

## 1.1. Main Objectives

---

The main objectives of this study are as follows:

1. To develop a Nonlinear Multiscale Principal Component Analysis (NLMSPCA) methodology for process monitoring that is able to effectively detect abnormal situations during their early development stage and to give a preliminary diagnosis of the cause of the problem.
2. To develop NLMSPCA monitoring software as a Matlab toolbox that incorporates the whole NLMSPCA methodology for step-by-step development and easy application.
3. Application of the NLMSPCA methodology to real nonlinear multivariate chemical process data so that the performance of the NLMSPCA methodology can be tested and validated.

In order to develop and explain the NLMSPCA methodology, an overview discussion is given on the following topics:

1. Linear principal component analysis
2. Nonlinear principal component analysis
3. Wavelet analysis
4. Input-training neural networks
5. Autoassociative neural networks
6. Statistical performance monitoring charts
7. Non-parametric density estimation

During these discussions those features that make these topics so significant for process monitoring will be highlighted since it is these features that will be combined to form the final NLMSPCA methodology. A discussion on abnormal situation management (ASM) is also included in order to emphasize and justify the need and significance of this work and also to put it into perspective with the global ASM methodology.

---

## 1.2. Background

---

In modern process plants controlled by distributed control systems, the role of operators has changed from being primarily concerned with control to a broader supervisory responsibility: analyzing operational data, identifying unusual conditions as they develop and responding rapidly and effectively by taking corrective actions.

Any action taken on a process operation generally relies on a description of the state of the operation or events that are occurring. Timely and correct interpretation of data through improved process monitoring and fault detection will lead to improved quality, reduced cost, safer operations, and waste reduction (Kosanovich et al., 1996; Davis et al., 1995). However, there are significant obstacles to using data for process monitoring and fault detection, including the sheer volume of the data, large numbers of variables, process noise, and the non-stationary tendency of the process data due to process and monitoring sensor drift.

## 1.3. Data Volume

---

The role of the operator has become a more challenging task than before because of the overwhelming volume of data operators have to deal with (Chen et al., 1999) due to chemical processes becoming increasingly measurement rich. Large volumes of data are recorded and are often not used until the process has undergone a significant upset. Although there may be hundreds of measurements in a typical chemical process, there are relatively few events generating this information.

High dimensional data analysis is becoming increasingly common as new problems are placing greater demands on computing resources. With high dimensional data, it is difficult to understand the underlying structure: it is difficult to "see the wood for the trees." Additionally, the storage, transmission and processing of high dimensional data places great demands on systems. This data can be very useful for process monitoring if the appropriate tools are applied. Hence, it is desirable to reduce the dimensionality of the data, whilst maintaining as much of its original structure as possible.

## 1.4. Process Shifts, Drift and Poor Process Data

---

Under ideal conditions a process would be stationary, i.e. retain the same mean and covariance structure over time. However, this is rarely observed over a long period of time so that most processes will exhibit non-stationary behavior over a long enough period. The process data may exhibit large amounts of normal systematic variation on

several time scales. This normal process drift is continuous on some time scales and discontinuous on others while variations due to faults can be relatively minor in comparison.

When a process suffers an out-of-control situation, the process behavior and normal process variation can be manifested in a variety of unnatural patterns such as cyclic, trend, systematic and sudden shift patterns. The root causes of process deviations and poor process data quality, as shown below are (Ghanim & Jordan, 1996):

(1) Trend:

- Tool, device wear.
- Aging.
- Human factors.
- Operator fatigue.
- Poor maintenance.
- Gradual change in standards.
- poorly or uncalibrated instrumentation
- high noise levels

(2) Stratification:

- Incorrect calculation of control limits.
- Oscillation caused by poorly tuned controllers etc
- The misplacing of a decimal point.
- Any causes for mixtures.

(3) Mixture:

- Error in plotting.
- Incomplete operation.
- Change in method of measurement.
- Carelessness in setting.

- Change to different kind of parameters material.

### (4) Cycle:

- Seasonal effects.
- Operator fatigue.
- Rotation of people on the job.
- Difference between gauges.
- Difference between day and night shifts.

### (5) Systematic:

- Difference between shifts.
- Difference between assembly lines.
- Systematic manner of dividing the data.
- Presence of a systematic variable in the process.

### (6) Sudden shift:

- New operator.
- New inspector.
- New machine/process.
- New machine/process settings.
- Out of design conditions (e.g. weeping or flooding)
- Temporary unstable phenomena caused by condition changes (e.g., change of crude oil or utility system)

The result is that it is normal for the process data to show considerable variation over time. This variation is often much larger than changes due to process faults. It has also been observed that the process mean shows more erratic behavior than the process covariance, i.e. how the process variables co-vary.

## 1.5. Current and Desired Practice

---

Fault detection in the petrochemical industry is routinely done with preset upper and lower limits for each variable in the petrochemical industry. However, the method sometimes does not detect faults in a short time, and furthermore, some kinds of faults are fully missed or are only found after a long delay. Operators usually take a succession of plant data as a trend (i.e. a slow-changing behavior) and unconsciously neglect fast-changing components as noise. Furthermore, they neglect the fact that significant information about faults is also contained in high-frequency components of measured data (Daiguji et al., 1997). The result is that most of these variations and especially the root causes of these variations cannot be observed or detected by current monitoring systems. Therefore, techniques are needed that are able to detect any form of process variation and systematic changes, and are also able to guide in the investigation of the root causes of these process deviations.

Without proper pre-treatment, the necessary interpretation is difficult, if not impossible. Gross data must be eliminated or modified and noise levels reduced. In many cases, critical information occurs over short duration, and hence, is difficult to detect. Rioul and Vetterli have described how wavelets can be used to pre-process data in order to better locate and identify significant events (Davis et al., 1995). Combining this type of data pre-processing with multivariate statistics holds great promise for generating useful insights into the problem of process monitoring, data analysis, and data interpretation.

A wide variety of data treatment methods and chemometrics techniques are available for application to process data, however, it is often not apparent what methods will be useful in meeting monitoring and fault detection goals (Wise, et. al., 1996). These applications can be roughly divided between those directed at maintenance of process instruments, e.g. calibration, and those concerned with maintenance of the process itself, e.g. statistical process control and fault detection. The focus of this study is on the latter. For this study principal component analysis (PCA) modeling methods, which are commonly used for multivariate statistical process control (MSPC), are used and modified to be robust over long time periods in the presence of process drift while remaining sensitive to faults.

## 1.6. Data Analysis and Process Monitoring

---

The terms data analysis and process monitoring, as used in the context of process applications, collectively refer to the interpretation and evaluation of sampled process measurements. Data analysis as used in this work is intended to describe how data are manipulated and used together with fundamental understandings to infer the state of a

---

physical process. Monitoring, on the other hand, refers to the classification of the data based upon a calibration model of expected behavior so that abnormal situations can be detected and fault modes isolated. Figure 1.1 is a simplified view of the on-line process monitoring activity.

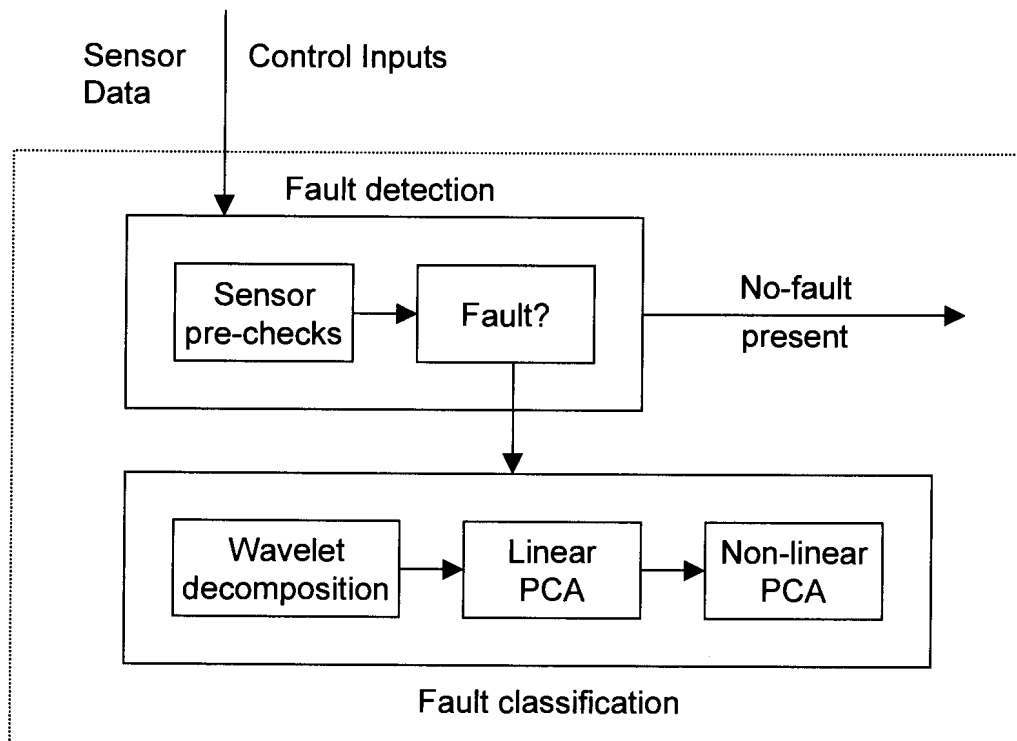


Figure 1.1 On-line process monitoring activity

### 1.7. Summary and Overview

Due to the aspects discussed in the previous sections of this chapter significant research has been done in recent years in more advanced techniques for multivariate process performance monitoring because of its increasing strategic importance. This research delivered promising results and followed the approach of reducing the dimensionality of the data by summarizing the data in terms of a smaller number of latent variables which are linear and nonlinear combinations of the original variables (Bakshi, 1998; Dunia and Quin, 1998; Dunia et al., 1996; Jia et al., 1998; Kosanovich and Piovoso, 1997; Nounou and Bakshi, 1998; Shao et al., 1999; Tong and Crowe, 1995; Wang et al., 1999). The most popular techniques are linear and nonlinear principal component analysis (LPCA and NLPCA). However, these analyses only concentrated on one or neither of the aspects of multiscale decomposition and NLPCA.

This study presents the non-linear multiscale principal component analysis (NLMSPCA) methodology which is an effort to combine the best of these techniques, with a few adjustments, to detect deterministic changes and extract those features that

represent abnormal operation. It combines the ability of non-linear PCA to decorrelate the variables by extracting both linear and non-linear relationships with that of wavelet analysis to extract deterministic features and approximately decorrelate autocorrelated measurements.

