VTT INFORMATION TECHNOLOGY

RESEARCH REPORT TTE1-2000-19

MODUS-Project
Case Study ODT2

Drive System Monitoring: Requirements and Suggestions

30.6.2000

Esa Rinta-Runsala

VTT

Version history

Version	Date	Author(s)	Reviewer	Description
0.1-1	2.6.2000	E. Rinta-Runsala	kiv	First draft to the steering group
1.0	30.6.2000	E. Rinta-Runsala	kiv	First version

Contact information

Esa Rinta-Runsala
VTT Information Technology
P.O. Box 1201, FIN-02044 VTT, Finland
Street Address: Tekniikantie 4 B, Espoo
Tel. +358 9 4561, fax +358 9 456 6027
Email: Esa.Rinta-Runsala@vtt.fi
Web: http://www.vtt.fi/tte

Last modified on 30 June, 2000

Abstract

A starting point for the case study of a paper machine monitoring is presented along with a requirement analysis. The monitoring is carried out by analyzing drive control measurements and making inferences about the state of the machine based on the measurements and their abnormal changes. The limitations to usefulness of the measurements posed by the properties of the measurement equipment are presented. The demands of different target groups of the monitoring tool and the essential states of the monitoring of the process are also presented in the requirement analysis. Moreover, some attention is paid to the important properties of the paper making process requiring especially good adaptivity and generality from the monitoring method.

The rest of the report is devoted to presenting the monitoring process based on feature extraction from the measurements, feature selection, and inference machine. Additionally theoretical background and properties of some common feature extraction methods, including descriptive statistics, trend analysis, principal component analysis, Fourier analysis, and wavelet analysis, are presented.

Tiivistelmä

Raportissa esitellään paperikoneen monitoroinnin tapaustutkimuksen lähtökohdat ja vaatimusanalyysi. Paperikoneen monitorointi toteutetaan valvomalla koneen telojen sähkökäytöistä tehtäviä mittauksia, ja tekemällä mittauksista päätelmiä koneen tilasta ja sen epänormaaleista muutoksista. Mittauksista saatavaa hyötyä rajoittavat kuitenkin tietyt mittauslaitteiston ominaisuudet, jotka esitellään raportissa. Vaatimusanalyysissä käydään läpi myös monitorointityökalun eri kohderyhmien tarpeet ja prosessin valvonnan kannalta olennaiset tunnistettavat tilat. Lisäksi kiinnitetään huomiota paperikoneen valvonnan kannalta tärkeisiin ominaisuuksiin, jotka vaativat monitorointimenetelmiltä erityisesti hyvää adaptiivisuutta ja yleispätevyyttä.

Raportin loppuosassa esitellään mittauksista tehtävä piirreirroitus ja -valinta sekä päättelykoneeseen perustuva monitorointiprosessi. Lisäksi esitellään joidenkin yleisten piirreirroituksen menetelmien teoreettista taustaa ja ominaisuuksia. Piirreirroituksen menetelmistä käsitellään tilastolliset tunnusluvut, trendianalyysi, pääkomponenttianalyysi, Fourier-analyysi sekä aallokeanalyysi.

Table of Contents

Nomenclature

$t,\ \tau,\ T$	: time variables
ω	: frequency variable
a	: scale variable
σ	: standard deviation
i	: $\sqrt{-1}$
{., ..., .}	: set
$w(.)$	: window function
$g(t,\ \sigma)$	: Gaussian filter
$h_{j,k}(t)$	: wavelet function at time t with parameters j and k
$f(t)*g(t)$	: convolution of $f(t)$ and $g(t)$
$f^{*}(t)$	: complex conjugate of $f(t)$
$\mathbf{X}^{T}$	: transpose of matrix $\mathbf{X}$
$\widetilde{X}$	: projection of X
sign(x)	: sign of x
[x]	: greatest integer less than or equal to x
2D	: 2-dimensional
PC	: principal component
PCA	: principal component analysis
STFT	: short time Fourier transform

1 Introduction

The purpose of this case study is to enhance the monitoring of a paper machine as well as to give some insight for further development of RapidBase active time series database. This report presents the case study, its requirements and limitations. The target is a paper machine whose rolls are controlled by electrical drive controls. The idea to enhance the monitoring is to measure some signals from the drive controls and to draw conclusions about the state of the machine from them. The most important question is how to extract the information from the measurement data for reliable conclusions. A wide range of feature extraction methods is presented and their fitness for the case is considered.

In Chapter 2 the requirements and the special features of the paper machine monitoring are presented. The limitations of the data acquisition methods are also given. Chapter 3 presents some common feature extraction methods as well as their theoretical background. In the last chapter the features are considered in the light of the case study.

2 Requirement Analysis

This requirement analysis is based on interviews of a paper machine drive control supplier, paper machine operators and electric maintenance personnel as well as some technical documents of paper machines and drive controls. The analysis is meant to be a synthesis of the knowledge acquired to guide in the paper machine process state monitoring.

2.1 Presentation of the Case Study

The case study is focused on a paper machine and its drive controls. A typical paper machine consists of sections like wire, press, dryer, coating, calender and reel sections. The sections have multiple drives controlled by speed, torque or tension measurements and the drives in a section could be linked together e.g. with gears or a common felt. All the sections and their controls form a so-called speed reference chain in which the speed reference of a drive is always taken from the former drive i.e. the speed references affect downstream.

The assumption of the study is

> **The measurements of drive controls can help in making inferences about the state of the paper machine.**

There are various ways the drive control measurements can help in making the inferences. In a conventional situation, operator monitors the measurements and makes the inferences by himself. However, the processes of a paper machine are very complicated and they contain a lot of information and measurements. The operator is not able to monitor all the measurements and some kind of data refinement is needed. The refinement can be based on monitoring of the measurements or it can be calculating of some indicators from the signals.

One can also think of a more sophisticated system where the inferences are made by some inference machine. In this kind of system the user gets an announcement only when the inference machine decides the situation is worth mentioning. The goal of the case study is to

1) Study and test suitable methods to refine the information of the paper machine drive controls and

2) Study and test suitable techniques for a machine based inference for paper machine states.

This kind of monitoring system could be used to help to better understand the reasons for abnormal states of a paper machine.

There are three kinds of target groups for the monitoring system. First, the operators of the machine, second, the electric maintenance personnel of the paper mill and third the maintenance personnel of the drive control supplier. These groups need two kinds of information about the paper machine and its states. The operators need real-time monitoring of the process to detect the possible abnormal or undesired states of the process before more severe malfunctions like a paper web breakage occur. On the other hand, the maintenance personnels of the mill and the drive control supplier need long term information about the process behavior to identify the possible causes of faulty situations like a web breakage.

RapidBase active time series database provides a suitable framework for this kind of intelligent data processing. Intelligent tools could retrieve data from the database, calculate the indicators needed, make the inferences, notify the operator if needed and finally save both the indicators and inferences in a database for future use.

In the case study the problem posed is approached both by analyzing process measurement data and collecting expert knowledge. The possible data sources for measurements are a drive control test environment with steel rolls or some real paper machine with appropriate drive controls. The analysis of the data consists of applying of different kinds of feature extraction methods to discriminate process states from each other. Collecting of expert knowledge includes interviews of the process supervisory personnel to identify the situations when the process is in abnormal state.

2.2 Limitations Posed by Data Acquisition

There are several measurements available from the drive controls. Those measurements include speed, torque and temperature of the drives as well as tensions in many places along the paper web. Typically the changes in torque and speed values are in a scale of dozens of milliseconds and they are much faster than the changes in tensions occurring in a scale of hundreds of milliseconds.

The technique used to conduct the measurements limits the sampling frequency to a couple of measurements per second. This sets limitations also to the monitoring of the process states. For example, some fast transient states of the drive torques cannot be detected. Another property of the data collection method is the difficulty to synchronize the observations of different measurements. Due to buffering and asynchronous sampling during the collection process two changes can have the same timestamp although one could have happened hundreds of milliseconds after the other. These facts result the case study to focus on identification and monitoring of slower process changes in scales of seconds instead of milliseconds.

In addition to a slight temporal inaccuracy of the measurements there is also a doubt about the reliability of the values of some speed measurements. Essentially, the speeds mentioned are circumferential speeds of the rolls but they are measured as angular velocity and the circumferential speed is calculated assuming constant roll diameter. The roll diameter, however, is not a constant but is affected by the roll temperature, pressure, and angular velocity. Thus the actual speed of the roll surface along the paper web may differ from the measured one. This introduces problems to inferences based on the speed measurements since the measurements can change though the actual speed would stay the same.

