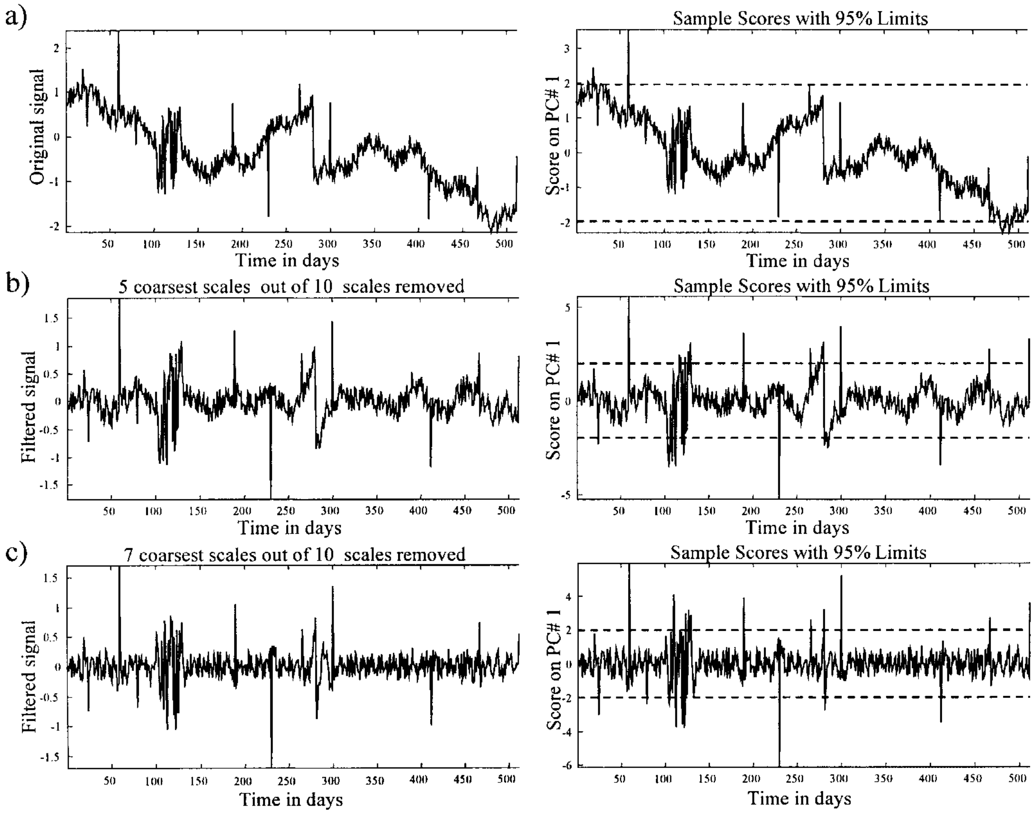

Figure 3. An example of the masking effect and detrending. An example of rank-one PCA models for the artificial signal already presented in Figure 2 and the corresponding confidence limits for the **X** block scores. (a) The original non-filtered signal, (b) the first filtered signal and (c) the second filtered signal. The two filtered signals represent reconstructed signals where low-frequency scales and thus events have been filtered out using MRA. Each signal was mean centered and scaled to unit variance before PCA modeling. The purpose of such a treatment is to focus on transient high-frequency events and to remove the masking effect of low-frequency events.

be updated or a different approach must be undertaken. The former has been well explored, while the latter and thus process dynamics, autocorrelated measurements and multiscale events have not been of much interest.

To treat multiscale events, new approaches must be undertaken. For this purpose, wavelets have proven useful. They have been applied to denoising [21], data rectification [22], time series modeling [23], trend analysis [24,25] and data compression [26]. All these applications result from several desirable features. For example, wavelets are localized in both the time and the frequency domain [27]. Wavelets automatically adapt to high- and low-frequency components. For the former a small window is used and for the latter a large window. This is an advantage compared to Fourier analysis. In more detail, wavelets capture differences from the data (from signals) and represent these differences at different scales. For instance, all the wavelet coefficients would be zero for a constant steady state signal. Wavelets are also well suited for aperiodic signals. They can extract both deterministic and stochastic components. In general, deterministic components are represented by large wavelet coefficients that are localized in either the time or the frequency domain, while stochastic components are represented by smaller coefficients distributed more evenly at different scales or time periods [28]. Based on normality considerations, these coefficients can also be used to estimate noise levels and thresholds for denoising. Moreover, wavelet coefficients typically exhibit less dependence than does the original data set, as documented by Bakshi [28], Bilen [29] and Luo *et al.* [30]. Based on this important property, wavelet coefficients can be manipulated, e.g. by retaining only a part of the wavelet coefficients. Depending on the objectives, signals can be filtered in different ways, one of which has been exemplified in Figure 3. In this work we want to monitor transient high-frequency faults and disturbances and, at the same time, to filter out seasonal variations and other long-term fluctuations. In other applications, signals are often smoothed by removing noise and other high-frequency stochastic components. In conclusion, wavelets are very efficient in detecting characteristics such as discontinuities, trends and even heteroscedasticity. In addition, wavelets can offer means to investigate other mathematical methods and can also be combined with other methods. For instance, Starck *et al.* [31] have combined wavelets and other data analysis techniques such as PCA, time series modeling, recurrent neural networks and self-organizing maps, while Bakshi [23] has, very interestingly, developed multiscale PCA.

Briefly, the present work introduces two novel approaches, i.e. one for exploratory data analysis and another for process monitoring. In this way it is possible to take into account both the multivariate and the multiscale nature of the particular process. These approaches are realized by combining PLS and MRA. In the first approach a standard PLS model is scrutinized with MRA. In this way, both high- and low-frequency events occurring at different scales are recognized. Normally good monitoring data consist only of the common-cause variation. In this case, especially periodic seasonal fluctuations and long-term drifting introduce problems. These low-frequency variations mask and interfere with the detection of small and moderate-level transient phenomena. As a result, the confidence limits become too wide, as can be seen from Figure 3(a). By detrending, this relatively common problem of autocorrelated measurements can be avoided (see Figures 3(b) and 3(c)). Thus the second PLS model is constructed based on the filtered measurements. In practice, data are represented using fixed-size moving windows that are decomposed, investigated and manipulated. The filtering is done by removing the low-frequency scales representing low-frequency components, such as seasonal fluctuations and other long-term variations, after which a standard PLS model is computed and applied to real-time monitoring.