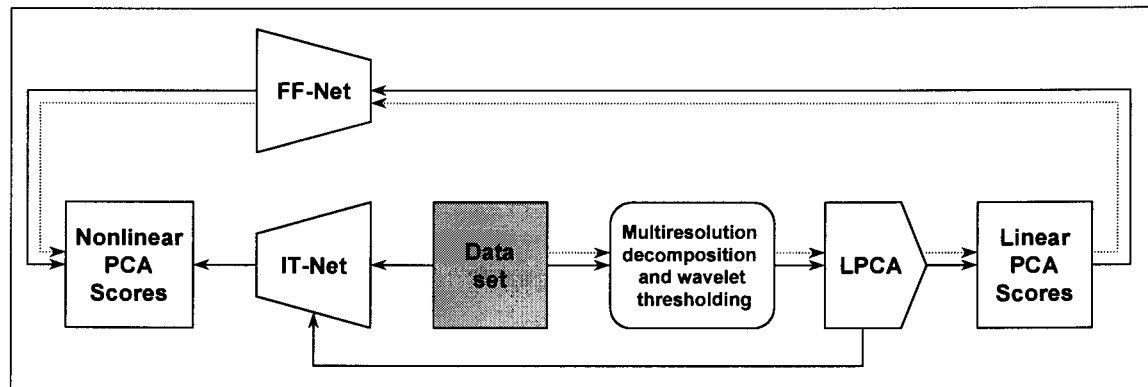


Figure 1.2. The NLMSPCA methodology (— Steps in the design phase; --- Steps in the implementation phase)

The non-linear multiscale methodology illustrated in Figure 1.2 consists of first decomposing each variable on a selected family of wavelets. Level dependent thresholding is then applied to the wavelet coefficients to select a smaller subset of wavelet coefficients. Thresholding of the coefficients at each scale identifies the region of the time-frequency space and scale where there is a significant contribution from the deterministic features of the signal. Details and approximations in the time domain are reconstructed from the thresholded and nonthresholded wavelet coefficients. The thresholded and nonthresholded details and approximations are treated separately. The nonthresholded details and approximations are combined. PCA is performed independently on the thresholded details and approximations at each scale and consists of both linear and non-linear PCA so that the process of extracting linear and non-linear correlations from the data can be performed separately. The same procedure is repeated on the combined nonthresholded details and approximations. For both linear and non-linear PCA an appropriate number of loadings are selected. Applying linear PCA results in a new set of uncorrelated ordinates. By retaining sufficient data variability, the underlying non-linear structure is not compromised and only those linear principal components associated with noise are discarded. Since the structure of the noise is not known a-priori, cross-validation as discussed in Section 7.10 is generally applied to assist in validating this.

Non-linear PCA is performed based upon the input-training neural network (IT-net) approach. Internal network parameters are trained using the Levenberg Marquardt algorithm while network inputs are updated using an extended backpropagation algorithm. This combined training approach results in faster convergence than just using backpropagation alone. After training the IT-net another network is trained that maps the observed data to the reduced data. An autoassociative network is then

constructed by combining the mapping network and the IT-net. The non-linear principal component scores are identified from the input layer of the IT-net. The advantage of this method is that both linear and non-linear correlations can be extracted from the process data to obtain a more parsimonious description of the original data. This method results in a conformance and generalized conformance principal component model.

Performance monitoring charts consisting of SPE and bivariate non-linear principal component scores plots with data-driven, non-linear control limits are derived to facilitate the comprehensive and robust occurrence of non-conforming operation. Detection limits for the scores and model residuals are computed at each scale from data representing normal operation and are calculated using the non-parametric technique of kernel density estimation.

The signal is reconstructed to the time domain and the scores and residuals for the reconstructed signal computed. The actual state of the process is confirmed by checking whether the signal reconstructed from the coefficients violates the detection limits of the PCA models. Since the reconstructed signal in the time domain is generated from the large wavelet coefficients, this approach integrates the task of monitoring with that of extracting the signal features representing abnormal operation, with minimum distortion and time delay. Consequently, there is no need for a separate step for prefiltering the measured variables. Furthermore, since the covariance matrix for all the scales together contain all the scale dependent information, the final detection limits to confirm the state of the process also adapt to the nature of the signal features. NLMSPCA transforms conventional single-scale linear PCA to a nonlinear multiscale modeling method, which is better suited for modeling data containing nonlinear contributions that change over time and frequency.

For on-line monitoring, the NLMSPCA algorithm is applied to measurements in a moving window of dyadic length.

A problem existing control chart displays are faced with is the space they occupy, limiting the display to only a few graphs at a time. A new approach is presented which allows the information to be displayed by univariate and bivariate control charts of the principal component scores and time-series plot of the squared prediction error (SPE), to be viewed in a compact manner so that the same information contained in multiple graphs can be viewed on a single display.

This advanced on-line process performance monitoring scheme is illustrated through application to a nonlinear multivariate chemical process. A complete toolbox has been created in Matlab to facilitate the design and testing of the advanced process monitoring scheme.



## 2.1. Introduction

---

Process monitoring and fault detection forms part of a much larger topic called Abnormal Situation Management (ASM). What follows may be regarded as an unnecessary long introduction to abnormal situation management. However, it is very important in the sense that it provides a background and bird's eye view over a subject for which I find it impossible to determine even estimated boundaries and puts this research topic of process monitoring in perspective to the global topic of ASM. Furthermore, it also provides some ideas for further research topics. I'm sure another ten years of intense research by a vast number of researchers can be spent on the subject of Abnormal Situation Management. The ASM Solution Anatomy model accompanying this work was developed from information collected from various sources (Anderson and Vamsikrishna, 1996; Bullemer and Nimmo, 1998; Cochran and Bullemer, 1996; Embrey, 1986; Harrold, 1998; Lorenzo, 1991; Musliner and Krebsbach, 1998; Nimmo, 1995, 1996, 1998a, 1998b; Rothenberg and Nimmo, 1996; Sticles and Melhem, 1998), including Internet searches, and represents a "generic" ASM solution.

Abnormal Situations have always challenged operations personnel, and they likely always will. Abnormal Situation Management is a particular challenge at this point in history because increased demands for higher efficiency and productivity have motivated the aggressive application of increasingly complex processes. The tremendous increases in the sophistication of process control systems through the development of advanced sensor and control technologies, and highly integrated approaches to production planning have led to productivity levels only dreamed of by previous generations of process engineers. The persistent paradox in the domain of supervisory control is that as automation technology increases in complexity and sophistication, operations professionals are faced with increasingly complex decisions in managing abnormal situations. However, the capacity of human operators to deal with this complexity, and the sophistication of their tools and user support technologies, has remained essentially unchanged and has not kept pace with the task demands imposed by abnormal situations. These sensor and control technologies have not eliminated abnormal situations and will not in the future. Consequently, operations personnel continue to intervene to correct deviant process conditions. Thus, the focus of this program is to develop collaborative decision support technologies that will significantly improve abnormal situation management practices.

Venkat Venkatasubramanian, professor at Purdue University's School of Chemical Engineering (Lafayette, Ind.), compares chemical plants with people who have a very

---

complex illness. "One or two doctors are unable to diagnose the illness. It takes a team of specialists each looking at the symptoms, each developing an opinion, performing additional tests, and then conferring with team members to reach a final conclusion." Similar to an ill patient, diagnosing a complex chemical process requires combinations of mathematical models, expert systems, neural networks, statistical techniques, and operations personnel, each working to independently diagnose an abnormal situation, with final diagnoses developed through cooperative problem solving.

## 2.2. What is an abnormal situation?

---

No standard definition of ASM exists. Although individual perceptions of abnormal situation management vary, there is consensus that "normal" and "abnormal" represent two distinct modes of operation. Abnormal Situations comprise a range of minor to major process disruptions or series of disruptions that cause plant operations to deviate from their normal operating state and in which operations personnel have to intervene to correct problems with which the control systems cannot cope. The nature of the abnormal situation may be of minimal or catastrophic consequence. A disturbance may simply cause a reduction in production; in more serious cases it may endanger human life.

Furthermore, abnormal operations are more likely during transition events such as startup and shutdown. Errors in situation assessment can be a source of abnormal situations, assumptions can direct plant personnel down the wrong diagnostic path and due to the response times required to correctly deal with a situation the problem may escalate.

## 2.3. Significance

---

To appreciate the significance of ASM one has to focus on the costs that accumulate with plant "hiccups," interruptions, unscheduled shutdowns, equipment failures, small losses of containment and quality problems. It is believed that solving these less dramatic disturbances potentially could yield a very high payback for companies. Estimates compiled by the ASM consortium (Harrold, 1998) indicate that elimination of all abnormal situations in petrochemical plants alone could add 5% to profits.

## 2.4. Goal

---

The goal as explained here represents a long-term goal. The Abnormal Situation Management approach is not just another attempt to introduce an "expert" artificial

---

intelligence device. Its success will hinge on its design as an embedded element in industrial automation system technology—integration is not enough. This long term goal is to drastically decrease the total costs of preventable process disruptions—saving industry millions of rands—by developing technologies that will offer better methods for informing operators, aiding operators during process disruptions, and preventing process disruptions in the first place.

This system should improve operator performance and offer a new challenge to operations by having the ability to interact with operations and production goals through the control system. The system should understand operations and maintenance rather than individual process variables. It should draw on other management techniques, such as incident investigation reports and the plant's corporate memory, as sources of knowledge. Useful design structures from process hazard analysis need to be captured within the system and used as rules for maintenance and operations activities.

The system should also address the communication issues identified in the site studies and provide solutions for plantwide communication, from the field to the control room. The existing industrial automation system technology from a wide selection of suppliers does not take into account casual users of the system. The same man-machine interface is provided for all users. The Abnormal Situation Management System should have the intelligence to recognize a user and provide information suitable for that person's discipline and knowledge of the industrial automation system.

Research should also address issues such as the impact of using a predictive plant state estimator on the alarm philosophy and man-machine interface. It should also incorporate an understanding of process operations and production goals and their relationship to safety, quality, environmental, and economic conflicts.

A comprehensive approach to the design of the human-machine system interaction is needed so that operations personnel receive information appropriate to their needs, while at the same time appropriate members of the operations staff are able to collaborate to solve the problem as a team. Individual needs vary as a function of a large number of variables: the current situation, the task being performed, individual preferences and styles—and others yet to be determined. In order to serve these needs, the information requirements need to be carefully assessed, not just for the current job functions present in existing plants, but for the job functions that will evolve as better decision aids become available and operators receive more support.

Systems must evolve so that the operator is not routinely swamped with information, aggravated by the user interface, required to use error-prone techniques to enter data, or exposed to situations in which being misled is even a remote possibility. The system must completely prevent adverse consequences from happening when the interaction of individuals predictably leads to misunderstandings, misperceptions, and mistakes. It

must also reduce, by orders of magnitude, the level of what post-incident review teams always label "human error."