Another kind of a limitation is posed on the real-time availability of the monitoring by the complexity of a paper machine. There are several hundreds or thousands of signals to be measured in the paper machine but only a couple of hundred of them could be monitored at a time because of the restrictions of the monitoring system. In time the scalability of the monitoring system is enhanced but at this moment only a part of the machine can be monitored at a time. Thus, the nature of the measured time series is periodic and there are long periods of missing or incomplete data in the time series.

The last point mentioned here is the availability of so called abnormal data i.e. the data of the abnormal situations like malfunctions or faulty process states. Since the process of papermaking is optimized and enhanced quite thoroughly, there are a little abnormal data available for analyzing. Thus analysis of the data has to concentrate on finding the properties of the normal data and the appropriate means to identify the deviations from the normal situation.

2.3 The Needs of Users

The needs of users are twofold. First, the efficiency requirements and the ways the system should report its detections and second, the important states the system should detect. The following section concentrates on these matters.

2.3.1 Information Obtained from the Monitoring System

A common need for all user groups of a paper machine monitoring system is the comparison between the current state and the past states. Usually an experienced paper mill personnel can identify some problematic states of the process but the process is simply too complicated for a more thorough analysis by human operators. A computer is a suitable tool for this kind of task.

Essentially the monitoring task is detection of abnormal changes. This is true for all the user groups, operators, electric maintenance personnel, and the drive control supplier personnel. The operators need an instant announcement when the process enters a clear and perhaps a predefined undesired state and the maintenance and drive control supplier personnel need information about all the unexpected changes in the process states.

The operators need their information in real-time to prevent more serious states like a paper web breakage. However, the states concerned are not so temporally critical that they should be detected in seconds or milliseconds instead of minutes. As an output to operators besides a notice about the abnormal state the monitoring system should tell the reason of the abnormality i.e. the deviant variables on which the conclusion about the abnormality is based on. This can help the operators to identify the underlying factors of abnormal behavior.

The time requirements of maintenance and supplier personnel are less strict than the ones of the operators. Unlike the operators, the maintenance personnel does not need a real-time monitoring and the detected changes are written into a log file or database for future use. Because the log is mainly used to detect reasons for unexpected states, there should be as many unexpected process state changes as possible. This is quite contrary to the needs of operators when only the most significant and clearly undesired changes are reported. Because the log files are saved long periods for analysis purposes, the data should be as compact as possible. A step towards compactness is the principle to save the measurements of the last minutes before a web breakage with high sampling frequency and use lower sampling frequency for at other times.

The last thing mentioned here is the interpretability of the system output. The reports and the reasons for detection of an abnormal state has to be clear enough to be understood by the users. In practice this means the features used to discriminate the states should be as intuitive as possible.

2.3.2 Monitored Process States

Different parts of the paper machine process are sensitive to different kind of abnormal states. Wire, press, dryer, coater, calender, and reel sections have different characteristics and states. Generally could be said the wet end of the paper machine is harder to monitor with drive control measurements than the dry end. This is because in the wet end there is more external factors affecting the process than in the dry end. In the wet end factors like the quality of the paper pulp cannot be measured from the drive controls and it has to be taken into account in a different way like importing it to the monitoring model from some other system.

In general states with changing measurement values without a respective change in set points should be detected. It should be noted that an altered set point may affect many process variables because of their relatedness. Comparisons between drive controls within a same section or between the measurements of consecutive days could be ways to implement the change detection adaptively.

Wrong load sharing is one of the most important process states to be identified with comparison of torques of drives within a same section. As a matter of fact the load sharing is so important that also the draw profile of the whole paper machine should be monitored. One example of a faulty draw profile is a situation where a reel is drawing so strongly that the previous roll has zero torque. This stretches the paper affecting the quality of the paper and may also result in scission of the paper.

Another general state to be identified is a presence of abnormal frequencies. They may result e.g. from defects in wire or felt or from synchronous oscillations of wipers in a coater. These kind of abnormal frequencies have an effect on the paper quality and they have to be removed as soon as possible.

The monitoring of these changes could be conducted through a so-called control window. The window sets control limits and the operator is notified or an entry is written to a log after an observed value crosses a control limit. Moreover, the control windows should be dynamic in sense that in normal but uncommon states like in closing of a nip the control limits may be looser than normally. However, this requires an external input about when the process is going through some uncommon state.

2.4 Adaptivity Requirements

The paper machine is a very complex system as stated earlier. In addition to huge number of factors concerned, the process changes almost constantly. The mechanical parts of the machine wear out and they are replaced and motors are warming. Moreover, the process line can be changed by adding, changing or removing some sections of the line like dryers, coaters or nips. These changing circumstances make a paper machine quite a challenging system to monitor. The unavailability of abnormal data is mentioned in Section 2.2 and the changing circumstances guarantee there is very seldom or never enough data for estimation of very complex models of a paper machine. On the other hand, the model would soon be obsolete because of the changing nature of the process.

Basically the process can change in three ways. First, short term changes like motor warming after runup or fluctuations due to day and night temperature changes. Adaptation needed for these kind of changes could be realized by some sort of calibration of the monitoring model after an unstable period. The calibration could be operator initialized or due to some sort of rule that calibrates the model when the process has been stable long enough. A disadvantage of a rule-based calibration is possibility of a persistent and stable undesired state which could be used as a calibration reference by accident. In the best case the monitoring model takes the slow transient states into account but this is not always possible.

The second kind of change is a long term change resulting from a mechanical wear-out of machine parts. Wear-outs have two sides. On the other hand, normal wear-out should be tolerated and adapted into but severe wear out should be notified for a replacement of the machine part. The adaptation to slow changes could be achieved by comparing measurements to values on the day before. However, this solution is not acceptable if there has also been a change between the days e.g. due to change of a paper type or some other instantaneous change mentioned in the next paragraph.

The third kind of change is instantaneous in its nature. "Instantaneous" changes are due to changes in a paper type, set points of drive parameters, replacements of the outworn machine parts or changes in the process line. These changes are easy to adapt with aid of external indicator variables which tell there has been an instantaneous change in the process. In case indicator variables cannot be used the operator could receive a false message about the detected abnormal state. However, after instantaneouts change the operator can easily determine the detection is a result of changed process circumstances and it could be ignored.

An especially difficult situation is confronted when an instantaneous change in circumstances results in a totally different and new process state. The challenge is to apply a reasonable method to decide if the new state is normal or not. One solution is to use heuristic rules for the decision making but the actual normality or abnormality of the new state has to be left to the user to avoid the danger of adaptation to an abnormal state.

The generality of the monitoring model, i.e. the ability to monitor different sections of the paper machine with different input variables and state identification requirements, is also one kind of adaptivity. The different sections of paper machine are sensitive for different undesired states as stated in Section 2.3 and the identification of different states may also need different measurements. Moreover, the mill personnel may like to monitor different values in different mills. These generality requirements set high demands especially for the state identification model behind the inferences. The model should be flexible enough to accept varying number of input variables with different types of variables like differences, deviations, or frequency components.

All these situations give rise to an increased need for adaptivity of the intelligent monitoring system and they should be considered carefully when deciding which methods are used for feature extraction and for intelligent inference machine.

3 The Monitoring Procedure

Computer based monitoring and state identification consists of several phases between the raw process data and the final inference about the state of the process. This procedure is illustrated in Figure 1.

The first step is the measuring of the process. There are often many possible variables one can measure so the first decision is to select the most appropriate ones for the state identification. The selected variables are called measurement variables. Next the continuous process signals are sampled by taking measurements of them in some specific frequency resulting in the measurements. The signals and the respective measurements are assumed to change according to the underlying process state and a certain type of measurements are assumed to be characteristic to a certain process state. However, the characteristic features of the process are not always easy to distinguish from the measurements with background noise, measurement error or some insignificant factors disturbing the measuring. Thus one has to refine the data somehow to get a grip on the useful characteristic features.

Additionally, the reasonable number of measured variables to be used in monitoring is limited. A human observer cannot monitor and interpret more than a few measurement variables at a time. A computer, on the other hand, is able to monitor a large number of variables but even an artificial intelligence cannot make reliable inferences from a large number of measurement variables without a vast amount of history data to base the inferences on. The growing need of data due to the growing number of measured variables is a phenomenon called "curse of dimensionality" [Bellman 1961] and it is long known in many fields of science. The term dimensionality refers to an idea that every measurement variable can be thought to form one dimension of a many dimensional measurement space and the process states, i.e. the measurements at a specific time instant, are thought as a point in the measurement space. The curse of dimensionality is the major reason for reducing the number of variables used to monitor the process.