THEORY

Partial least squares/projection to latent structures (PLS)

Very often in industrial applications the data are severely corrupted by noise and collinearities among a high number of variables. To treat these problems, it is convenient to apply latent variable models, particularly partial least squares (PLS) modeling. PLS maximizes the covariance between process variables and responses. In PLS the matrix **X** (process variables) is decomposed and modeled in such a way that the information in **Y** (responses) can be predicted as well as possible. In addition, PLS uses only the variation in the **X** matrix that is significant in predicting the variation in the **Y** matrix. Moreover, one does not assume that the **X** variables are free of noise as in MLR. Therefore noise and insignificant variations are not used in modeling. In PLS the objective is achieved by decomposing both matrices into a set ofloadings (**P** and **Q**), loading weights (**W**) and scores (**T** and **U**). Loadings and loading weights refer to variables, while scores refer to objects. The first few latent variables (LVs) usually capture most of the systematic variation, while the last LVs mostly consist of noise and variations that are not related to **Y**. Importantly, the LVs are orthogonal to each other. These features together make it possible to compress information in the presence of collinearity and redundancy. Detailed theory is given elsewhere [1–3].

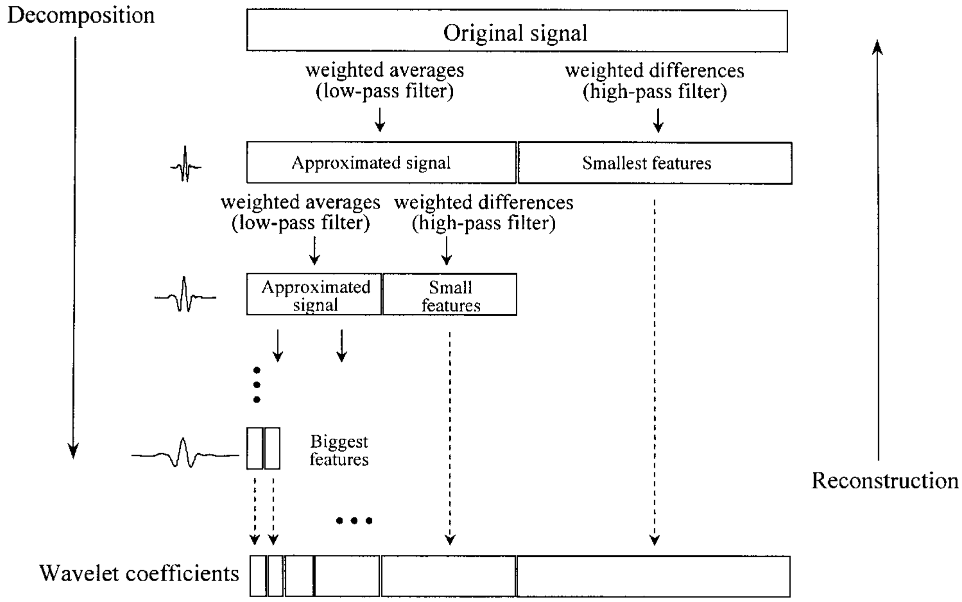

Figure 4. The basic idea in multiresolution analysis with dyadic sampling [26,40].

Wavelets and multiresolution analysis (MRA)

To take into account the multiscale appearance of different events, it is necessary to decompose signals. For this purpose, discrete wavelet decomposition (DWT) and multiresolution analysis (MRA) form a promising and flexible alternative [27]. An orthogonal MRA algorithm provides the means to decompose signals, whether original or inferential. In Figure 4 the principle of MRA has been illustrated. Briefly, the original signal is recursively decomposed at a resolution differing by a factor of two from the previous resolution [27]. This is realized by computing weighted differences (highpass filter) and weighted averages (lowpass filter). The differences are stored as wavelet coefficients. In decomposition, smallest features (differences) are first extracted, resulting in an approximated signal (see scale 1). From this approximation, new features (differences) are extracted, resulting in an even coarser approximation (see scale 2). This continues until the average signal has been approximated (scale *n*). In the other direction the differences between successive scales are added to the average signal, resulting in signals each at a resolution twice as great as the previous approximation [27]. If all wavelet coefficients are used, this reconstruction perfectly composes the original signal.

The biggest practical problem seems to be the selection of an appropriate type of wavelet function. Unlike the Fourier transform, the wavelet transform is not unique [27]. There are different types of wavelet functions that can be applied. The simplest wavelet is the Haar wavelet [32], which is short, orthonormal, irregular and asymmetric, with a compact support and a minimal number of vanishing moments. The advantages of orthonormality include efficiency, invertibility, concision, perfect reconstruction and ease of computation. As a drawback, scales can be somewhat arbitrary. Regularity, on the other hand, is a measure of smoothness that is often associated with ease of interpretation [27,32]. Support means the smallest closed set outside which the function vanishes identically, while compact means that the closed set is bounded. Thus the wavelet is zero outside a bounded interval

[32]. Compactness is usually related to computational efficiency. The number of vanishing moments is also related to computational efficiency. The higher the number of vanishing moments, the more information will be concentrated in a smaller number of wavelet coefficients [27]. Often the choice of wavelet function is a compromise between compactness and smoothness of the different types of wavelets [27]. Nesic *et al.* [33] and Staszewski [34] have also given some guidelines for the selection of wavelet functions.

In general, the simplest wavelet, Haar, is only suitable for step functions [27,32,34]. If it is used for smooth signals, many wavelet coefficients are needed to approximate smooth features. For this reason, smoother wavelet functions are preferred. Of the two most frequently used wavelets, called Symmlets and Daublets, Symmlets are used in this work. Both these functions are orthonormal, with a compact support and a high number of vanishing moments. However, both of them are longer and smoother than the Haar wavelet and are capable of capturing smooth low-frequency features [33,34], such as periodic seasonal fluctuations and long-term drifting. Symmlets are preferred over Daublets because they are less asymmetric [27], resulting in easier interpretation.

SOFTWARE

The PLS model was calculated using PLS Toolbox Version 1·5 [35]. MRA was carried out using WaveLab Version. 701 [36]. Both toolboxes work under Matlab 5.x[37].

EXPERIMENTAL

Process data

The process data were collected from an activated sludge wastewater treatment plant that purifies effluents from the Enso Publication Papers Oy Ltd paper mill in Summa, Finland. Twenty process and control variables, the **X** block, were used to model four purification efficiency-related variables, the **Y** block. The **Y** block consists of the diluted sludge volume index (DSVI) and the reductions of chemical oxygen demand (COD), nitrogen and phosphorus. The process and the variables are explained elsewhere in more detail [15,38].