There should be no such thing as a break down in lock-out, tag-out procedures—the user-machine system interaction model should utterly prevent such things from being possible. There should never again be coloured text on clashing coloured backgrounds on operational displays—the user interface development tools should make it very clear to the developer why such a design is inappropriate. Users should never again have difficulty in navigating from one display to another, should never again be able to enter a value for a set point that is outside the controller's capabilities, or ever again perceive the data from one unit as coming from another. And, looking to the future, decision support systems must never act like a back-seat driver when what the user needs is a helpful child—or vice versa.

When an abnormal situation is detected, operations and engineering teams must dynamically diagnose the root cause and correct the failure whilst trying to continue to meet the safety, environmental and production goals. At the same time they must track the underlying chain of events that led to the root cause(s) of the abnormal situation. As the abnormal situation evolves, some goals may need to be shed (that is, product quality, throughput, efficiency) if they compete with more critical goals (environmental or human safety).

The plant personnel should have a clear and up-to-date understanding of the types of abnormal situations recently experienced by their plants, the identified root cause and understanding of the incident investigation, and understanding of the correct steps to resolve this problem. Some plants have a variety of opinions on what was the root cause and generally lack understanding of the sources of abnormal situations and their impact on plant productivity.

Another goal is to enhance the ASM initiative to provide ways to detect and correct human errors before an undesired consequence occurs. Solution components for this problem are also beginning to emerge, but there is little consensus yet as to how to apply them. Operator intent recognition can help systems act in task-specific ways. Task modeling can help online information systems provide relevant (as opposed to canned) support. Tailored user interface displays can ensure that colour-deficient users can differentiate key data, users preferring graphs can see lots of graphs, and users needing quantitative information can see lots of appropriate numbers. And, user-centered design methodologies can ensure that this whole problem area is addressed in an empirically rigorous way when the analytically rigorous methods are lacking.

It is not a question of whether all these needs can be achieved, but rather a question of how long before they are achieved. Most of the technology is already available and just needs to be utilized and adapted.

## 2.5. Background: Chemical Plants and Control

A typical chemical plant will have of the order of 1000 readable “points” and a few hundred writable “points”. In addition to PID control loops, industries use distributed control systems (DCS) to simultaneously control thousands of process variables such as temperature and pressure and which can be programmed with numerous “alarms” that alert the human operator when certain constraints are violated (e.g., min/max values, rate limits). Control systems can be designed, programmed, and tuned to provide automated control for normal or near-normal operation. The major human role in this control is to supervise these highly automated systems. This supervisory activity requires: monitoring plant status; adjusting control parameters; executing pre-planned operations activities; and detecting, diagnosing, compensating and correcting for abnormal situations. The operator has a view of the values of all control points, plus any alarms that have been generated. The actions the operator is allowed to take include changing set points, manually asserting output values for control points, and turning on or off advanced control modules. Figure 2.1 gives an illustration of a typical control approach without abnormal situation management.

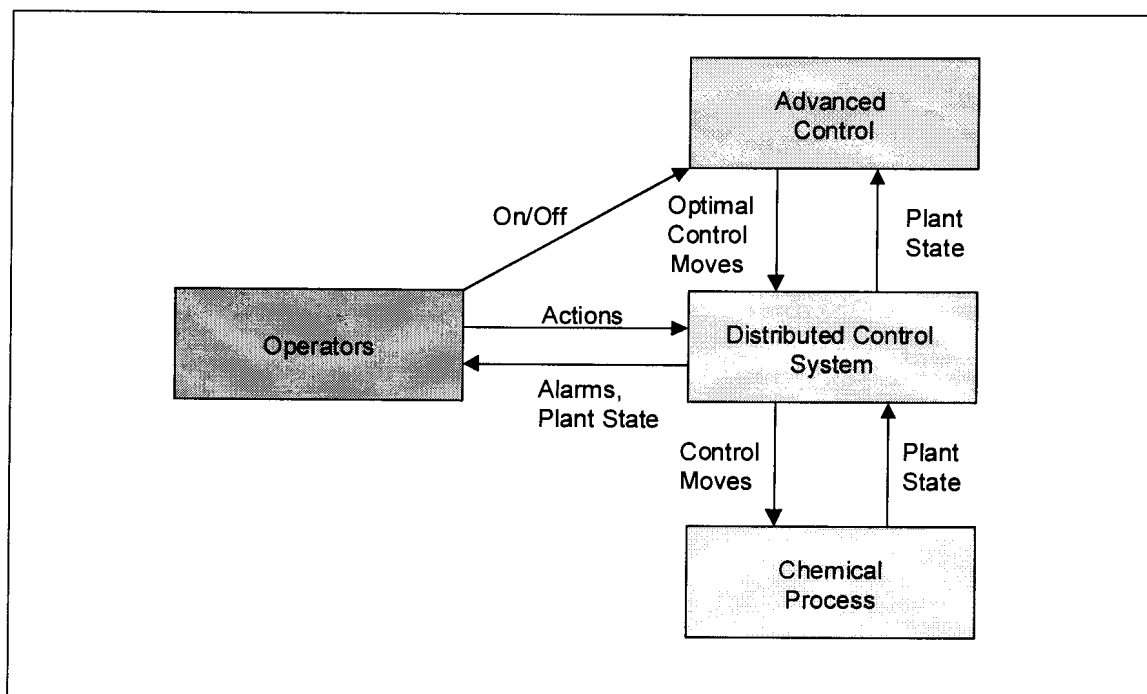


Figure 2.1. Control without Abnormal Situation Management

When the process becomes unsafe, safety instrumented systems designed to initiate a process shutdown, take over. But between normal operation and shutdown, processes can deviate into abnormal situations lasting a few minutes, or several days. Often deviations are undetected because automatic control readjusts the process. When an abnormal situation comes to the operator's attention, the common response is to place

loops in manual, reduce feed and energy streams, and manually attempt to return the process to a normal (steady) state—all the time searching for the initial cause of the problem. Frequently, the switch from automatic to manual control only worsens the situation, and a shutdown follows.

## 2.6. Past and Current Work

---

Previous approaches using technologies to assist operations in identifying and managing abnormal situations evolved large, specialized applications. These applications compared theoretical process models to real-time plant operations and generated alerts, recommendations, and predictions. Some success has been achieved with these solutions, but a lot of "care-and-feeding" is required to keep them current with ever-changing plant operations. Also, some systems use linear models that can ignore the nonlinearity and limitations of real equipment, and results in developing false predictions of equipment or process responses.

Attempts to integrate knowledge-based systems with plant operations have been few in number and mildly successful, mainly due to the complexities associated with:

- Integrating multiple proprietary platforms;
- Keeping the knowledge base current with ever-changing plant operations;
- Identifying and implementing models and methods best suited to handle the variety of complex problems of chemical process plants; and
- Getting all the operations "experts" to agree on what actions to take once the problem has been identified.

To address the problems associated with process disturbances, several industry leaders have joined forces with Honeywell to form the Abnormal Situation Management Consortium with the aid of a National Institute of Science & Technology Advanced Technology Program (NIST-ATP). Participating in the consortium are: Amoco, Chevron, Exxon, Mobil, Novacor Chemicals, Shell, Texaco and two software suppliers—Gensym and Applied Training Resources. This group is the offspring of the Alarm Management Task Force formed in the late 1980s to address problems associated with alarm functions in industrial automation systems and to suggest alarm-management enhancements. That group's work resulted in an important set of new features—defined and requested by users of the system—being included in the latest software release for Honeywell's TDC 3000X system, Release 500. The consortium estimates that by addressing the situations that are directly preventable, the losses attributable to abnormal situations can be reduced by 64 percent.

## 2.7. Methodology

---

If we are to address the problem and prevent incidents and provide tools for operators to perform more efficiently in abnormal situations we must understand the root causes of these incidents and the steps that need to be taken to eliminate or prevent escalation from an abnormal condition to a major catastrophe. The control system design needs to move from a reactive mode to a predictive mode and a long time before an alarm is initiated the system must predict the event using the latest state estimation tools.

The methodology of this research field needs to involve studying plants, reviewing previous years' history of plant incidents for different plants and sharing "best management practices". A systematic and statistical review of these incidents, together with interviews of operations personnel, can identify root causes of incidents, including problems introduced by today's industrial automation system technology and enabling technologies and the impact of system integration. Visits to sites also need to include human factor and personal performance reviews and research into how people and systems communicate. Today's offering of object-oriented software designs, relational databases, modular software development and maintenance tools, open communication standards, and acceptance of PCs makes development and deployment of knowledge-based ASM applications possible, but users still need to understand what they need and want.

This solution requires that technical challenges be overcome in three strategic areas:

- **Human-machine interaction:** A comprehensive approach to the design of the human-machine system interaction is needed so that a single user interface environment provides operations personnel with information appropriate to their needs, while at the same time supporting the collaboration of appropriate members of the operations staff in solving the problem as a team.
- **System architecture:** To provide accurate, timely support in abnormal situations, a system architecture needs to be developed composed of multiple processing modules, data bases and knowledge bases. These various software modules must communicate their conclusions with each other in real time and must remain coordinated among themselves and with human operators. Many past efforts have failed because this problem alone is so challenging.
- **System customization:** A major practical challenge in collaborative decision support technologies is configuring their capabilities to the idiosyncratic and dynamic nature of the plant processes and operations. Aspects of the software modules will need to be customized with specific knowledge about the

operations, equipment, personnel, and procedures of a specific site. Acceptable solutions will need to be self-adaptive or easily customized by plant personnel.

The system needs to be developed in a layered architecture based upon an opened standard, and so to enable it to run on any DCS which supports that standard.

Applications need to work together to determine the current state of the plant, decide upon the most appropriate goals to pursue, develop plans for pursuing those goals, and for executing those plans and monitoring the execution process. In addition, applications need to be responsible for communicating with plant personnel and for monitoring the Abnormal Situation Management System itself.

## 2.8. Sources of abnormal situations

---

Whilst major catastrophes are of concern they are fortunately infrequent and the major costs can be attributed to loss in production, quality problems, economic and conversion efficiency, equipment replacement and a collection of environmental issues.

The problems identified as contributing to abnormal situations falls into two major areas: human performance and performance of the industrial automation system and associated control equipment.

A lot of inspiration can be found in the excellent work done by Don Lorenzo for the Chemical Manufacturers Association, Inc. in his work "A Manager's Guide to Reducing Human Errors Improving Human Performance in the Chemical Industry". In this book Lorenzo states:

*"Historically managers in the CPI have found human errors to be significant factors in almost every quality problem, production outage, or accidents at their facilities. One study of 190 accidents in chemical facilities found the top three causes were insufficient knowledge (34%), procedure errors (24%), and operator errors (16%). A study of accidents in petrochemical and refining units identified the following causes: equipment and design failures (41%), operator and maintenance errors (41%), inadequate or improper procedures (11%), inadequate or improper inspection (5%), and miscellaneous causes (2%). In systems where a high degree of hardware redundancy minimizes the consequences of single component failures, human errors may compromise over 90% of the system failure probability".*

Safety groups estimate that human performance has been responsible for 80 percent of catastrophic incidents. The consortium's study identified several key personnel areas that hinder effective management of abnormal situations. These include: procedures not being followed, procedures that are too complex or unusable, lack of knowledge or



understanding, insufficient time to make effective decisions, and "information overload." In general, these are the results of poor context sensitivity and a lack of effective communication between the system and the people interacting with it.