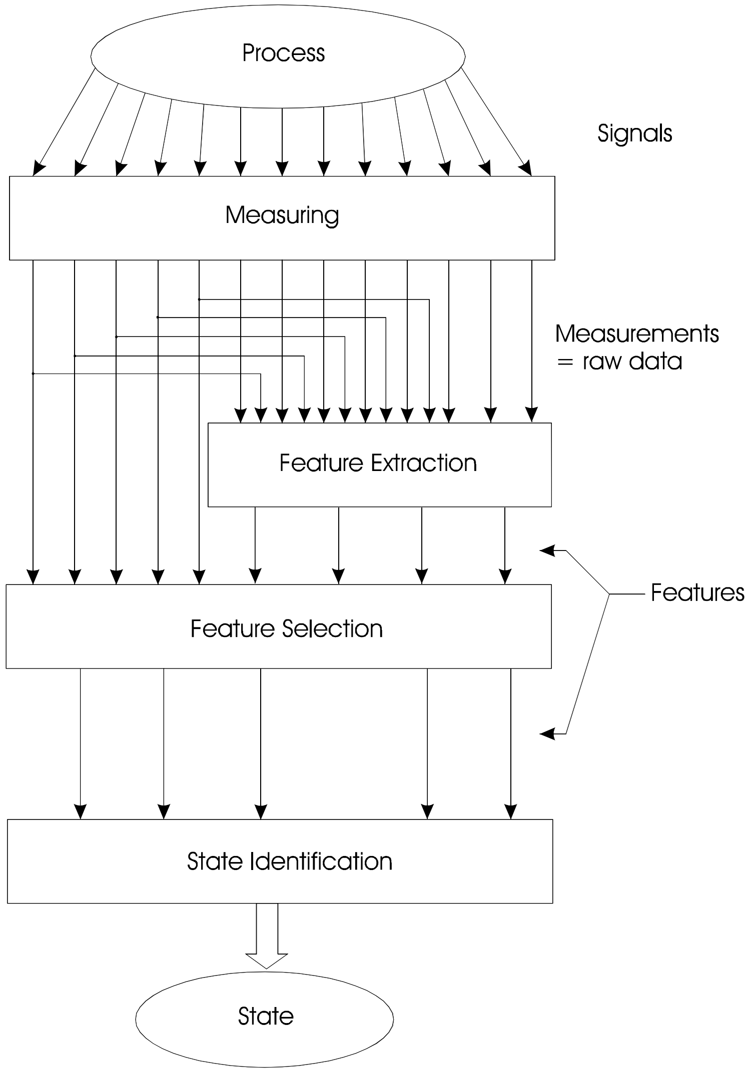

Figure 1. The process of monitoring and state identification.

In identification or recognition tasks the variables used to discriminate the process states from each other are called features. The term feature is used here to refer to a very wide set of variables. A feature can be e.g. a single measurement, difference of two signals, power of a frequency band, or a figure based on more complex calculations on one or more measurements.

There are basically two methods for reducing the number of variables and improving their interpretability: feature extraction and feature selection [Pudil & Novovičová 1998]. Feature extraction is a method to transform the original variables into other preferably more descriptive ones. Some feature extraction methods are described in Chapter 4. On the other hand, feature selection is essentially selecting of a subset of the original feature set. For discrimination of the states the features are selected based on their discriminative value.

The final step in monitoring is the state identification based on the extracted and selected features. The states mentioned here can be so simple as normal and abnormal states, or they can more specific ones e.g. a slip of the paper, too worn felt, or faulty drive. The identification, i.e. inference about the state, can be done by several techniques like neural networks, fuzzy rules, or statistical analysis. However, the properties of artificial intelligent inference techniques are beyond the scope of this report.

4 Feature Extraction Methods

This chapter presents some common feature extraction methods including descriptive statistics, trend, Fourier and wavelet analysis and their theoretical aspects.

4.1 Descriptive Statistics

The descriptive statistics such as mean (i.e. arithmetic mean), median, standard deviation, kurtosis, and skewness are perhaps the easiest way to monitor a process. Calculating these statistics is elementary and their interpretation is straightforward. Change in a sample mean tells that the process is not in a stable state and an increased standard deviation or maximum deviation from the mean may indicate disturbances in the process. Kurtosis and skewness are a little bit more difficult to interpret and they are less used than the mean or standard deviation.

The practice of using charts called Shewhart control charts for monitoring of development of descriptive statistics dates back to the 20s and they are used especially in quality control [Grant & Leavenworth 1996]. Monitoring with the Shewhart chart is simple. Based on process knowledge some control limits are specified and sample means or other statistics to be controled are drawn in the same chart with control limits as points of time series (Figure 2). If the measured statistics crosses a control limit, the process is out of control.

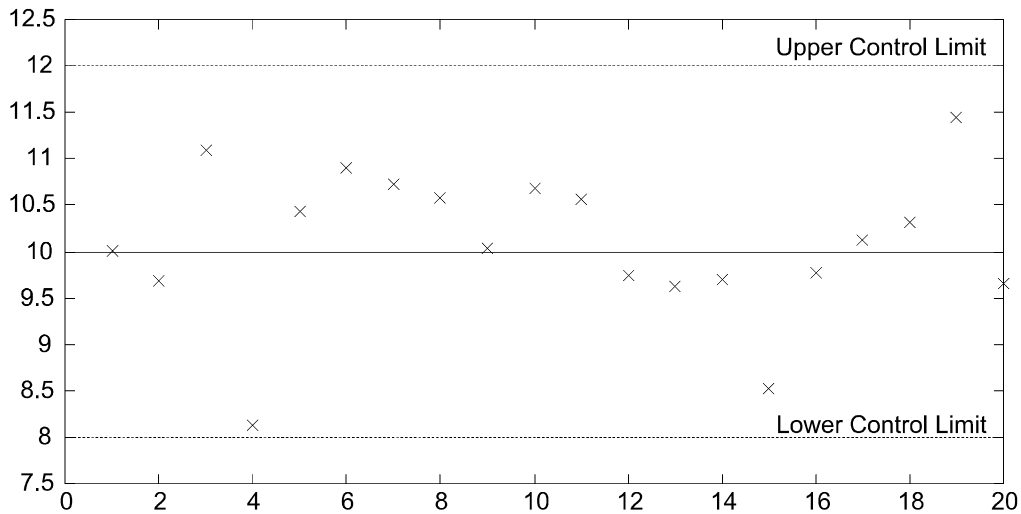

Figure 2. Example of Shewart control chart with control limits 12 and 8. The process is in control for the whole period of study.

Descriptive statistics are still valuable features in monitoring, but they are inadequate measures of some temporal characteristics like periodicity. Thus more sophisticated methods for feature extraction are needed.

4.2 Trend Analysis

Trends are perhaps the most intuitive features of time series. A human observer can distinguish upward or downward trends from a graphic presentation of noisy data quite easily, but the task is much more difficult for a computer.

The first thing is to define a trend. In this report trend is a pattern or structure in 1-dimensional data. The definition is adapted from [Kivikunnas 1999]. Since the task of finding trends in data is quite fundamental there are quite many methods for trend analysis. Some of them, including regression analysis, triangular representation, dynamic time warping, and wavelet-based methods are presented in [Kivikunnas 1999]. This section is focused on the triangular representation method [Cheung & Stephanopoulos 1990a], which is the basis for a couple of more sophisticated trend analysis techniques [Bakshi & Stephanopoulos 1994a], [Vedam & Venkatasubramanian 1997]. Additionally some general notions about trends are made.

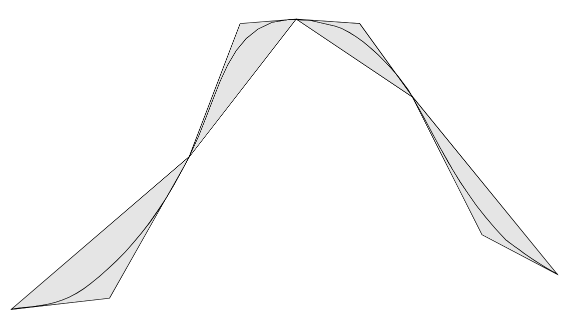

Figure 3. Example of process trend and triangular representation.

At its simplest a trend can be upward, downward, or constant. Basically this means observing the first derivative of the underlying function. In a more thorough analysis also derivatives of a higher degree could be taken into account. In so-called triangular episode representation the first and the second derivatives are calculated and an episode is defined as a time interval in which the signs of derivatives remain constant [Cheung & Stephanopoulos 1990a]. An example of such episodes is in Figure 3. Altogether, the variation of derivative signs gives seven basic types of triangular episodes called primitives presented in Figure 4. The time series is represented as a series of these primitives and trends are described by them.