Analyzing PLS models via MRA—exploratory data analysis

The variables in both the **X** block and the **Y** block were mean centered and scaled to unit variance before PLS modeling. The data set contained daily values for each of the **X** and **Y** variables starting from 1 January 1995 and ending on 28 February 1997, with a total number of 787 objects. In this first example the whole data set was used in exploratory data analysis. The number of latent variables was determined via a cross-validation procedure, in which a continuous block of 10% of the calibration data was set aside in turn and 90% of the samples were used in modeling. A leave-one-out cross-validation would have been too optimistic because of the autocorrelated measurements. Based on the predictions of the left-out samples in cross-validation, seven latent variables were retained in the model. After constructing the PLS model, multiresolution analysis was applied to the **X** block scores. In this way, each latent variable was interpreted individually. In MRA a Symmlet wavelet with 10 vanishing moments was applied. The procedure has been sketched in Figure 5(a) and documented in detail in Figure 6(a).

Before analyzing the score vectors one by one through MRA, they were padded to a dyadic length. In this work, so-called level padding was tested. This padding scheme was accomplished in such a way that missing values ($1024 - 787 = 237$ dummy objects) were divided into two groups (118 and 119 dummy objects). The first group was given values equal to the first real object of the window of the 787 real objects. Similarly, the last real object determined values for the second group of dummy

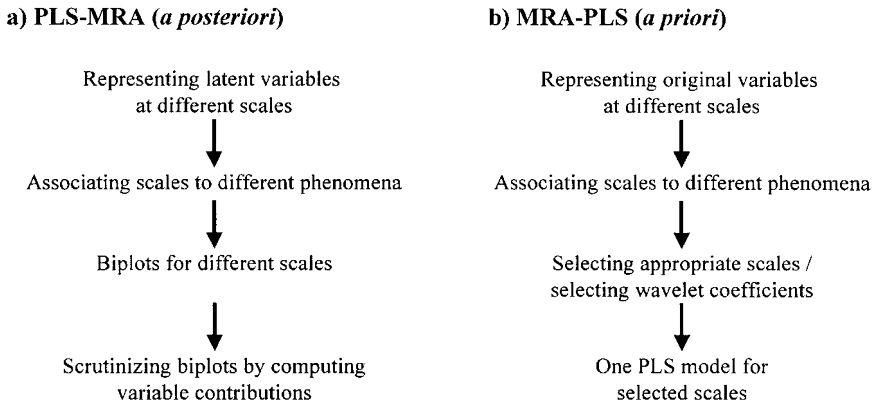

Figure 5. Two novel approaches (a) to investigate latent variable models and data and (b) to focus on certain scales (events) by filtering out the undesired and disturbing scales (events).

objects. Figure 7 exemplifies this level padding in general by using an artificial signal. If the vector is not in a dyadic length, then it must be padded either to the dyadic length or to the next larger dyadic length, resulting in an additional scale. In this work the need for the additional scale was made on a rather heuristic basis to avoid boundary problems. These problems have been described by many authors [26,39]. Figure 8 illustrates the boundary problem with and without the additional scale by

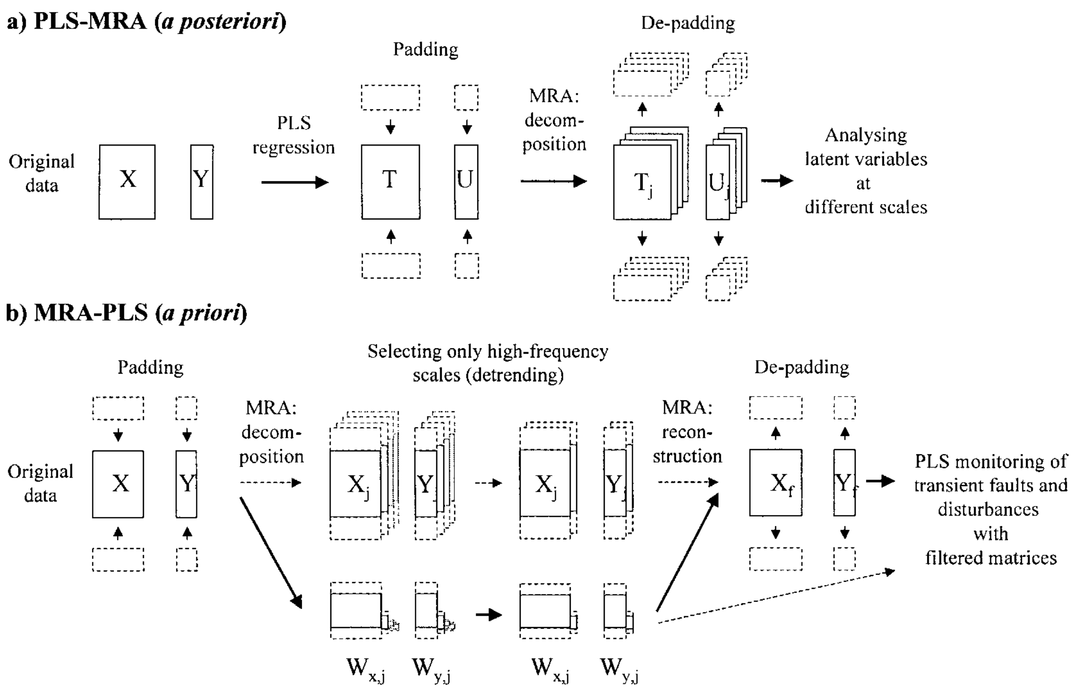

Figure 6. Detailed approaches (a) to investigate latent variable models and data and (b) to focus on certain scales (events) by filtering out the undesired and disturbing scales (events).

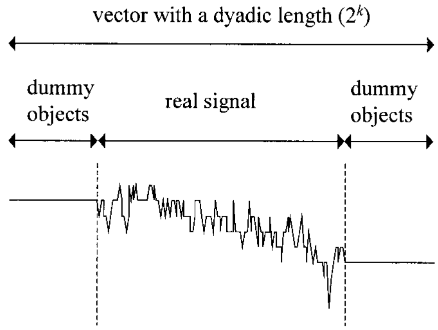

Figure 7. Level padding of a real signal to a dyadic-length signal.

using an artificial signal. One can very clearly observe undesired boundary problems in the detailed signals presented in Figure 8(a). If this approach without any padding scheme is to be used in monitoring where the present and the most recent values are on the right side of the moving window, one can be sure that faulty alarms will result. Figure 8(b) presents the corresponding results of level padding with no boundary problems. The additional scale is due to the excessive use of dummy objects, which will ensure error-free performance.