Errors in sensor reading and valve positions cause a significant burden on the operations team. Operators have made poor judgment calls because the automated system reflects one value and the local traditional instrumentation registered a different value. The operator will often put trust in the device that is right most of the time especially if the other has maintenance or historical problems. Often the correlation between one process value and other variables are significantly complex, a good engineer may be able to discern that a pressure variable is incorrectly reading low given that a temperature is currently very high. Poor judgment on the part of the operator may result in erroneous diagnostics with potential catastrophic consequences.

Often, varied opinions lead to the development of multiple uncoordinated initiatives to address symptoms of a problem, whilst the root cause has not been correctly identified. The operations team believes that problems are caused by mechanical failures and the engineering teams are convinced that equipment failures are due to operational problems.

A contributing factor that does not raise the profile of this situation, and in some ways masks the problem, is the lack of measurement. This is especially true of the short upset, that may affect quality or cause slight loss of production, but which has a significant effect on net profit. Most large incidents are investigated, but the financial losses are often not recorded, making it difficult to help see the true cost in loss of product, quality restrictions, accident and injury expenses, and insurance reimbursement for damaged equipment or property. Currently only the obvious process variables are monitored like pressure, flow and temperature. However, there are other non-process variables that could provide needed diagnostic information such as noise, smell, real-time video images, infra red cameras for hot spots and many others that good field operators use every day using their human sensors.

Often escalation is caused by a series of "hidden" multiple failures in different systems. The skill level of the individual diagnosing and correcting these failures can have a significant impact on the success or disaster scenario. During a disturbance, when the highest degree of concentration is crucial, operators are currently faced with high noise levels and interference from outside sources such as phone calls, people traffic through the control room, unhelpful observers and lack of access to the control system due to the heavy traffic generated by alarms.

The largest contributor still remains the problem of time. For example, normal operation of a polyethylene process is relatively slow, but during abnormal operation a run-away reaction can cause very fast actions and there is no room for delay or error in correcting problems.

---

The Union Carbide's Bhopal Plant accident (Nimmo, 1996) started out as a minor problem and eventually escalated. The operator was trained and understood the actions needed to make the plant safe. As he implemented the procedures he soon discovered that backup systems were not available, cooling systems had been stripped down for use in other working parts of the plant, the flare stack was under maintenance and he was not aware of the full extent of what was in commission and was not available. When things went wrong it was not from the operators' wrong choices, but from their inability to take the correct action. That incident was based on a series of unfortunate circumstances and lack of management of change and coordination of information.

## 2.9. The solution

---

Developing a complete ASM solution requires implementing two parts, or layers as illustrated in Figure 2.2. The first layer validates incoming data and generates advisories of what is happening during an abnormal situation. The second layer predicts where the process is likely to go if current conditions persist. Some ASM solutions describe "closing-the-loop" between the ASM solution and the process. This is a form of supervisory control, with provision for the operations team to remain part of the diagnosing and prescribing process. While not all ASM solutions include all pieces of both layers, most provide the following pieces for constructing the advisory layer.

**A control system interface** that uses robust, real-time communication standards, such as OPC (OLE for process control), gateways to proprietary systems, or custom written application program interfaces, is necessary to obtain information from the control system about process measurements, valve positions, device status, etc.

**Sensor validation** to quickly detect sensor malfunctions or failures is critical to the integrity and acceptance of the ASM solution. For example, "failed" sensor input signals remain below a minimum value longer than a defined period, while "frozen" sensor input signals do not exceed the expected noise band for a period of time.

Jack Stout, president of Nexus Engineering (Kingswood, Tex.), explains, "The advanced diagnostics available in 'smart' transmitters and digital valve controllers is valuable in validating individual sensors. Many control systems can alarm, based on these diagnostic errors. ASM solutions differ by requiring sensor validation to include establishing sensor relationships to produce 'signatures' of equipment module and/or process unit performance. Informing the operations team that a pump has tripped because of cavitation, and that an empty vessel caused the cavitation, is a simple example of ASM sensor validation, alarming, and messaging."

**Point retrieval** of real and calculated process variable information is important in developing ASM solutions. Real process variables include temperatures, flows, pressures, analyzer results, control valve positions, etc. Calculated process variables include outputs to valves, totalized volumes, on-line material and energy balance calculations, etc. Combining real and calculated information is critical in developing performance "signatures".

**Message handling and viewing** must provide accurate, concise, and timely information about the current and future state of the process. ASM solution message complexity can vary from single line text messages to context sensitive help systems, allowing the operations team to view the appropriate level of detail. Some ASM solution message handlers automatically "pop" the initial alert on the operator's screen. After that, navigation buttons for cause-and-effect, details, procedures, and trouble-shooting are available.

**Alarm handling** that alerts the operations team of escalating circumstances during an abnormal situation requires advanced alarm management. Merely generating alarms, as many control systems do, is inadequate. As processes move through varying operational states, the operations team must remain focused on the task at hand. Spending time to work through complex alarm scenarios and then implementing advanced alarm management techniques will help the operations team to be more effective during a crisis.

**Incident history archives** are files of past process performance data. Initially the data may come from an existing data historian and can be used to playback past situations (good and bad) for testing the expertise of the ASM solution. Rolling data archives combine information collected by the point retrieval module and the sensor validation module into files that allow other modules to work with "smoothed" data.

**Custom and generic displays** are the operation team's window into the ASM workings. Custom displays are one-of-a-kind displays created specifically for a particular part of the process. Generic displays are templates for repetitive process areas (i.e., tank farms) with relevant data mapped into the display based on operator or event occurrences.

Combined, these pieces form the **advisory layer** to provide the operations team with early-warnings of a process' current health. However, the ASM solution requires additional sophistication to predict where the process is going.

The ASM solution **prediction layer** should develop equipment and plant signatures during normal operations and compare these to current operating signatures. Elements of this layer especially benefit by mixing mathematical models, neural networks, and statistical techniques to implement a solid ASM predictive layer.

For illustration purposes, the predictive layer consists of two parts: modeling, and planning and executing.

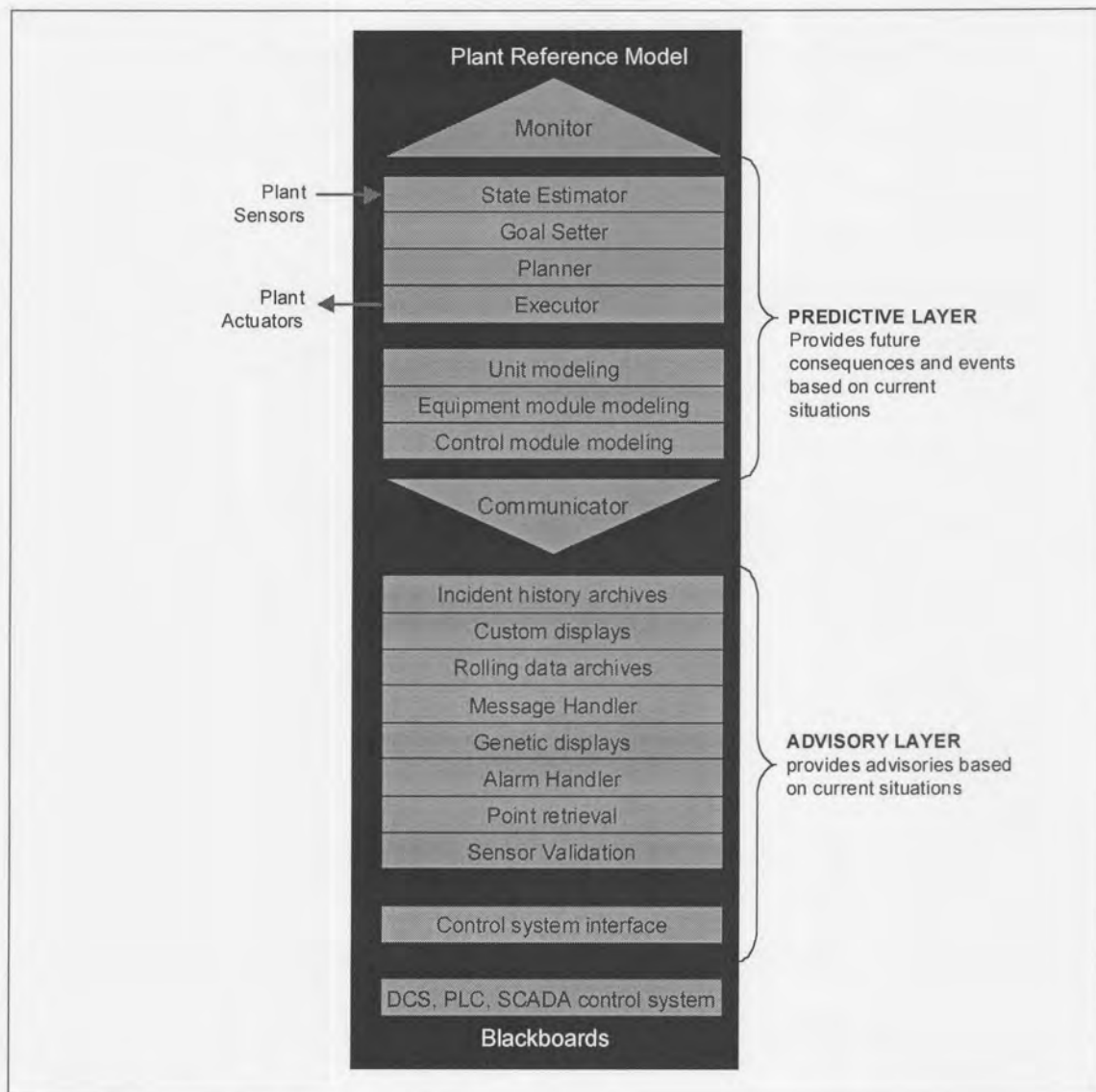


Figure 2.2. The ASM solution structure

### Modular objects

ASM problems are so complex that no single mathematical modeling technique is appropriate for each piece of plant equipment. Applying the appropriate model is easier when plant equipment is viewed as individual objects. For example, the model most appropriate for centrifugal pumps may differ from the model chosen for gear pumps. Developing models in an object-oriented programming environment to match plant objects, makes assembly and maintenance of the larger, more complex process models easier.

**Control module** (measurements, valve outputs, etc.) modeling allows development of sensor related calculations. For example, a "rate-of-change" calculation may be a more appropriate model for a temperature measurement than working directly with the process variable.