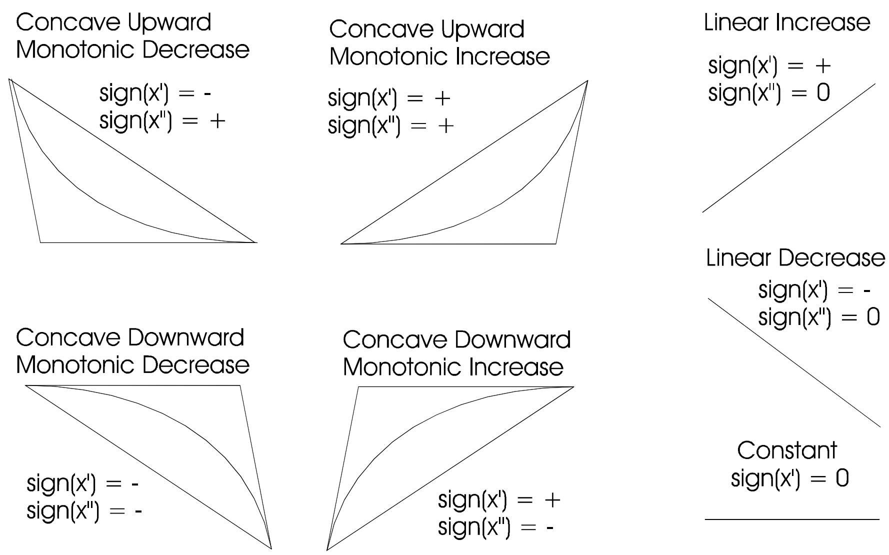

Figure 4. Primitives of triangular representation.

The process signal is a result of many underlying phenomena [Cheung & Stephanopoulos 1990b] as illustrated in Figure 5. Each of these phenomena acts on their own scale, in the case of the figure 5 the effect of noise is smaller than the effect of periodic disturbance and on the other hand the effect of malfunction is greater than the effect of disturbance. To be able to concentrate on interesting phenomena without disturbing details the analyzer has to smoothe the signal appropriately. However, smoothening may also destroy or distort interesting properties of the signal. For example, smoothening may flatten the peak caused by the malfunction in figure 5 enough for the malfunction not to be detected.

This problem is especially relevant in trend analysis and the decision of an appropriate scale of observation is crucial. If the scale is too small, the trends cannot be distinguished from insignificant noise and on too large a scale the trends are flattened out. The appropriate scale and number of details is totally case dependent. The smoothing is conventionally done with filtering or curve fitting but they have some disadvantages.

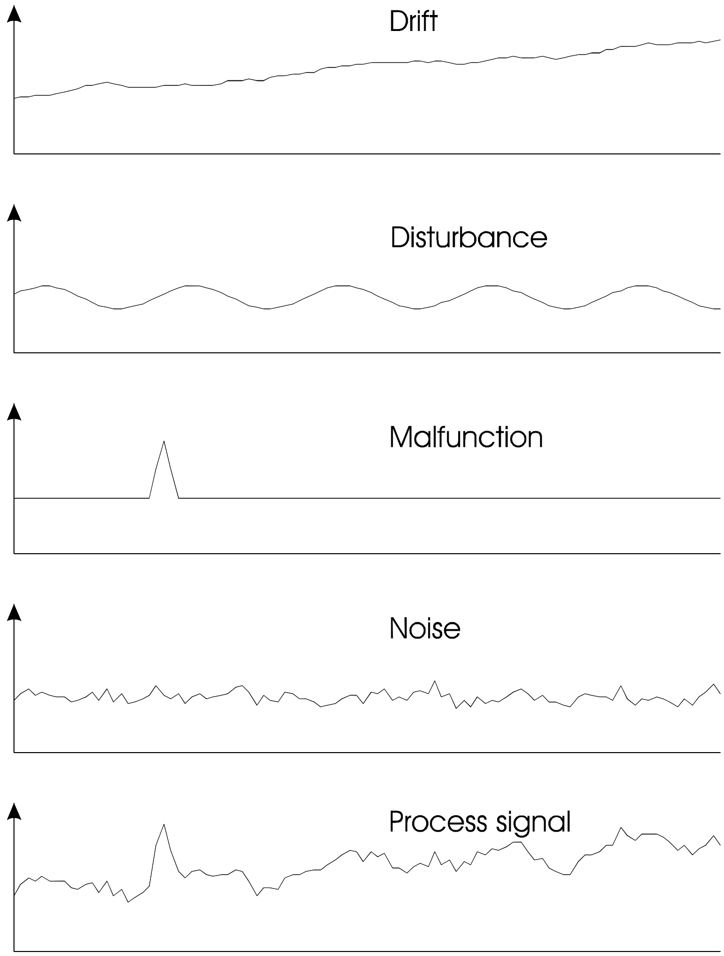

Figure 5. Some components of a process signal.

The problem of scaling is presented in Figure 6 with different smoothings with Gaussian filter. The Gaussian filter is chosen because it is the only filter resulting hierarchical structure of a scale for any input [Babaud et al. 1986]. In this context hierarchical means that the inflexion points of a higher scale are persistent in lower scales.

However, smoothing of the data to different scales is computationally quite intensive since it is essentially calculating of the convolution $F(t,\sigma)$ [Cheung & Stephanopoulos 1990b]

$$F(t,\sigma) = f(t) * g(t,\sigma) = \int_{-\infty}^{\infty} f(\tau) \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\tau)^2/2\sigma^2} d\tau \qquad (1)$$

where $f(t)$ is the signal and

$$g(t,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2} \qquad (2)$$

is the Gaussian filter with a standard deviation σ. The problem of finding a suitable scale for trend analysis is thoroughly considered in [Cheung & Stephanopoulos 1990b].

Triangular episodes as features of a process have a couple of good properties. Firstly, they are a compact way to describe the process behavior since only the important information of the trend is preserved. Secondly, the episodes have a clear intuitive meaning for an analyzer. The behavior of the process can be described also linguistically based on triangular episodes [Cheung & Stephanopoulos 1990a]. For example, a trend with the first derivative positive and the second derivative negative could be said to approach some value.

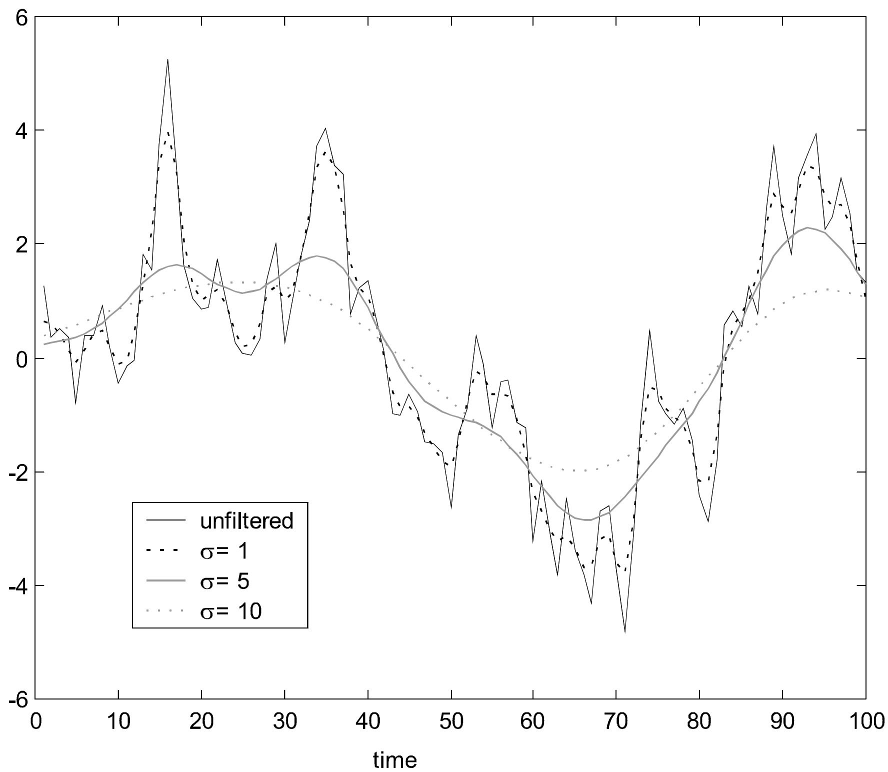

Figure 6. Signal filtered with Gaussian filters and variable standard deviations σ. *The number of peaks and valleys varies considerably as a function of* σ.