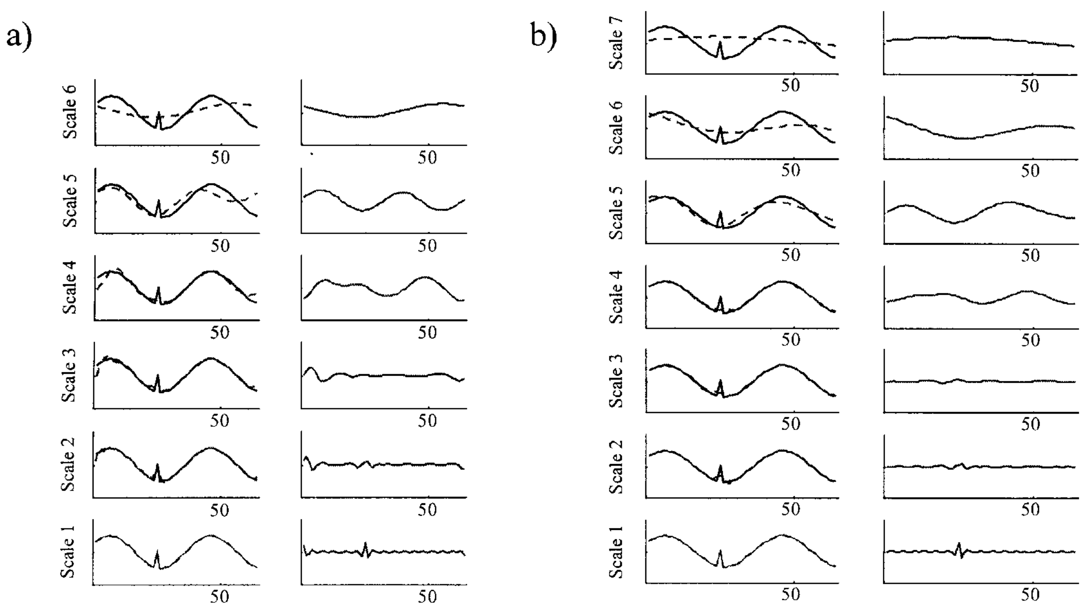

Figure 8. The boundary problem with and without the additional scale using another artificial signal. In part (a) the left-hand column shows the reconstruction of this artificial signal scale by scale without any padding (since the vector already has a dyadic length). The right-hand column shows the corresponding scale-wise detailed signals. These details add up to the original signal. Part (b) presents the corresponding results of level padding with no boundary problems. The additional scale is due to the excessive use of dummy objects, which will ensure error-free performance.

Using MRA as a pretreatment method prior to PLS modeling and monitoring

In the next step, original data were filtered to remove undesired low-frequency disturbances. In this second example, MRA was applied to both the **X** and **Y** blocks. Each original variable was analyzed and filtered separately using MRA, although some of the variables showed no severe drifting or seasonal fluctuation. This is not serious, since those variables that have no or only negligible low-frequency variations have already zero or close-to-zero wavelet coefficients corresponding to these low-frequency scales. In principle, one can choose whether all the variables or only some variables are filtered. This shows importantly the flexibility of this particular approach, i.e. one can pretreat the variables in many different ways. This enables a tuning of the monitoring, which can really be beneficial and result in increased sensitivity towards certain process faults and disturbances.

Again a Symmlet wavelet with 10 vanishing moments was applied. In this example the non-dyadic moving window of 150 objects was used and then vectors were padded to dyadic length using level padding. The length of the window was selected on a somewhat arbitrary basis, although it was ensured that the size was enough to capture seasonal fluctuation from one season to another. After the level padding, MRA was carried out, resulting in eight scales ($2^8 = 256$), the two coarsest of which were ignored. To ensure correct performance, these two scales were checked and it was discovered that the removal of these scales was enough to remove seasonal fluctuations and drifting without disturbing the monitoring of transient high-frequency events. After the padding, the filtered vectors were reconstructed. A removal of dummy objects was carried out before computing the PLS regression. Then filtered variables in both the **X** and **Y** blocks were mean centered and scaled to unit variance before PLS modeling. The corresponding procedure has been sketched in Figure 5(b) and documented in detail in Figure 6(b). Again the number of retained LVs for the calibration set of the first 150 objects was determined using cross-validation, in which a continuous block of 10% of the calibration data was set aside in turn and 90% of the samples were used in modeling. Based on the predictions of the left-out samples in cross-validation, four latent variables were retained. The cross-validation was easy to carry out because it was done in the time domain. To simulate real monitoring, separate test sets were used. These were formed from the remaining 637 objects by computing contiguous non-overlapping moving windows with 150 objects or less. After this the test sets were filtered in the same way as the calibration set and then mean centered and scaled based on the mean and standard deviation vectors of the calibration set.

A conventional PLS model was also computed based on the first 150 objects and compared to the wavelet–PLS approach. The conventional PLS differed from the wavelet–PLS approach in that there was no MRA or filtering. Original data were used as such and only mean centered and scaled to unit variance based on the calibration set's mean and standard deviation vectors. The minimum PRESS value was obtained with seven latent variables for the conventional PLS model.

RESULTS AND DISCUSSION

Exploratory data analysis: scrutinizing PLS models via MRA

The first example briefly explains how a scale-wise interpretation of latent variables can be done and what one can benefit from it. Figure 9(a) shows the score values of the first four latent variables. The other three LVs are not shown here since they did not contain any interesting frequency information, and because it is usually not reliable to analyze the minor LVs. Let us now concentrate on the first four LVs individually. According to the autocorrelograms shown in Figure 9(b), the first LV clearly indicates seasonal fluctuation. Accordingly, scale 8 was associated with this seasonal fluctuation. In the second LV the same scale was also represented, while other higher-frequency scales also contributed to this LV. These new scales were associated with the peaks observed at the score values.