**Equipment module** (pumps, on/off valves, exchangers, headers, etc.) modeling combines control module models with equipment status to form mixed expression logic formulas. For example, combining the process variable value of a flowmeter in a calculation with the on/off status of a pump to determine if a flow rate should be present, avoids a low flow "nuisance" alarm when the pump is stopped.

**Unit modeling** combines control and equipment module calculations to form mathematical models of equipment, such as distillation columns, fluidic catalytic-crackers, fractionators, waste-heat boilers, and compressors.

The top layer of the anatomy diagram introduces very innovative concepts, especially for many chemical operations. But, as chemical complexity (and product value) increases, as quality demands continue to toughen, as pressure to reduce emissions builds, and as demands to "stay-on-line" echo through chemical operations, innovative thinking transforms good performing companies into great performing companies.

"Closing-the-loop" of an ASM solution requires very specialized functions, such as state-estimator, goal-setter, planner, executor, communicator and monitoring modules.

**State estimator modules** can determine the current process state, such as improving, staying the same, or getting worse, based on information provided from the lower layers of the anatomy at varying levels of abstraction, by fusing diverse sensor data and other available information (e.g. prior control moves, known malfunctions, human observations).

**Goal-setter modules** gather and maintain information relevant to quality and production goals established prior to the abnormal situation occurrence. It decides which of the currently-threatened operational goals should be addressed.

**Planner modules** develop and recommend recover-plans to address threatened goals selected by Goal Setter after refining multiple test results from current and historic knowledge of the process represented in the modeling and advisory layers.

**Executor modules** close-the-loop, monitor success, execute plans, and update other Abnormal Situation Management System components in progress towards goals.

**Communicator modules** communicate effectively with multiple plant personnel including DCS operators and field personnel located outside the control room.

**Monitor modules** observe the performance of the Abnormal Situation Management System components and may adjust or adapt the system's behavior in response to observed performance.

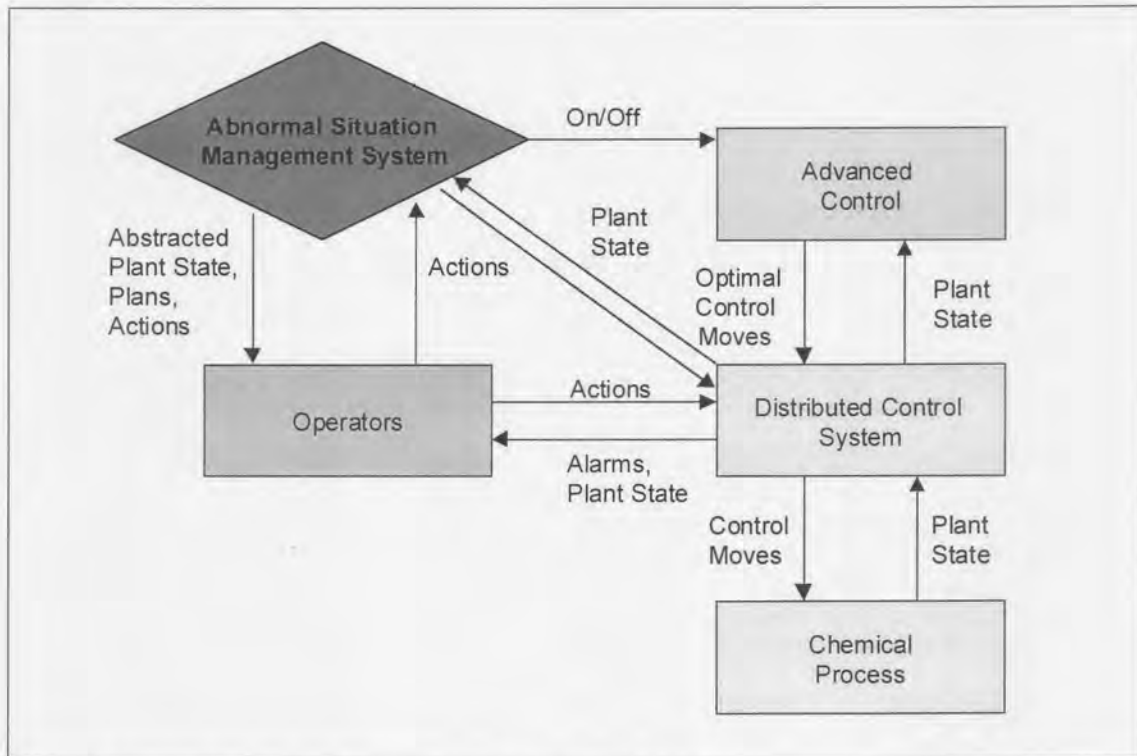


Figure 2.3. Control with Abnormal Situation Management

These functions interact by exchanging information on shared blackboard data structures. The Plant Reference Model blackboard captures descriptions of the plant at varying levels of abstraction and from various perspectives, including the plant's physical layout, the logical processing layout, the operational goals of each component and the current state and suspected malfunctions. Figure 2.3 shows how the Abnormal Situation Management System interacts with the existing system.

Abnormal situation management solutions are specialized applications of expert systems designed to work like the plant's best operator, on their best day, every day. These systems never get bored, distracted, or take a break; they remember what happened last week, last month, and last year, and provide accurate, consistent information, even in the heat of "battle".

One of the first initiatives within the solution proposal should be to provide a knowledge base of previous incidents. This system should allow the capturing of operating experience into a form of corporate memory. The learning and information from experience constitutes a fundamental source of requirements complementing those articulated by persons and organizations with an interest in the system and traditional

standards and regulations. It is well known that accidents often occur that could have been prevented by knowledge lurking in the corporate memory but forgotten and not applied. It is this knowledge that is to be fed back into the design process. The aim would be to extract this knowledge and apply it into an incident recall module. Once we have a better monitoring and investigation system we can apply the learning to design, operations and maintenance. This information could only be of use if the system understands the context of the operations:

- starting up
- shutting down
- operating normally
- what are the production targets
- what maintenance activities are being implemented or planned.

Hence, one of the strengths of solution should be its planning and goal setting ability. When contemplating ways to improve human performance Lorenzo (Lorenzo, 1991) states " there are two basic types of errors that managers must address:

- (1) Errors whose primary causal factors are individual human characteristics unrelated to the work situation and
- (2) Errors whose primary causal factors are related to the design of the work situation.

By providing resources necessary to identify and eliminate error-likely situations, managers can improve the performance shaping factors (PSFs) and dramatically reduce the frequency of human errors. This strategy Lorenzo calls the work-situation approach, and it involves the following elements:

- Implementing good human factors engineering of control systems, process equipment, and the work environment
- Providing clear, accurate procedures, instructions, and other job aids
- Providing job-relevant training and practice
- Providing ways to detect and correct human errors before an undesired consequence occurs
- Providing avenues for workers to achieve their social and physical needs

Therefore, the solution approach should form an integrated part of the current control system and have good human factors built into the control system. This will involve a new way of implementing control schematic diagrams. A style guide and implementation strategy needs to be developed for the next generation control system. More attention is recommended in control room design and the integration of different supplier's equipment. Also, we need to discover what new control screens are required and which of the more traditional need to be suppressed.

Another key issue is the configuration of the alarm system and the use of colours, symbols, priorities and how the alarm will be filtered e.g. context sensitive. In the past little attention was given to the design of the alarm system. Without proper consideration a process monitoring device is added to the control/monitoring system and on the surface it is a logical and justifiable action. During normal operating conditions this device does not cause any conflict, and may be a useful addition to the operator. However, during an abnormal situation this device is low priority and often becomes a nuisance to the operator trying to manage priority alarms.

What is required is some form of intelligence to put the alarm in context with the plant situation, eliminate unnecessary information and forward meaningful information to the operator and avoid information overload. For example, a temperature of 500 °C is appropriate under normal operations and any deviation by plus or minus 10 degrees should be annunciated but during plant shutdown the temperature may vary by new parameters, hence, the rules associated with the alarm need to change. Some processes actually cause the temperature to go outside the range of the transmitter and a new alarm is generated (BAD PV). This message is not very helpful because the operator now has to assume that the process is operating outside the range of the transmitter. This message could also mean that the transmitter is not working correctly and the temperature has not changed. To resolve this problem we need more sensor diagnostic information and better maintenance tools. The introduction of SMART Sensors has made a significant contribution to industry and has provided on-line calibration services and better diagnostics, however, more is required.

Sensors have made a significant contribution to error detection, improved reliability, and maintainability, however, what is really required is not just raw data but useful plant information. Measurement devices that are always suspect should be removed as their contribution may be only negative. When a loop shares different technology every effort must be made to ensure consistency and common calibration.

If the control system is well designed it will anticipate and prevent many situations from occurring. However, when the control system cannot maintain control, many plants are equipped with safety devices to ensure that the plant can be shutdown to a safe state. It is currently the job of the operations and engineering teams to identify the root cause of the situation and execute compensatory or corrective action in a timely and efficient

---



manner. A disturbance may simply cause a reduction in production; in more serious cases it may endanger human life, hence, the requirement for mechanical relief systems and automatic shutdown equipment which will always mitigate any failure of the control and ASM systems.

Training is recognized by most manufacturers as a major consideration in these situations. However there is not always a good understanding of the impact of poor training, hence the problems are not always eliminated. Manufacturers find it difficult to justify high-fidelity simulators or find the time needed for adequate training. Hence, operators lack confidence, gain experience in only normal operations which can contribute to difficulty in taking the correct actions within the time constraints imposed by an abnormal event.

DCSs are not incompatible with problem-based alarming: Indeed, mass balance analyses, expert systems, and statistical diagnostic techniques are becoming more widespread, albeit very slowly. What is needed to accelerate this trend is better ways to combine and aggregate data, better tools for easier, perhaps even automatic, development of such problem monitors, and higher-level, more comprehensive representations for plant equipment and processes.

### 3.1. Introduction

---

Since every abnormal situation is unique, it is difficult to study abnormal situation management, and in this case process monitoring, as a single subject. The best way to approach abnormal situation management is to study the theory and to then address a specific case. In this study process monitoring, as the first step to abnormal situation management, will be applied to a single nonlinear process and will lay the foundation for further investigation and development.

The advanced process monitoring methodology was applied to a real industrial process in order to evaluate its application capabilities. Due to the proprietary nature of the industrial example, only a cursory explanation of the industrial process is provided. The sensitive names of the process have been substituted with imaginary names and only normalised and standardised data are displayed. The data used, however, are real; the results of applying the methods are presented and discussed for process monitoring.

For the purpose of investigation a current problem in the steam export system at Company A was investigated since it contains all the interesting and important aspects of a typical abnormal situation.