Due to the computational intensiveness of the previous approaches in multiscale trend analysis several alternative methods are applied to trend analysis. In [Bakshi & Stephanopoulos 1994a], [Vedam & Venkatasubramanian 1997] and [Flehming et al. 1998] trend analysis is approached with an aid of wavelets. [Bakshi & Stephanopoulos 1994a] use wavelets to identify interesting trends on different scales resulting on a filtered signal with the most stable trends of different scales. [Vedam & Venkatasubramanian 1997] constructed a trend identifier neural network which uses wavelet representation of the signal to classify a trend into primitive classes of Figure 4. On the other hand [Flehming et al. 1998] used wavelets as denoising tool and fitted low order polynomials on the resulting data.

4.3 Principal Component Analysis

Principal component analysis (PCA) is a technique to form new variables from multidimensional data in such a way the new variables are uncorrelated [Jackson 1991]. An important consequence of this fact is that most of the information, i.e. variation, is preserved in fewer variables [Bishop 1995]. Consider an n–component vector $\mathbf{X}$ of random variables $X_1, X_2, \ldots, X_n$ in n-dimensional space. Now one can form new variables Y_k e.g. as a linear combination of the old variables $X_1, X_2, \ldots, X_n$:

$$Y_k = \sum_{j=1}^{n} c_{kj} X_j = \mathbf{c_k X} \qquad 1 \le k \le m \tag{3}$$

where c_{kj} are the coefficients of the variables X_j, $\mathbf{c_k} = (c_{k1}, c_{k2}, \ldots, c_{kn})$ and m is the number of new variables. Writing the Equation (3) into a matrix form gives

$$\mathbf{Y} = \mathbf{CX} \tag{4}$$

where $\mathbf{Y}$ is an $m \times 1$ vector of new variables and $\mathbf{C}$ is an $m \times n$ matrix of coefficients c_{kj}, $1 \le k \le m$, $1 \le j \le n$. If $m = n$ and the matrix $\mathbf{C}$ is orthonormal, the resulting coordinate system is just a rotation of the old one [Bishop 1995]. In this kind of linear transformation, all the information of the old variables X_j is preserved.

As stated, however, most of the information could be preserved also with fewer new variables. For example, if observations X_j are in a 3-dimensional space, but they are organized in a plane, the third dimension would be completely redundant and the observations could be mapped into a 2-dimensional plane without loss of information. Although there usually are no completely redundant dimensions in data, the observations can be mapped into a lower dimensional space preserving most of the information. The transformation from an n-dimensional space into an m-dimensional one is done by multiplying the original observations X_j by an $m \times n$ transformation matrix $\mathbf{C}$ where $m < n$. The amount of information preserved is determined by the matrix $\mathbf{C}$. Choosing of $\mathbf{C}$ is an optimization problem to be considered next.

Essentially, the process of reducing the dimensionality of the data is projecting of the data points to another coordinate system. The amount of information lost during the reduction process can be understood as a distance between the original data point and the projected one. Thus, a conventional way to choose the matrix $\mathbf{C}$ is to minimize the sum of squares of differences over the data set. Consider a case where an n-dimensional vector $\mathbf{X}$ is decomposed to two sums with inverse of an $n \times n$ transformation matrix $\mathbf{C}$, the first sum representing the projection to principal components and the second one the residual

$$\mathbf{X} = \mathbf{K^T Y} = \sum_{j=1}^{m} Y_j \mathbf{c_j} + \sum_{j=m+1}^{n} Y_j \mathbf{c_j} \qquad m < n \tag{5}$$

In principal component analysis with less than n principal components, the residuals Y_j, $m+1 \le j \le n$, are replaced with constants b_j.

$$\widetilde{\mathbf{X}} = \sum_{j=1}^{m} Y_j \mathbf{c_j} + \sum_{j=m+1}^{n} b_j \mathbf{c_j} \qquad m < n \tag{6}$$

Now the error E between the projection $\tilde{\mathbf{X}}$ and the original $\mathbf{X}$ is

$$E = \sum_{j=m+1}^{n} \left(Y_j - b_j\right)\mathbf{c_j} \tag{7}$$

Squaring (7) and summing over all the observations give the error of PCA associated to the whole data set. Minimization of the whole error with respect to b_j results finally to the solution that the principal components (PC), i.e. the vectors $\mathbf{c_k}$ or the rows of the transformation matrix $\mathbf{C}$, are eigenvectors of the covariance matrix of the observations [Bishop 1995]. It can be shown that the same result can be achieved as a singular value decomposition of the original data matrix or as a spectral decomposition of a covariance matrix [Sharma 1996].

Graphically the principal components are orthogonal (i.e. perpendicular) to each other. Additionally, the first one takes the direction of the largest variation of the data, the second one the direction of the largest residual variation and so on. A 2D example of this is presented in Figure 7. The new variables Y_k called principal component scores are projections to the PCs as seen in the figure.

There are many uses for PCA in multivariate analysis. The usual use of PCA is as a dimension reduction technique to avoid the curse of dimensionality. In this case the objective is to find $m<<n$ PCs that contain as much information from the original data as possible without losing relevant information. In PCA the loss of information is measured by amount of unexplained variance. Usually the smallest PCs are considered as noise but there is no single method to determine the number of relevant PCs [Jackson 1991] and the appropriate number is case-specific.

PCA could also help in the interpretation of data. By looking at the correlations between original variables and PCs, one gets an idea which variables are influential in which components. This way one can assign indicative labels to the PCs to help in interpreting the principal component scores [Sharma 1996].

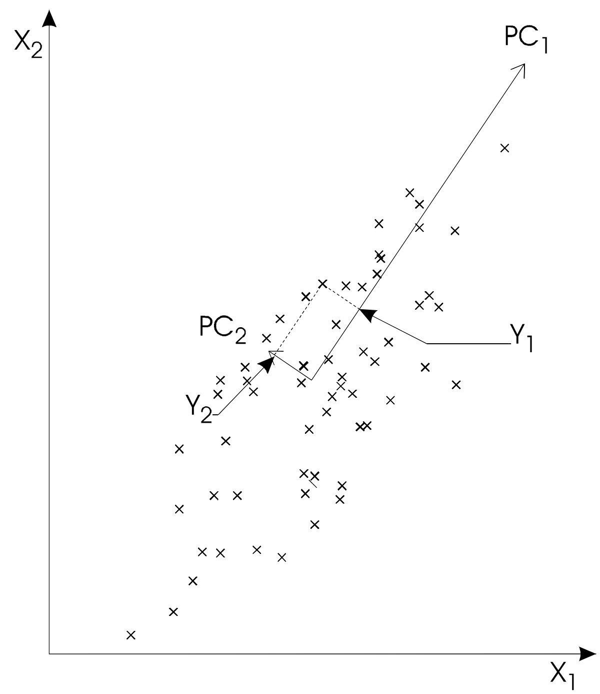

Figure 7. Example of a 2-dimensional data and its principal components (PC_1 and PC_2).Y_1 and Y_2 are the principal component scores i.e. projections of data points to PCs.

There are a couple of matters to consider when deciding whether to use PCA or not. Firstly, there has to be some covariance in the data for PCA to be appropriate. Without covariance the old variables are independent and PCA should not be used [Jackson 1991]. Additionally, depending on the case the reduction of dimensions with PCA may result in substantial loss information [Sharma 1996]. If the final goal of the analysis is to classify the states, the analyzer has to take seriously the danger of losing discriminating dimensions during the dimension reduction [Bishop 1995].

Traditional way to conduct monitoring with PCA is to monitor the process by calculating the residuals or errors associated with PCA [Jackson & Mudholkar 1979]. If the error exceeds a predefined confidence limit, the process is assumed to be out of control i.e. in a faulty state.

The disadvantage of this kind of use of PCA is that the observer assumes the PCs (i.e. covariance matrix) stay the same during the monitoring. This condition is quite limiting in many industrial applications. Despite of this fact PCA has become a popular method in statistical process control and it has been applied to variety of real world problems [Kourti et al. 1996].

4.4 Fourier Analysis

Fourier analysis is a way to represent a function in a frequency domain. In Fourier analysis, function *x(t)* is represented as a superposition of sines and cosines or complex exponentials. The theory was originally developed for continuous functions but the theory is later on expanded to discrete phenomena [Wei 1994]. Since the real world measurements result in a discrete time series the following explanation is focused on discrete analysis.