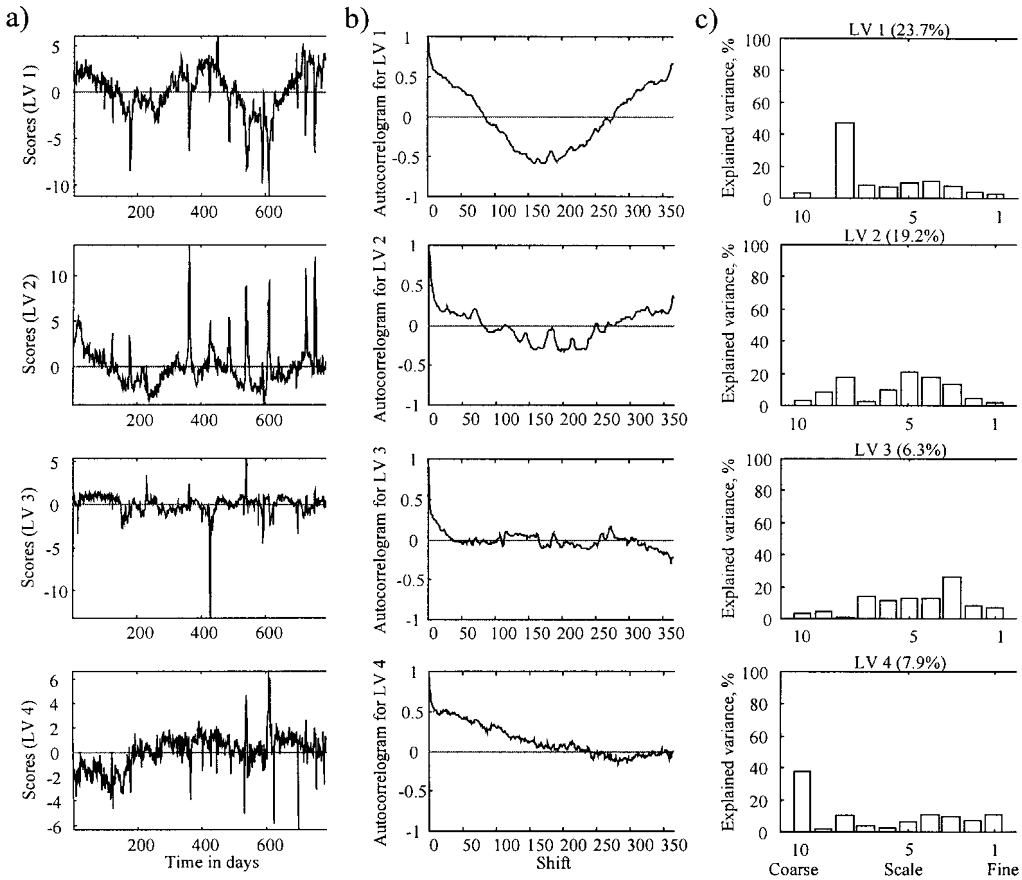

Figure 9. Different latent variables capture different variable intercorrelation structures and events occurring at different scales. The (a) **X** block scores and both the corresponding (b) autocorrelograms and (c) 'scalegrams' [40]. In this case, autocorrelograms indicate roughly what kinds of events are explained by each latent variable (LV), while scalegrams explain in more detail at individual scales what kinds of events are present in each LV. Note that the scalegram is not the same as the common scalogram that is a time–frequency representation of signals.

The third LV lacked almost completely the first three low-frequency scales. The interpretation of this LV is difficult to explain with the scale representation only. Instead, one also needs to look at the loadings and loading weights, which facilitates the interpretation. On the other hand, the fourth LV indicates again a scale-dependent event where the coarsest scale contributes to the LV. This can be again verified from both the score values and the autocorrelogram. Actually, this event was due to a change in the way the process was operated. This change was made after the first winter and it can also be verified from the corresponding score values. As one can see, it is very fruitful to analyze LVs in this way and see how different events occur at different scales and with different rates. In this way it is relatively easy to recognize characteristic time- and scale-dependent features of the process under investigation. The next question is quite obvious, i.e. which variables cause all these time- and scale-dependent events? Interestingly, the same approach can be extended to the original variables as

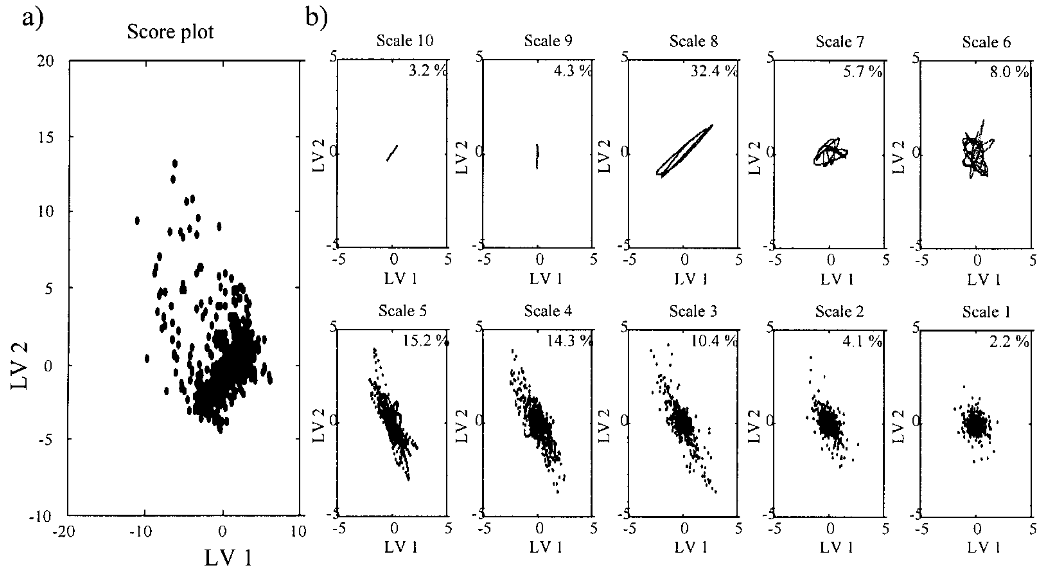

Figure 10. Decomposition of a score plot. Part (a) presents a standard PLS score plot of the first two latent variables. Part (b) gives the corresponding scale-wise representation of the standard score plot. Scale 10 corresponds to the extreme of the low-frequency scales, while scale 1 corresponds to the extreme of the high-frequency scales.

shown in our previous paper [40]. This results in identification of those variables that contribute not only to the latent variables but also to the events extracted by the LVs.

A very popular visualization technique in chemometrics has been that of using score plots or, if superimposed with loadings or loading weights, biplots. Figure 10(a) shows a simple score plot. At a single glance one observes that there is a large dense group of objects elongated towards the upper right-hand corner and a smaller group of objects pointing from this larger group to the upper left-hand corner. A further analysis reveals that the former, the elongation of the dense group, is due to seasonal fluctuation and the latter due to stoppages of the paper plant during 'national holidays'. These variations completely mask the real process faults. While the stoppages are not a problem because they are known to occur at certain points in time, the seasonal fluctuation and the other low-frequency variations are very detrimental. This clearly addresses the issue of detrending, i.e. the removal of these low-frequency events, which is described in the next subsection. As mentioned before, the score plots can reveal interesting things. According to the MRA of the score values, it is also possible to present the score plots at different scales to extend the simple trend plots such as those presented in Figure 9(a). Figure 10(b) shows such a decomposition of the score plot where scales 10–8 correspond to low-frequency variations and seasonal fluctuation (scale 8). Scales 7 and 6 are interesting again in that they point to relatively slow events that vary within a couple of months. These events are believed to result from both the relatively short-term process drifting and the corrective actions made by the operators. The score plots at scales 5–2 are completely masked by the stoppages during 'national holidays'. If the corresponding points are removed, more intersting events are to be seen, such as transient process faults. Scale 1, on the other hand, represents mainly noise, but also very quick changes (peaks, etc.). One should note that, for instance, step changes and peaks are often seen at multiple scales. Importantly, scale 1 also suggests that there is noise present in the first two LVs. This is because the noise is unstructured and hence present in all LVs.