### 3.2. Objective

---

At the time of the investigation Company A was busy building a new plant that would put a greater demand on the steam export system. The steam distribution system currently provides in the complete steam demand at Company A. However, with the new plant this demand will increase substantially. When any situation in the plant causes a decrease in the steam production, the steam export system won't be able to supply in the whole steam demand. Selective supply will then need to be applied since some processes will be more sensitive to a decrease in steam supply. A decrease in steam supply to the new plant for example will cause it to shut down.

The problem operators are faced with is the high nonlinearity that exists between steam production and steam distribution. No current accurate model exists that can relate the steam production to the steam distribution to a specific plant. This has the

effect that the influence of an upset in the steam production on a specific plant cannot be accurately anticipated in order to take preventative action with the result that an upset in the steam supply is only discovered when it is too late.

It will be to great advantage if any upset to the steam export system can be anticipated in advance in order to either take the necessary preventative actions to prevent it, or if it is not possible, to minimise the effect it would have on the whole system. In order to do this the cause of the upset needs to be identified as early as possible. The effect of the upset also needs to be quantified in order to quantify the preventative or impact minimisation actions.

So the main objective of the Abnormal Situation Management scenario under investigation would be to minimise any effect on the steam export from Process A. However, since only the process monitoring part is investigated this will not be possible yet. It should however be possible to identify the specific abnormal situation before it is noticed by the current alarm system or operator and identify the root cause of this abnormal situation in the steam export system. Quantification will also be partly possible. Only single faults will be investigated. In this study the objective is to confirm the abnormal situations identified since it was known prior to investigation from the plant history data.

### 3.3. Plant description

---

In order to understand the nonlinearity between steam production and steam distribution and why it is so difficult to generate a process model or to detect abnormal variation in the steam production or supply, one needs to look at least at an overview process description. The steam production and supply form a network throughout the whole factory. Appendix D gives an overview impression of the whole factory illustrating that the factory consists of a magnitude of separate process units linked with each other.

The most important fact to keep in mind is that during normal plant operation all the units are monitored independently. From the process description one gets a general idea of the multitude of interactions and the sheer magnitude of the process that needs to be monitored. These interactions cause many variables to be highly correlated. A general problem faced with observing such a magnitude of variables is deciding first which abnormal situation objective needs to be met (i.e. early detection of decrease in steam availability) and secondly which variables to monitor to meet this objective. These variables should be most representative of the whole process. Thirdly one needs to decide where the central monitoring system is going to be located since it will include variables from different process units. Although only a few variables are

selected in the end for monitoring purposes, there are many other factors that have an effect on these variables and on the normal operation of the plant. Therefore, the NLMSPCA system should be robust enough to detect abnormal operations despite other changes or disturbances occurring.

Since each unit is controlled separately (lack of plant wide control) it is currently almost impossible to determine when an abnormal situation is starting to occur. If two separate situations are developing in Unit 1 and Unit 2, without affecting the normal operation of these units, the effects will be carried over to unit 3 unnoticed. If, for example the two situations together have an abnormal affect on Unit 3 it will only be noticed after being carried over to Unit 3 which is some time after the initial 'symptoms' occurred in Unit 1 and Unit 2.

### 3.4. Scenarios

---

The advanced monitoring system was formally evaluated in five scenarios of which one was selected for discussion. These scenarios included sudden and unexpected malfunctions, problems originating in process equipment and the process itself.

The first set of influences investigated was that of a cutback in pure gas (PG) and reformed gas or fresh feed to Process A. Its effect on the steam export and in particularly its effect on the 43 bar steam export and export to gasification was investigated. For this purpose five sets of data were used during the month of November 1998. Firstly, data representing normal operation was gathered and used for training the system. For investigation purposes each data set represented an upset which caused a cutback in either the pure gas (PG), reformed (RG) gas or both.

#### Case Study 1

Date : 5 November 1998

6:30 Substation caught fire at Coal handling

Cut 700 000 m<sup>3</sup>/h on tailgas

Also affects gassification

Gassifiers 13-22 and 37 to 46 are shut down

7:15 PG Train 3 from Rectisol out of control due to insufficient gas

7:20 PG Train 5 from Rectisol out of control due to insufficient gas

**Case Study 2**

Date : 10 November 1998

15:22 Signal on Methane reforming compressor disappeared which caused train 5 to trip. Pure gas was cut by 100 000 m<sup>3</sup>/h.

16:17 Train 5 back in operation.

**Case Study 3**

Date : 17 November 1998

23:58 Oxygen 4 trip

00:05 Cut 100 000 m<sup>3</sup>/h pure gas

00:10 Trip 5 reformers

**Case Study 4**

Date : 20 November 1998

13:20 50 000 m<sup>3</sup>/h pure gas import from East

13:22 Fire wash feed pump tripped (train 4)

15:52 Rectisol train 4 trip

Cut to 180 000 m<sup>3</sup>/h

**Case Study 5**

Date : 26 November 1998

Loose compressors at cold separation (cooling compressors). This causes the feed to methane reforming to be halved (cutback on reformed gas). The other half (90 000 m<sup>3</sup>/h) that does not go back to Process A is flared.

For discussion in this report, case study 3 was selected. A second case study chosen for discussion did not involve an upset to the process itself, but involved the identification of an error in some calculation procedures after replacement of two control valves in the steam export system which influenced other parts of the system.

### 3.5. Background Process Information

---

#### 3.5.1. PROCESS OBJECTIVE

The overall objective of the steam system is to distribute steam at High, Medium and Low pressure to consumers in the factory for use among other as an energy source. This is done by producing high-pressure steam (40 bar and 43 bar) with boilers and Process A, and letting this down to medium pressure (8 bar) and low pressure (4bar).

#### 3.5.2. PROCESS OVERVIEW

The boilers produces superheated steam at 40 bar and Process A produces saturated steam at 43 bar. This is distributed to consumers and letdown to the 8 bar and 4 bar headers. Most important to notice is that Process A needs a fixed amount of the steam that it produces for internal use. Only the excess steam is exported. If an upset in the steam production is caused, Process A will first satisfy its own internal demand before exporting steam. This problem is addressed by example in Section 3.9 and will provide a better understanding of how the steam distribution network operates. Figure 3.1 gives a schematic that puts Process A and the steam production system in perspective to the rest of the plant. Figure 3.2 gives an illustration of the steam distribution network.

The 40 bar superheated steam is letdown to 8 bar via two letdown stations each with a desuperheater. The 40 bar also supplies the 4 bar with steam via three letdown stations each with a desuperheater. The major consumers of 40 bar superheated steam are Gasification, Oxygen plant, Power Generation and Process A.

The 43 bar saturated steam is letdown to the 8 bar header via 4 letdown stations and to the 4 bar header via 1 letdown station. Each of these letdown stations has a condensate knockout-drum. The major consumers of 43 bar saturated steam are Chemical Work-up, Phenosolvan and Gasification.

The 8 bar header receives steam from the 40 and 43 bar headers and supplies steam to the 4 bar header via three letdown stations each with two letdown valves in parallel and a desuperheater. Consumers of 8 bar include Benfield, Phenolsovan, and Rectisol.

The 4 bar header receives steam from the 40 and 43 bar headers. In the case of a high pressure on this header steam is vented to atmosphere via 4 vent valves. Consumers of 4 bar steam include Rectisol, Benfield and Chemical Work-up.

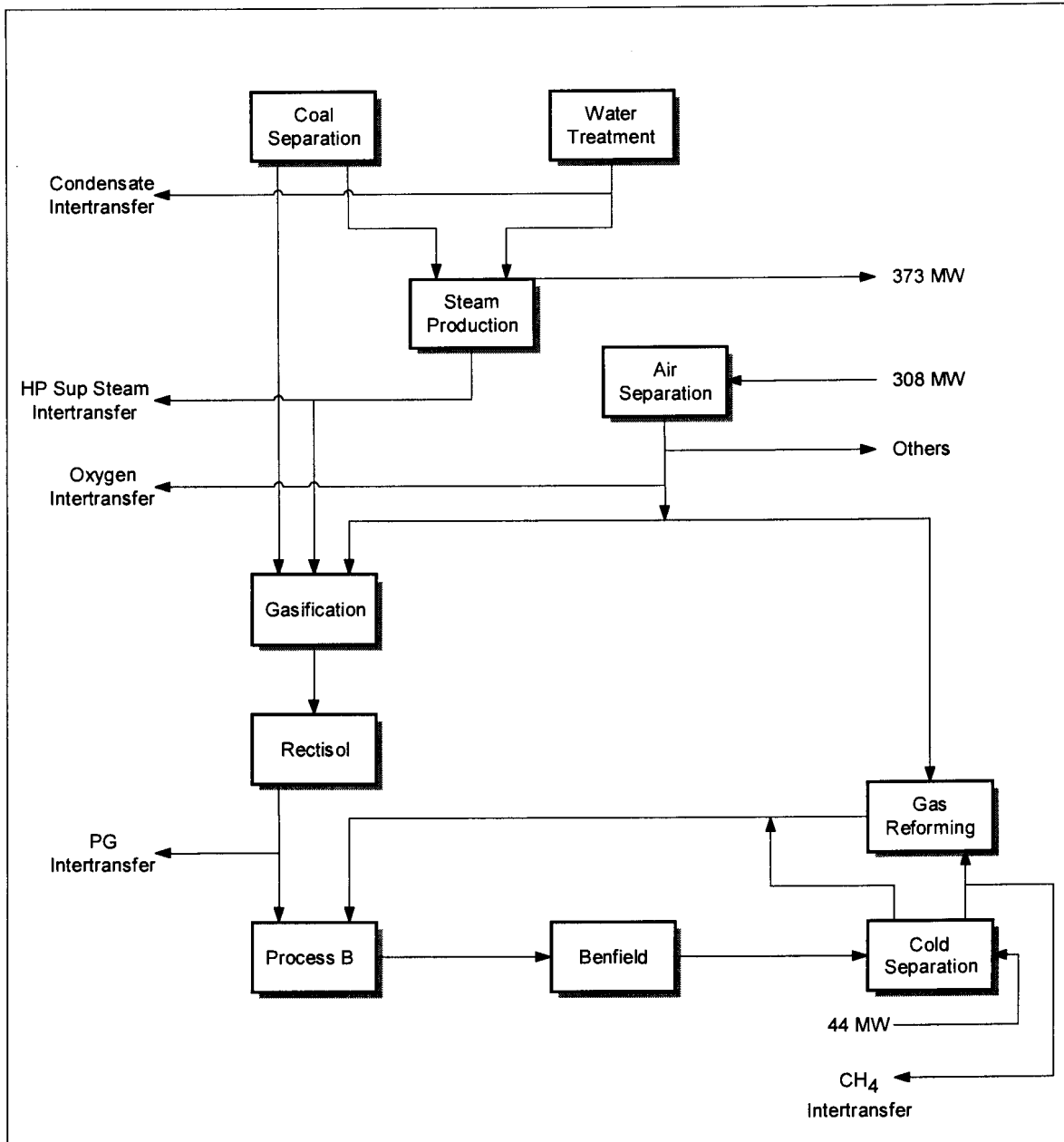


Figure 3.1. Process schematic of the most important process plants

### 3.6. Control Objective

The objective of controlling the steam letdown stations is to ensure a stable pressure on the headers and reliable temperature control when desuperheating.