Consider trigonometric sine and cosine functions sin($2\pi kt/n$) and cos($2\pi kt/n$) defined in a finite number of n points with $t = 1,2,...,n$. With $k = 0,1,2,...,[n/2]$, where $[x]$ is the greatest integer less than or equal to x, one can define a system

$$\{\sin(2\pi kt/n), \cos(2\pi kt/n) : k = 0,1,...,[n/2]\} \tag{8}$$

The System (8) contains exactly n nonzero functions and additionally it is a collection of orthogonal functions [Wei 1994]. A set of vectors is called a basis if every vector in the space can be represented as a linear combination of the vectors of the basis. From vector analysis it is known that in an n-dimensional space every set of n orthogonal vectors forms a basis. Thus also the set (8) forms a basis and any given sequence of n numbers $x(t)$ can be written as a linear combination of its elements

$$x(t) = \sum_{k=0}^{[n/2]} \left(a_k \cos(2\pi kt/n) + b_k \sin(2\pi kt/n) \right) \qquad t = 1,2,...,n \tag{9}$$

The Equation (9) is known as a Fourier series of $x(t)$ and a_k and b_k are Fourier coefficients calculated from

$$a_k = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{t=1}^{n} x(t)\cos(2\pi kt/n), & k = 0 \text{ and } k = n/2 \text{ if } n \text{ is even,} \\ \dfrac{2}{n}\displaystyle\sum_{t=1}^{n} x(t)\cos(2\pi kt/n), & k = 1,2,...,\left[\dfrac{n-1}{2}\right] \end{cases}$$
$$b_k = \frac{2}{n}\sum_{t=1}^{n} x(t)\sin(2\pi kt/n), \qquad k = 1,2,...,\left[\frac{n-1}{2}\right] \tag{10}$$

The Fourier coefficients tell the relative power of a frequency $\omega_k = 2\pi k/n$ and they can be considered as a similarity measure between the signal and the base functions (i.e. sines and cosines) [Rioul & Vetterli 1991].

The main disadvantage of Fourier analysis is its lack of ability to distinguish time variant characteristics of the sequence. The Fourier coefficients are calculated for the whole sequence and the decomposition to everlasting waves destroys all the time information in the signal. Thus the analysis of non-stationary signals and detection of transient phenomena require more than the ordinary Fourier transform.

One way to overcome this problem is Short Time Fourier Transform (STFT). The idea of STFT is to take Fourier transform in a neighborhood of each time sample thus mapping the signal into a time-frequency plane with a fixed resolution. In practice STFT is done by limiting the Fourier transform to a short time section by an analysis window [Nawab & Quatieri 1988]. The exact expression for discrete STFT at time t_0 and frequency ω is given by

$$X(t_0, \omega) = \sum_{n=-\infty}^{\infty} x(t) w(t_0 - t) e^{-i\omega n} \tag{11}$$

where *x(t)* is the signal and $w(t_0-t)$ is the analysis window. A typical analysis window for STFT is much shorter than the duration of the signal *x(t)*. The use of an analysis window is illustrated in Figure 8.

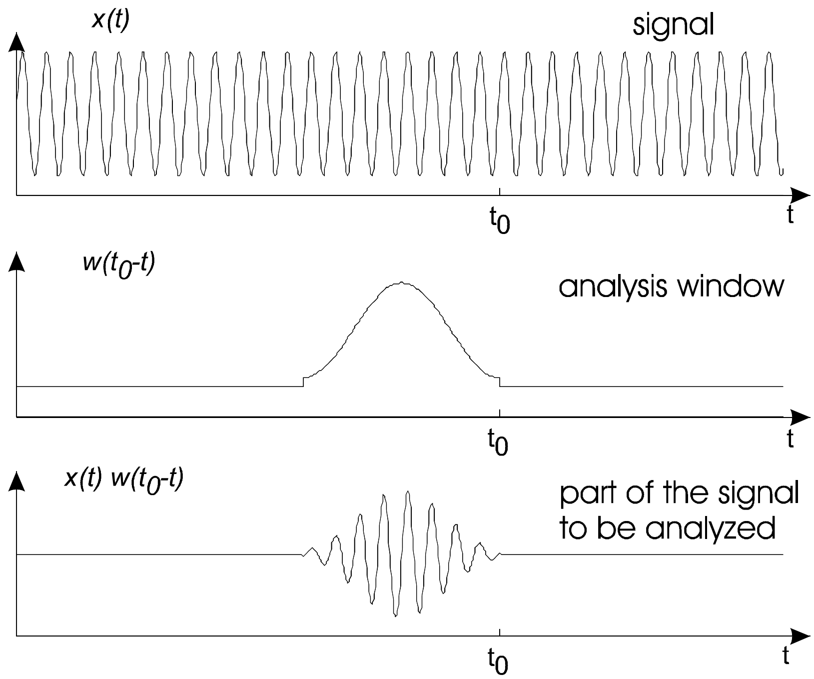

Figure 8. Use of an analysis window to limit the studied time period for STFT.

It is clear from the definition of STFT that the choice of the analysis window has an effect on the results of STFT. Both the form and the length of the window $w(t_0-t)$ affect the accuracy of the frequency representation. Due to these factors the question of the length of the window is quite fundamental. Since the goal of STFT is to get a better understanding of the time variant properties of the signal, one could try to use as short a window as possible to get as high a time resolution as possible. However, a shorter window also means less data for the Fourier analysis and this implies lower accuracy in the frequency representation. Thus, the choice of the window length is essentially a question of trade-off between time and frequency resolutions of STFT [Nawab & Quatieri 1988]. This connection is known in literature as Heisenberg-Gabor uncertainty principle [Flandrin 1999] and its consequence is that STFT does not solve the needs of the monitoring of non-stationary signals completely satisfactorily. Wavelet analysis is one solution to this problem and it is studied in more detail in Section 4.5. However, in cases where the resolution needed is known in beforehand and the scale of the study is the same for every phenomenon interested, STFT is a sufficient tool.

4.5 Wavelet Analysis

Fourier analysis is a method to find frequency characteristics of a signal over the whole period of observation. Short time Fourier transform is an attempt to introduce temporal aspects to the signal analysis. Wavelet analysis is another way to discover temporal characteristics of the signal and it has many advantages over STFT [Daubechies 1990], [Rioul & Vetterli 1991] & [Bakshi & Stephanopoulos 1994a].

The basic idea behind the wavelet analysis is to decompose the signal to a family of functions with a vanishing mean value [Daubechies 1990]. Let *h(t)* be some function with a vanishing mean value called a mother wavelet. This means *h(t)* is localized on some period of time. Now one can define a family of wavelets by shifting and scaling the mother wavelet:

$$h_{\tau,a}(t) = \frac{1}{\sqrt{a}} h\left(\frac{t-\tau}{a}\right) \qquad (12)$$

where τ is a parameter for shifting of the function and *a* is a scaling factor. With varying τ and *a,* the wavelet family covers the whole time-scale (time-frequency) plane. Since the local frequency of a wavelet depends on both the mother wavelet *h(t)* and its scaling, the term time-scale plane is preferred to time-frequency plane used with STFT [Rioul & Vetterli 1991]. Some shifted (translated) and scaled (dilated) wavelets are drawn in Figure 9 to clarify the concepts of scaling and shifting.

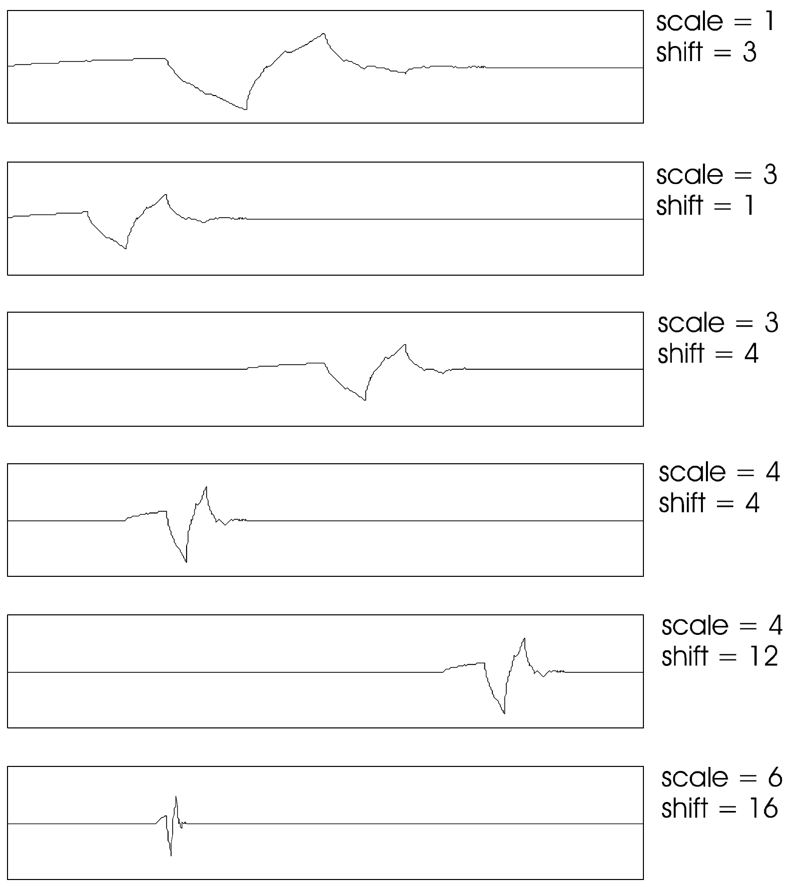

Figure 9. Daubechies-4 wavelets with different scalings and shifts (picture drawn with aid of WaveLab 802 library for Matlab [Donoho et al 1999]).