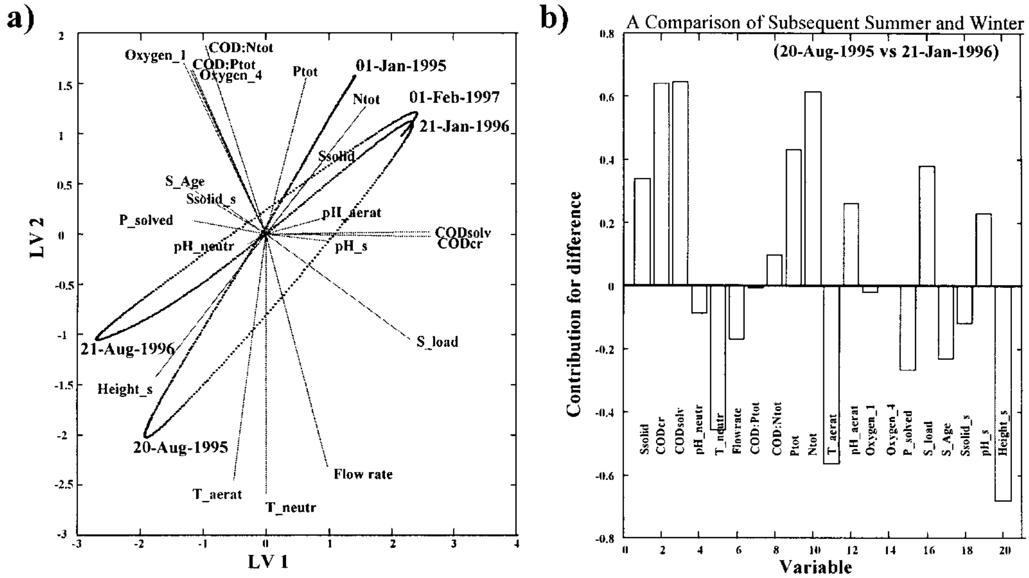

Figure 11. Scrutinizing score plots at different scales. Part (a) is a low-frequency construction of the score plot presented in Figure 10 (a). The reconstruction was based on the three lowest frequency scales. In Part (b) the differences between successive summer and winter have been exemplified by computing contributions for the differences between seasonal inflection points.

In Figure 11(a) a partial reconstruction of the first two LV score vectors is presented. The three coarsest scales, from 10 to 8, were used in the reconstruction. As discussed above, these scales corresponded to seasonal fluctuation and other low-frequency variations. From Figure 11(a) one can isolate seasonal inflection points and compare the differences between them. In Figure 11(b) this kind of comparison was carried out for successive summer and winter. The differences can be investigated by calculating variable contributions as shown in Figure 11(b). Similarly, other scales can be investigated, either individually or cumulatively, by computing the variable contributions for the differences. In this way it would be possible to identify features common to both the **X** and **Y** variables and to improve the product quality or process performance by reducing certain (identified) sources of variation in the **X** variables.

MRA as a pretreatment method prior to PLS modeling and monitoring

In Figure 12, two different approximate multivariate control charts have been computed for both the conventional PLS model and the wavelet–PLS model. Comparison of the two different PLS models indicates that the conventional PLS model fails rapidly right after the calibration set, i.e. objects 1–150 (see Figures 12(a) and 12(b)). This is mainly because these calibration objects are not representative owing to the seasonal fluctuation and other low-frequency variations. The calibration set covers only the first 5 months of 1995, i.e. winter and spring time. Summer and autumn are not represented. Additionally, successive winters differ from the calibration set because of the change in the way the process has been controlled by the operators. All these sources of variation are responsible for the failure of the conventional model. Contrary to the conventional PLS monitoring, Figures 12(c) and 12(d) reveal that the wavelet–PLS approach is not only free from the masking effects of the low-frequency events, but it also seems to be robust in the sense that no model updating