In the case of one letdown valve going out of operation no deadband must exist and when the valve is brought back into operation bumpless transfer must be guaranteed.

Temperature needs to be controlled in such a way to ensure that as little as possible condensate will be present in the headers.

The system must also ensure that when pressure is lost on one header it must not affect the other headers drastically.

### 3.7 Header Functional Description

---

#### 3.7.1. 43 BAR HEADER, LETDOWN TO 8 BAR AND 4 BAR HEADERS

The objective is to control the pressure on the 43 bar header by letting down to the 8 bar and 4 bar headers. This is achieved by utilizing a pressure controller to which the operator enters the desired pressure setpoint.

The steam is supplied by the Process B reactors at 43 bar and 256 °C. The steam is letdown to the 8 bar header via 4 letdown stations (2 existing and 2 new) each with a knockout drum and control valve. The letdown to the 4 bar header is accomplished with one control valve with a knockout drum upstream from the control valve (1 new station).

A direct acting pressure controller operates in split range, first opening 3 of the 4 valves letting down to 8 bar then switching back to the last valve letting down to 8 bar.

This scheme supplies all the excess steam available on the 43 bar header to the 8 bar and 4 bar headers.

#### 3.7.2. 8 BAR HEADER, LETDOWN FROM 40 BAR AND TO 4 BAR HEADERS

The objective is to control the pressure on the 8 bar header by letting down from 40 bar and to 4. This is achieved by utilizing a pressure controller to which the operator enters the desired pressure setpoint.

The main steam supply is from the 43 bar header. The 40 bar header will supply any additional steam needed via two letdown stations in split-range. The reverse acting pressure controller operates in split-range between letting down to 4 bar (0-55%) and the letdown from 40 bar (55-100%).

A feedforward signal from the outputs of the valves letting down from 43 bar to 8 bar via a summation block is utilized to act when the pressure in the 43 bar header changes. When this situation occurs pressure will be stabilized by either shutting the 8-4 letdown valves or be made up from the 40 bar header.



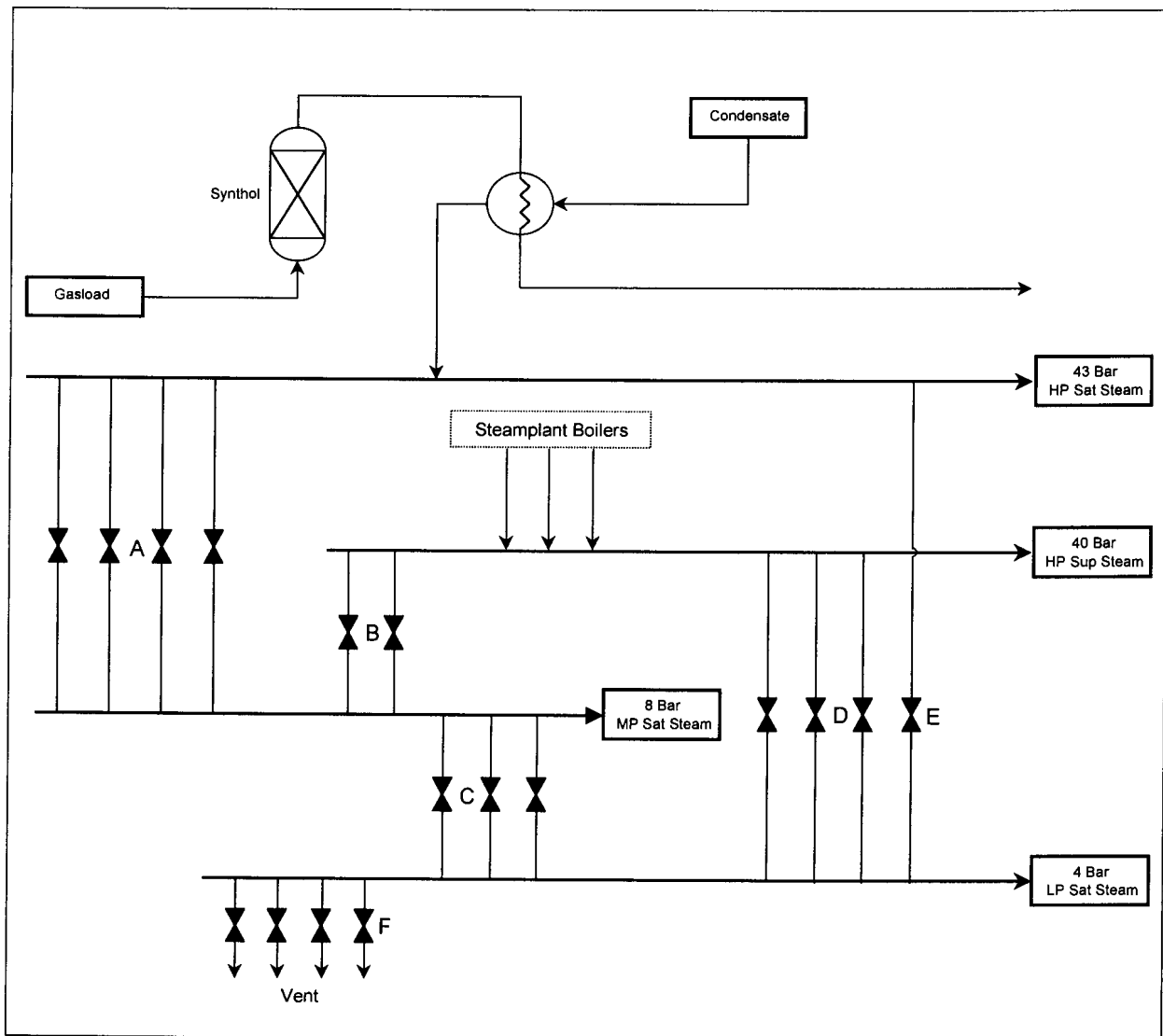


Figure 3.2 Steam distribution system

- A. 43 to 8 bar letdown
- B. 40 to 8 bar letdown
- C. 8 to 4 bar letdown
- D. 40 to 4 bar letdown
- E. 43 to 4 bar letdown
- F. 4 bar to atmosphere vent

If an under pressure situation occurs on the 40 bar an under pressure controller will override the pressure controller via an override low selector and close the valves letting down from 40 bar.

An over pressure situation on the 8 bar will cause the 6 valves letting down to 4 bar to open in split-range and relieve the situation. The valve on the northern side of the factory will be placed first in the split-range to alleviate the pressure drop problem in the northern side of the factory.

Temperature controllers on each of the letdown stations (except the new 8-4 letdown station) are used to control the amount of desuperheating.

The output of two of the valves has been characterized because the valves do not have a linear effect on the process i.e. when the one valve is 90% open and it is closed by 10% the other valves open 10% because of the bumpless transfer. When this occurs, a bump in the process is experienced because of the non-linear characteristic of the valve.

If the pressure in the 40 bar header drops, an under pressure controller will override the 8 bar pressure controller via an override low selector. When this happens a direct acting pressure controller will initialize to prevent windup. The output of one of the controllers is limited between 55 and 100%. This is done because when the controller reaches 55% both the valves letting down from 40 bar will be closed and any further reduction in output will have no effect. It is also desired that this controller does not influence the 8-4 letdown stations.

### 3.7.3. 4 BAR HEADER, LETDOWN FROM 40 BAR AND VENTING TO ATMOSPHERE

The objective is to control the pressure on the 4 bar header by letting down from 40 bar and venting to atmosphere by utilizing two pressure controllers to which the operator enters the desired pressure setpoint.

The 4 bar header receives feed from the 43 bar, 40 bar and 8 bar and vents to atmosphere. The steam is letdown from the 43 bar via one letdown station (new), from the 40 bar via three stations with desuperheating (existing), from 8 bar via three stations (two existing with desuperheating, 1 new without desuperheating) with two valves on each station and vents to atmosphere via four valves.

If the situation occurs where pressure on the 4 bar decreases, a reverse acting controller will increase the letdown from 40 bar in split range to increase pressure. If however an under pressure situation on the 40 bar system occurs at the same time a direct acting controller will override the pressure controller and close the letdown valves.

In the situation where the pressure on the 4 bar header increases the reverse acting controller will decrease the letdown from 40 bar until normal situation is reestablished.

## 3.8. Process B Reactor

---

### 3.8.1. GENERAL

The Process B reactor was designed as a replacement for the existing Train 8 CFB reactor which remains as a standby "swing" reactor. The Process B reactor makes use of an existing Train 8 quench column, product cooling train, and total feed compressor. The existing cooling train is debottlenecked by a quench column top pumparound cooler that preheats the total BFW to the Process A area. The Process B reactor has its own reactor coolant system.

### 3.8.2. OPERATION

The Process B reactor takes its total feed from the existing Train 8 CFB reactor inlet line. The gas enters the bottom of the Process B reactor through a gas sparger. It flows up through a distributor grid that supports the fluid catalyst bed. As the feed gas flows through the bed, hydrocarbons, water, and oxygenates are synthesized via the Fisher-Tropsch reaction. All reactor products are in the vapor phase at reactor conditions. Water and carbon dioxide are formed via the water-gas shift reaction. A mixture of water - and oil-soluble oxygenated hydrocarbons are byproducts. The reaction is exothermic. A portion of the heat of reaction heats the feed gas from the inlet temperature of the reactor to the operating temperature. The excess heat of reaction is removed by generating high pressure steam in the cooling coils.

Catalyst that is entrained from the bed with the gas stream is separated in internal cyclones and returned to the bed. The cyclones discharge effluent gas into a plenum from which the effluent gas exits the reactor. The effluent line ties into the existing Train 8 reactor effluent line upstream of the existing hot quench tower.

### 3.8.3. REACTOR COOLANT SYSTEM

The reactor coolant system removes the excess heat of reaction from the reactor by generating high pressure steam. BFW is fed to the steam drum through a level control valve to maintain the drum level. Saturated water is fed to the BFW circulation pumps. The BFW is pumped to the reactor cooling coils which have on/off valves on the inlets.

The reactor temperature is controlled by the operator varying the number of cooling coils in operation. Water is partially vaporized as it flows through the coils and a saturated water/steam mixture returns to the steam drum. The generated steam is disengaged from

the water in the top section of the drum. The steam exits through a demister and flows through a pressure control valve that maintains the steam generating pressure constant .