For real world discrete time signals, the time-scale plane of wavelet analysis has to be discretized. A simple choice for discretization of a scale is taking powers of a fixed dilation step $a_0>1$, $a=a_0^j$, with integer *j*. Now it is convenient to define the shifts τ relative to the scale of the wavelet: $\tau = k a_0^j \tau_0$, with integer k [Daubechies 1990]. These choices result to a discrete wavelet family

$$h_{j,k}(t) = a_0^{-j/2} h\left(a_0^{-j} t - k\tau_0\right) \tag{13}$$

With suitable parameters a_0, τ_0 and a mother wavelet $h(t)$ the discrete wavelet family $h_{j,k}(t)$ forms an orthogonal basis with some useful properties [Mallat 1989] and once again an arbitrary signal $x(t)$ can be represented exactly in this new basis as a superposition of wavelet functions [Chui 1992]:

$$x(t) = \sum_{j,k} c_{j,k} h_{j,k}(t) \tag{14}$$

where $c_{j,k}$ are wavelet coefficients calculated from

$$c_{j,k} = \int x(t) h_{j,k}^{\ *}(t) dt \tag{15}$$

where $h^*(t)$ denotes the complex conjugate of the function $h(t)$. As in Fourier analysis the coefficients $c_{j,k}$ can be understood as similarity measures between $x(t)$ and the wavelet functions $h_{j,k}(t)$ [Rioul & Vetterli 1991] and they can be used in pattern recognition as features describing the properties of the signal [Pittner & Kamarthi 1999]. Some other uses of wavelets in pattern recognition can be found in [Wu & Du 1996], [Chen et al. 1999] and [Liu & Ling 1999].

The power of wavelets is their ability to increase the resolution through scaling. By controling scaling and shifting the signal can be examined at different resolutions, an ability called multi-resolution [Mallat 1989]. This advantage over Fourier and short time Fourier analysis is clearly seen in Figure 10. In Fourier analysis, there is no temporal resolution and in short time Fourier analysis the resolution is fixed for the whole time-frequency plane. On the other hand in wavelet analysis the high frequency components are studied with higher time resolution than the low frequency components. This makes wavelet analysis a powerful tool for cases with phenomena of different scales like in Figure 5.

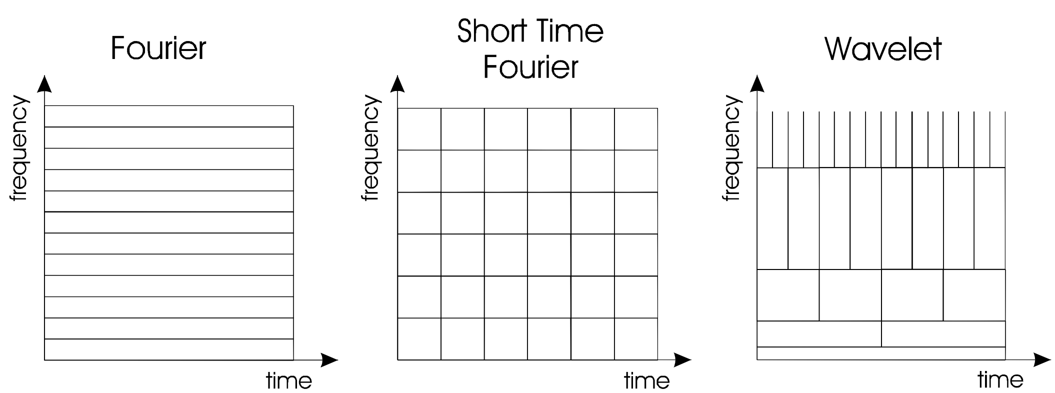

Figure 10. Comparison of Fourier, short time Fourier and wavelet analysis in time-frequency plane.

One topic to consider with the wavelets is the choice of a suitable mother wavelet since the different mother wavelets capture the properties of signals with different accuracy. The different wavelets need a varying number of coefficients to be calculated to represent a signal with a certain accuracy. For example, the most simple wavelet called Haar wavelet does quite a poor job approximating a function but a more advanced wavelets like Daubechies-4 of Figure 9 performs much better well (Figure 11).

As a conclusion wavelets are a useful tool for process monitoring and they have many desired properties. However, as with many other sophisticated tools one needs some time to learn to use them properly.

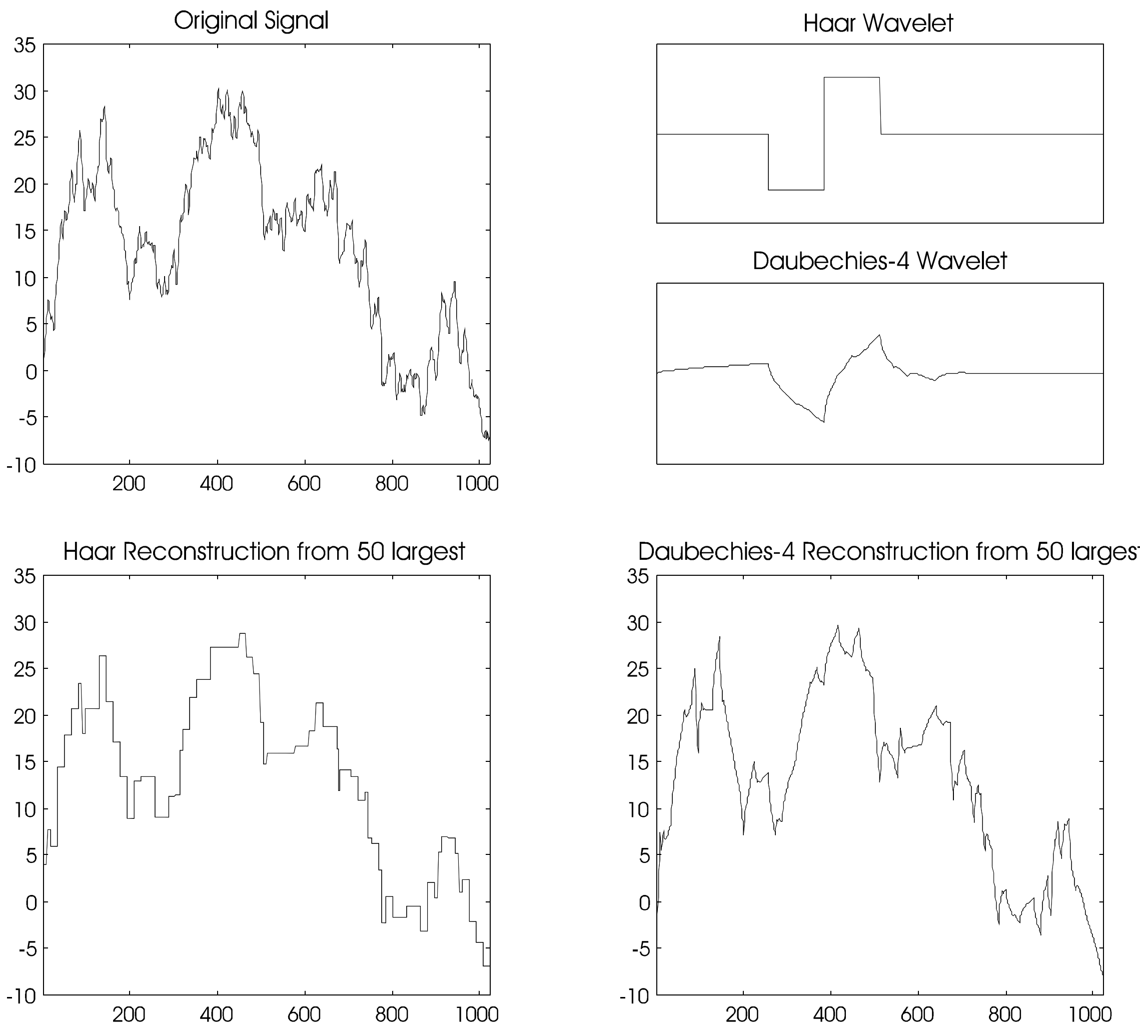

Figure 11. Signal (top left), Haar and Daubechies-4 wavelets (top right), and reconstructions of the signal with 50 largest wavelet coefficients of Haar (bottom left) and of Daubechies-4 (bottom right) wavelets (drawn with aid of WaveLab 802 library for Matlab [Donoho et al 1999]).