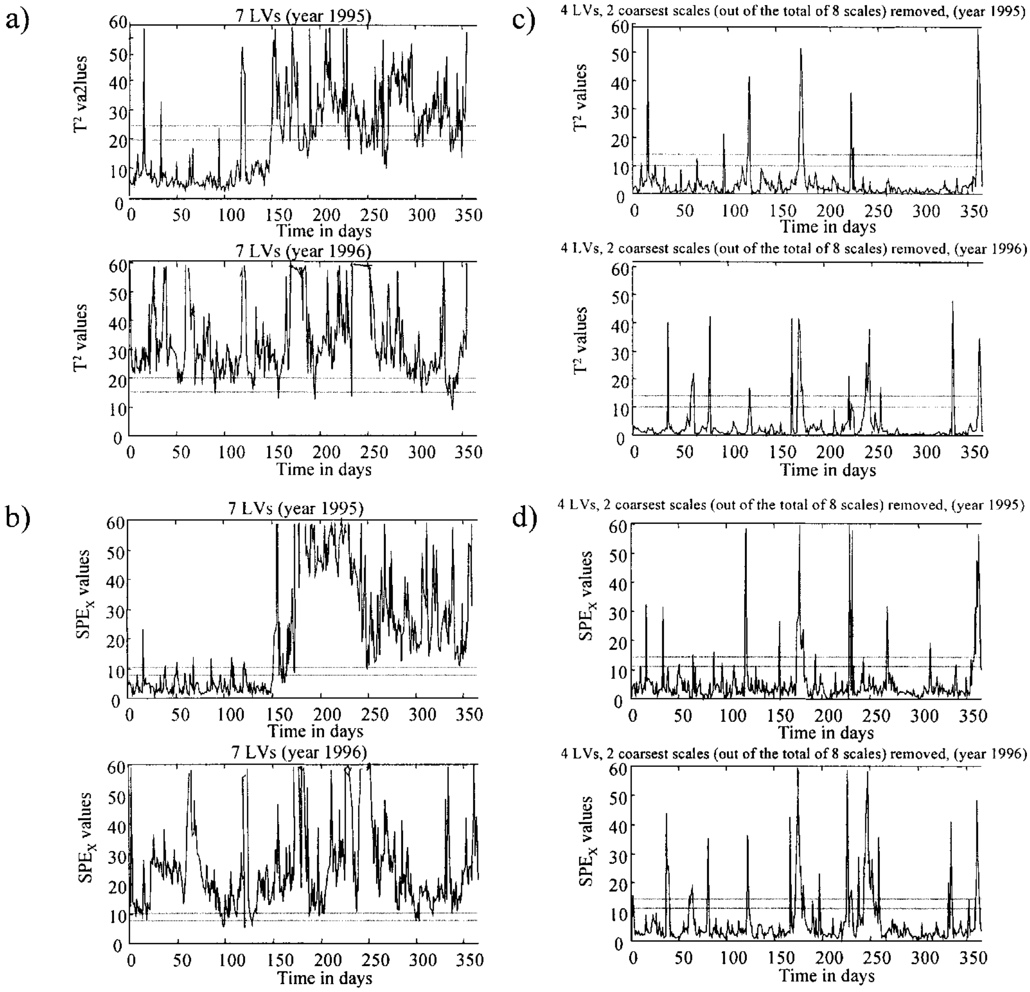

Figure 12. Comparison of monitoring performance of transient high-frequency events. Approximate multivariate statistical control charts for PLS ((a) and (b)) and wavelet–PLS ((c) and (d)) models. The calibration sets consist only of the first 150 objects of 1995. Hotelling's T^2 chart (a) and (c) monitors systematic variations in the latent variable space, while a SPE_X chart ((b) and (d)) monitors squared prediction errors of **X** block variables. A zooming on the *y* axis has been used to give same the *y* scaling (0–60) for every plot. This makes the comparison much easier. Notice that, because of this zooming, some of the peaks exceed the value 60 on the *y* axis.

is needed, at least for this particular data set. This is because the presented approach removes low-frequency fluctuations and results in a more stationary filtered data set that is more suitable for monitoring.

In Figure 13(a) it has been illustrated how a detected and isolated fault can be scrutinized in terms of computing variable contributions from the wavelet–PLS model for the detrended data. A deviant object (177) has been compared to the calibration set. In Figure 13(b), parallel co-ordinates [41–43] have been used to demonstrate one additional but extremely important point in every data analysis, i.e. the confirmation of the conclusions by comparing the results and the original data (in this case the filtered data). The deviant point is due to a process stoppage.

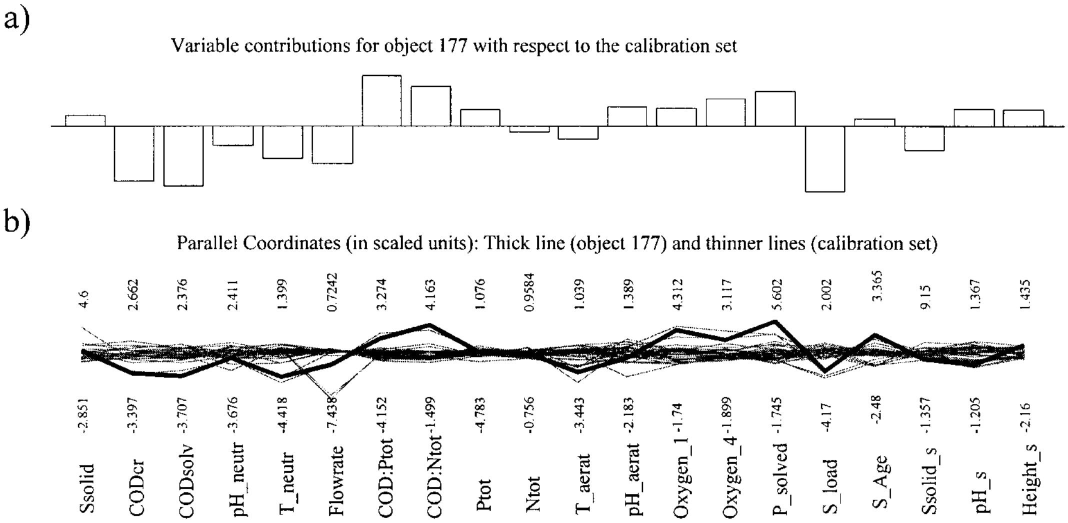

Figure 13. Both identification and visualization of aberrant variables. (a) Variable contributions have been computed for object 177. (b) Parallel co-ordinates are used to visualize the aberrant variables in object 177 compared to normal operating conditions (NOC). In the parallel co-ordinates system the object vector is a line connecting different axes, while each of them corresponds to a different variable. In this way one can quickly identify and visualize the levels of the variables, e.g. scrutinize outliers with respect to NOC.

CONCLUSIONS

In this work, both multiscale and multivariate data analyses were combined to yield improvements in both exploratory data analysis and process monitoring of continuous process data. This was shown by presenting two different approaches. In the first approach a multivariate PLS model was scrutinized using MRA. As a result, low-frequency variations, such as seasonal fluctuation, drifting and changes in the way the process was operated, were revealed and addressed. These variations result in autocorrelated measurements. It was also shown with an artificial signal how autocorrelated measurements can have detrimental effects on process monitoring by broadening the confidence limits and resulting in decreased sensitivity. In the second approach these low-frequency variations were filtered out using MRA. This is necessary since MSPM models should normally consist only of stationary common-cause variation in order to work properly. The comparison between conventional PLS and wavelet–PLS indicated clearly that the former failed to process future data, while the latter did not suffer from low-frequency phenomena. Thus the proposed wavelet–PLS approach was applied to the monitoring of transient short-term faults and disturbances.

To monitor the long-term variations, i.e. the low-frequency scales, one can either use MRA–PLS and instead of removing the low-frequency scales scrutinize these scales or, alternatively, use a PLS–MRA model into which the new data points are added. In MRA–PLS one should note that in principle both scales and individual wavelet coefficients can be pretreated and manipulated in many different ways and thus monitoring can be tuned to be sensitive to different kinds of process faults and disturbances. In order to do this, one has to have comprehension of the different events that can take place in the process and at which scales they are being observed. For this purpose, exploratory data analysis, i.e. PLS–MRA, is of special interest. In MRA–PLS the whole data set can be pretreated similarly or, alternatively, different variables can be treated differently once again to tune the monitoring in order to observe different kinds of events. In a real process environment the parallel use of different kinds of tuned models would benefit in better detecting, isolating and scrutinizing of both high- and low-frequency phenomena. Hence it is strongly believed that multiscale multivariate monitoring remains worthy of further investigation in the future.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to the Graduate School of Chemical Engineering and the Academy of Finland for financial support. The authors also wish to acknowledge both the Southeast Finland Regional Environment Center and Enso Publication Papers Ltd for providing the data.