5 Summary

Monitoring of a paper machine is difficult due to huge number of measurements taken. This gives rise to a need for automatic monitoring and state identification. In this case study the state identification is based on measurements from machine drive controls.

There are two kinds of demands for the monitoring tool. Firstly, a real-time detection of clear undesired states to prevent more serious malfunctions and secondly, some way to search for reasons for malfunctions happened. The first goal can be achieved by automatic alarms when a control limit is crossed and the second one by writing a log of every unexpected change detected. Because the users of the tool are not experts on a machine inference, the information restored should be as easily interpretable as possible.

Some measuring methods and equipment pose limitations for the monitoring. A limited sampling frequency and a difficulty to synchronize measurements of different variables prevents monitoring of phenomena in scales less than a second. Additionally, derived nature of some measurements, periodic monitoring due to huge number of signals, and the unavailability of large amounts of abnormal data have their effect on monitoring.

Essentially, the state monitoring is detection of changes. The most important changes to detect are the ones in which the set points or external conditions do not change. The changes could be detected by comparing measurements between values of successive days, or values of drive controls within a certain section. Some typical states due to abnormal changes are a wrong load sharing and a presence of harmful frequencies. The alarms caused by changes can be triggered with an aid of control windows, which set the limits for changes. However, the control windows should be flexible because some situations like closing of a nip require looser limits than normal drive.

The flexibility of control windows is one kind of adaptivity. The monitoring system needs also other kinds of adaptivity properties. Short term changes like warming up of motors, long term changes like wear-out of machine parts and instant changes like changes in set points need all to be adapted by the monitoring model. Different parts of a paper machine require also the model to be quite general to suit for different inputs and states.

The suggested monitoring method is based on state identifier using extracted and selected features. Feature extraction means calculating new more descriptive variables called features from the measurements and feature selection choosing of a subset of the original set of features. The state identifier is an inference machine using the selected features to make inferences about the state of the process.

The rest of the report describes some feature extraction methods, including descriptive statistics, and trend, principal component, Fourier, and wavelet analysis. Descriptive statistics are easy to interpret but they have quite a limited ability to characterize the process. Trends are also intuitive but without a reasonable scaling the phenomena behind the trend could clutter each other complicating the interpretation. Principal component analysis is a useful technique to reduce the dimensionality of the problem but the process is assumed invariable. Fourier analysis gives the frequency information of the process without time resolution and short time Fourier analysis with a fixed time resolution. Finally, wavelet analysis can be used to characterize the process with variable time-frequency resolutions and it can also help in scaling problems with trend analysis.

References

[Babaud et al. 1986]
Babaud, Jean; Witkin, Andrew P.; Baudin, Michel & Duda, Richard O. "Uniqueness of the Gausian Kernel for Scale-Space Filtering". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 1, pp. 26-33, 1986.

[Bakshi & Stephanopoulos 1994a]
Bakshi, B. R. & Stephanopoulos, G. "Representation of Process Trends – Part III. Multiscale Extraction of Trends from Process Data". Computers and Chemical Engineering, Vol. 18, No. 4, pp. 267-302, 1994.

[Bellman 1961]
Bellman, Richard. "Adaptive Control Processes: A Guided Tour". Princeton University Press, Princeton, New York, 1961.

[Bishop 1995]
Bishop, Christopher M. "Neural Networks for Pattern Recognition". Oxford University Press, 1995.

[Chen et al. 1999]
Chen, B.H., Wang, X.Z., Yang, S.H. & McGreavy, C. "Application of wavelets and neural networks to diagnostic system development, 1, feature extraction". Computers and Chemical Engineering, Vol. 23, No. 7, pp. 899-906, 1999.

[Cheung & Stephanopoulos 1990a]
Cheung, J. T.-Y. & Stephanopoulos, G. "Representation of Process Trends – Part I. A Formal Representation Framework". Computers and Chemical Engineering, Vol. 14, No. 4/5, pp. 495-510, 1990.

[Cheung & Stephanopoulos 1990b]
Cheung, J. T.-Y. & Stephanopoulos, G. "Representation of Process Trends – Part II. The Problem of Scale and Qualitative Scaling". Computers and Chemical Engineering, Vol. 14, No. 4/5, pp. 511-539, 1990.

[Daubechies 1990]
Daubechies, Ingrid. "The Wavelet Transform, Time-Frequency Localization and Signal Analysis". IEEE Transactions on Information Theory, Vol. 36, No. 5, pp. 961-1005, 1990.

[Donoho et al. 1999]
Donoho, David; Duncan, Mark Reynold; Huo, Xiaoming & Levi, Ofer. "WaveLab 802 for Matlab5.x". [http://www-stat.stanford.edu/~wavelab/], 1999.

[Flandrin 1999]
Flandrin, Patrick. "Time-frequency/time-scale analysis". Academic Press, San Diego, CA, 1999.

[Flehming et al. 1998]
Flehming, F.; Watzdorf R. V. & Marquardt, W. "Identification of Trends in Process Measurements Using the Wavelet Transform". Computers & Chemical Engineering, Vol. 22, Supplement, pp. S491-S496, 1998.

[Grant & Leavenworth 1996]
Grant, Eugene L. & Leavenworth, Richard S. "Statistical quality control", 7th ed. McGraw-Hill, New York, 1996.

[Jackson 1991]
Jackson, J.E. "A User's Guide to Principal Components". John Wiley & Sons, New York, 1991.

[Jackson & Mudholkar 1979]
Jackson, J. Edward & Mudholkar, Govind S. "Control Procedures for Residuals Associated with Principal Component Analysis". Technometrics©, Vol. 21, No. 3, pp. 341-349, 1979.

[Kivikunnas 1999]
Kivikunnas, Sauli. "Overview of Process Trend Analysis Methods and Applications". Workshop on Applications in Chemical and Biochemical Industry, Aachen, Germany, Sept. 15, 1999.

[Kourti et al. 1996]
Kourti, Theodora; Lee, Jennifer & MacGregor John F. "Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control". Computers & Chemical Engineering, Vol 20, Supplement, pp. S745-S750, 1996.

[Liu & Ling 1999]
Liu, B. & Ling S.-F. "On the Selection of Informative Wavelets for Machinery Diagnosis". Mechanical Systems and Signal Processing, Vol. 13, No. 1, pp. 145-162, 1999.

[Mallat 1989]
Mallat, Stephane G. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, pp. 674-693, 1989.

[Nawab & Quatieri 1988]
Nawab, S. Hamid & Quatieri, Thomas F. "Short Time Fourier Transform" in Advanced Topics in Signal Processing, Jae S. Lim & Alan V. Oppenheimer, eds., Prentice Hall, Englewood Cliffs, NJ, 1988.

[Pittner & Kamarthi 1999]
Pittner, Stefan & Kamarthi, Sagar V. "Feature Extraction from wavelet coefficients for Pattern recognition tasks". IEEE transactions on Pattern Analysis and machine Intelligence, Vol.21, No 1, pp. 83-88, 1999.

[Pudil & Novovičová 1998]
Pudil, P. & Novovičová, J. "Novel Methods for Subset Selection with Respect to Problem Knowledge". IEEE Intelligent Systems, Vol. 13, No. 2, pp. 66-74, 1998.

[Rioul & Vetterli 1991]
Rioul, Olivier & Vetterli, Martin. "Wavelets and Signal Processing". IEEE Signal Processing, Vol. 8, No. 4, pp. 14-38, 1991.

[Sharma 1996]
Sharma, Subhash. "Applied Multivariate Techniques". John Wiley & Sons, New York, 1996.

[Vedam & Venkatasubramanian 1997]
Vedam, Hiranmayee & Venkatasubramanian, Venkat. “A Wavelet Theory-Based Adaptive Trend Analysis System for Process Monitoring and Diagnosis”, 1997 American Control Conference, Piscataway/NJ, USA, pp. 309-313, 1997.

[Wei 1994]
Wei, William W.S. "Time Series Analysis; Univariate and Multivariate Methods". Addison-Wesley, Reading, MA, 1994.

[Wu & Du 1996]
Wu, Ya & Du, R. "Feature Extraction and Assessment Using Wavelet Packets for Monitoring of Machining Processes". Mechanical Systems and Signal Processing, Vol. 10, No. 1, pp. 29-53, 1996.