REFERENCES

1. Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. *Anal. Chim. Acta*, 1986; **185**: 1–17.
2. Jackson JE. *A User's Guide to Principal Components*. Wiley: New York, 1991.
3. Martens H, Naes T. *Multivariate Calibration*. Wiley: New York, 1993.
4. Negiz A, Çinar A. Statistical monitoring of multivariable dynamic processes with state-space models. *AIChE J.* 1997; **43**: 2002–2020.
5. Smilde AK. Three-way analyses. Problems and prospects. *Chemometrics Intell. Lab. Syst.* 1992; **15**: 143–157.
6. Chen G, McAvoy TJ. Predictive on-line monitoring of continuous processes. *J. Process Control* 1998; **8**: 409–420.
7. Wise BM, Gallagher NB. Multi-way analysis in process monitoring and modeling. *AIChE Symp. Ser.* 1997; **316**: 271–274.
8. Nomikos P, MacGregor JF. Multi-way partial least squares in monitoring batch processes. *Chemometrics Intell. Lab. Syst.* 1995; **30**: 97–108.
9. Wangen LE, Kowalski BR. A multiblock partial least squares algorithm for investigating complex chemical systems. *J. Chemometrics* 1988; **3**: 3–20.
10. MacGregor JF, Jaeckle C, Kiparissides C, Koutoudi M. Process monitoring and diagnosis by multiblock PLS methods. *AIChE. J.* 1994; **40**: 826–838.
11. Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemometrics* 1998; **12**: 301–321.
12. Ku W, Storer RH, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics Intell. Lab. Syst.* 1995; **30**: 179–196.
13. Norvilas A, Tatara E, Negiz A, DeCicco J, Çinar A. Monitoring and fault diagnosis of a polymerization reactor by interfacing knowledge-based and multivariate SPM tools. *Proc. Am. Control Conf.*, Philadelphia, PA, June 1998; 3773–3777.
14. Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis. Forecasting and Control*. Prentice-Hall, Englewood Cliffs, NJ, 1994.
15. Teppola P, Mujunen S-P, Minkkinen P. Partial least squares modeling of an activated sludge plant: a case study. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 197–208.
16. Teppola P, Mujunen S-P, Minkkinen P. Kalman filter for updating the coefficients of regression models. A case study from an activated sludge waste-water treatment plant. *Chemometrics Intell. Lab. Syst.* 1999; **45**: 371–384.
17. Teppola P, Minkkinen P. Possibilistic C-means clustering and fuzzy C-means clustering for monitoring of sequential process data in an activated sludge plant. *J. Chemometrics* 1999; **13**: 445–459.
18. Teppola P, Minkkinen P. Empirically defined expert/experience database based on an integration of possibilistic C-means clustering and multivariate process monitoring techniques. *Proc. 6th IAWQ*, Tampere, June 1999.
19. Kolodner J. *Case-based Reasoning*. Morgan Kaufmann: San Mateo, CA, 1993.
20. Gallagher NB, Wise BM, Butler SW, White Jr DD, Barna GG. Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: improving robustness through model updating. *Proc. IFAC ADCHEM*, Banff, June 1997; 78–83.
21. Alsberg BK, Woodward AM, Kell DB. An introduction to wavelet transform for chemometricians: a time–frequency approach. *Chemometrics Intell. Lab. Syst.* 1997; **37**: 215–239.

22. Bakshi BR, Bansal P, Nounou MN. Multiscale rectification of random errors without fundamental process models. *Comput. Chem. Engng* 1997; **21**: S1167–S1172.
23. Bakshi BR. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE J.* 1998; **44**: 1596–1610.
24. Flehmig F, von Watzdorf R, Marquardt W. Identification of trends in process measurements using the wavelet transform. *Comput. Chem. Engng* 1998; **22**: S491–S496.
25. Bakshi BR, Stephanopoulos G. Representation of process trends. III. Multiscale extraction of trends from process data. *Comput. Chem. Engng* 1994; **18**: 267–302.
26. Trygg J, Wold S. PLS regression on wavelet compressed NIR spectra. *Chemometrics Intell. Lab. Syst.* 1998; **42**: 209–220.
27. Hubbard BB. *The World According to Wavelets. The Story of a Mathematical Technique in the Making*. A. K. Peters: Wellesley, 1996.
28. Bakshi BR. Multiscale analysis and modeling using wavelets. *J. Chemometrics* 1999; **13**: 415–434.
29. Bilen C. Wavelet-based detection of outliers. *PhD Thesis*, University of Wyoming, 1998.
30. Luo R, Misra M, Himmelblau DM. Sensor fault detection via multiscale analysis and dynamic PCA. *Ind. Engng Chem. Res.* 1999; **38**: 1489–1495.
31. Starck J-L, Murtagh F, Bijaoui A. *Image Processing and Data Analysis. The Multiscale Approach*. Cambridge University Press: Cambridge, 1998.
32. Morgan D. Haar wavelets? *Embedded Syst. Program.* 1997; **10**(12) [Online]. Available: http://www.embedded.com/archive.htm [July 1999].
33. Nesic Z, Davies M, Dumont G. Paper machine data analysis and compression using wavelets. *TAPPI J.* 1997; **80**: 191–204.
34. Staszewski WJ. Wavelet based compression and feature selection for vibration analysis. *J. Sound Vibr.* 1998; **21**: 735–760.
35. Wise BM, Gallagher NB. *PLS Toolbox Version 1·5 for Use with Matlab™*. Eigenvector Research, Manson, 1997.
36. Wavelab Version. 701. [Online]. Available: http://playfair.stanford.edu/~wavelab [May 2000].
37. *MATLAB Reference Guide*. MathWorks: Natick, MA, 1997.
38. Mujunen S-P, Teppola P, Minkkinen P, Wirkkala R-S. Modeling of activated sludge plants treatment efficiency with PLSR. A process analytical case study. *Chemometrics Intell. Lab. Syst.* 1998; **41**: 83–94.
39. Depczynski U, Jetter K, Molt K, Niemöller A. The fast wavelet transform on compact intervals as a tool in chemometrics. II. Boundary effects, denoising and compression. *Chemometrics Intell. Lab. Syst.* 1999; **49**: 151–161.
40. Teppola P, Minkkinen P. Wavelets for scrutinizing multivariate exploratory models—interpreting models through multiresolution analysis. *J. Chemometrics* 2000; **14**.
41. Inselberg A. The plane with parallel coordinates. *Vis. Comput.* 1985; **1**: 69–91.
42. Inselberg A, Dimsdale B. Parallel coordinates. A tool for visualizing multivariate relations. In *Human–Machine Interactive Systems*, Klinger A (ed.). Plenum: New York, 1991; 199–233.
43. Spiegelmann CH. Plotting aids for multivariate calibration and chemostatistics. *Chemometrics Intell. Lab. Syst.* 1992; **15**: 29–38.