The steam separator separates any water that is carried over from the drum plus condensate formed by dropping the pressure. The steam then goes to the high pressure saturated steam header. Condensate is fed to the low pressure steam header. A continuous blowdown stream (for conductivity control) goes to the blowdown header.

### 3.9. Process A Reactors

---

#### 3.9.1. PROCESS A REACTORS AND UNIVERSAL EQUIPMENT

The process flow can be separated into the following steps:

1. Fresh feed gas is taken in at the Process B plant. This gas consists of a mixture of the following:
  - Pure gas from the Gasification and Rectisol plants,
  - Hydrogen-rich gas from the Cold Separation plant, and
  - Reformed gas from the Gas Reforming plant
2. The fresh feed gas that is added to an internal recirculation stream, is then compressed by the total feed gas compressor, heated and fed to the reactor.

#### 3.9.2. PROCESS A REACTOR TRAINS

1. The gas that enters the reactor picks up catalyst and carries it through a reaction chamber where the Fisher-Tropsch synthesis takes place. Two banks of cooling coils are provided to remove the heat from the reaction. High pressure steam is generated in the cooling coils and then exported via the steam drum. The catalyst is separated from the gas by means of five sets of cyclones. The reactor outlet gas is separated downstream into different products.
2. The reactor outlet gas is fed to the quench tower, where a light oil stream as well as a heavy oil circulation stream is injected into the gas stream.
3. Overhead vapours from the quench tower are now cooled and condensed in air conditioners followed by a shell-and-tube heat exchanger. This stream is then separated in the separations drum to form three main product streams.

- Uncondensed gas which is partially used as the internal recirculation stream, and partially as spare gas.
- Light oil, of which a large portion is recirculated to the quench tower and the net oil production, that is exported to the light oil stabilizing plant.
- Reaction water that is pumped to the water degassing plant.

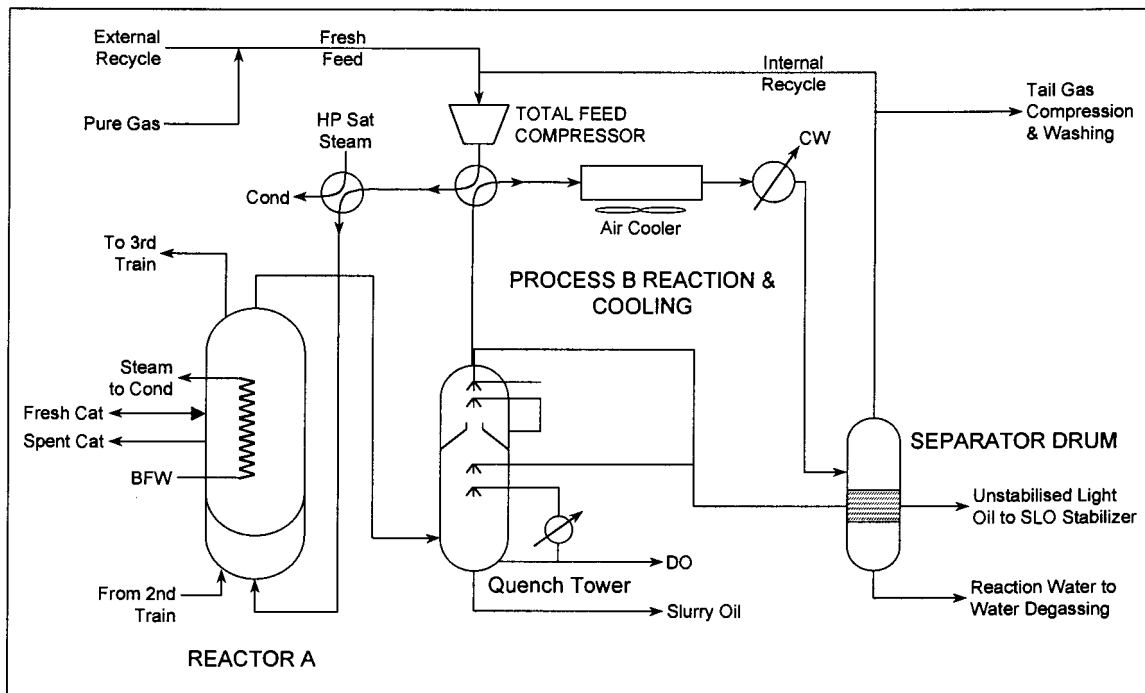


Figure 3.3. Process B reaction, cooling and recycle Loop

The flue gas is compressed by means of a centrifugal compressor and washed with water to remove the non-acid chemicals before it is transported to the plants further down.

Carbon monoxide is removed downstream from the Process A plant through the Benfield plant. The Cold Separation plant divides the gas into hydrogen-rich and methane-rich streams, as well as a C<sub>2</sub>-rich and three condensate streams.

A large part of the hydrogen-rich gas streams is recirculated from Cold Separation to Process A. The methane-rich stream is reformed with oxygen to manufacture carbon monoxide and hydrogen. This reformed gasstream is also recirculated to Process A. The reformed gas and hydrogen-rich gas streams are called the external recirculation.

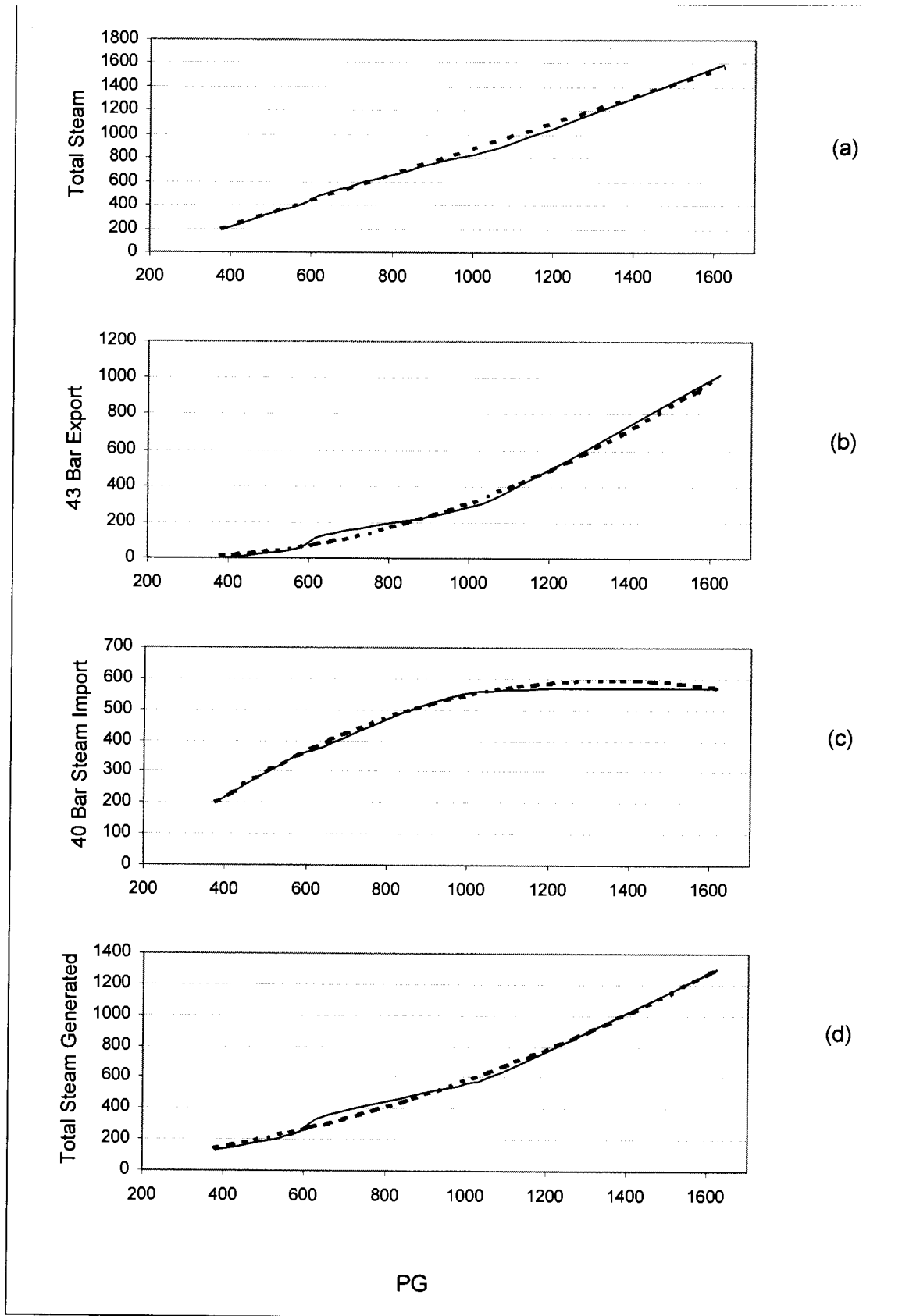


Figure 3.4 Nonlinear relationships between PG, steam production and steam export.

### 3.10. Steam Relationships

The figures in Figure 3.4 and Figure 3.5 will serve to illustrate the nonlinear relationships that exist between the gasloads, steam production and steam export. All the figures presented illustrate the relationship between various steam quantities and pure gas feed. The dashed lines are regression models fitted to the data for interest.

From Figure 3.4(a) we can see that there is a nonlinear relationship between the total steam being utilised in the system and the pure gas supply. This total steam is a summation of Figure 3.4(b) and Figure 3.4(c). Figure 3.4(b) clearly illustrates the nonlinear relationship between the PG-supply and the 43-bar steam export. Figure 3.4(c) illustrates the relationship between the PG-supply and the 40-bar steam import and Figure 3.4(d) gives the relationship between the PG-feed to and total steam generated by Process A. The influence on the steam export is clearly illustrated in Figure 3.5 generated from the data in Figure 3.4. From Figure 3.5(a) we can see that a 10% reduction in PG feed will cause a 24% reduction in 43-bar steam export and a 12% reduction in the total steam export.

A 50% reduction in PG feed will cause an 88% reduction in 43-bar steam export and a reduction of 40% in the total steam export.

From this it is evident that the problem lies with the 43-bar steam export. Thus, a small upset in the PG feed to Process A can have a huge effect on the 43-bar steam export, which in turn can have a major influence on the rest of the system since so many plants are dependent on the steam supply.

### 3.11. Process Variables

A list of all the process variables appear in Appendix C together with the calculated variables used in the investigation. From this list eight variables were selected that most accurately represent the system under investigation and is listed in Table 3.1. More detail will be provided in Chapter 5.

Table 3.1. Process variables used in the investigation (See Appendix C)

Variable number	Variable Description	Variable number	Variable Description
1	Total Rectisol Feed	5	Total Tail gas
2	Total Pure Gas feed	6	Total Steam Consumers
3	Total Reformed Gas Feed	7	Total Steam Letdown
4	Total Fresh Feed	8	Total Steam Export (Measured)

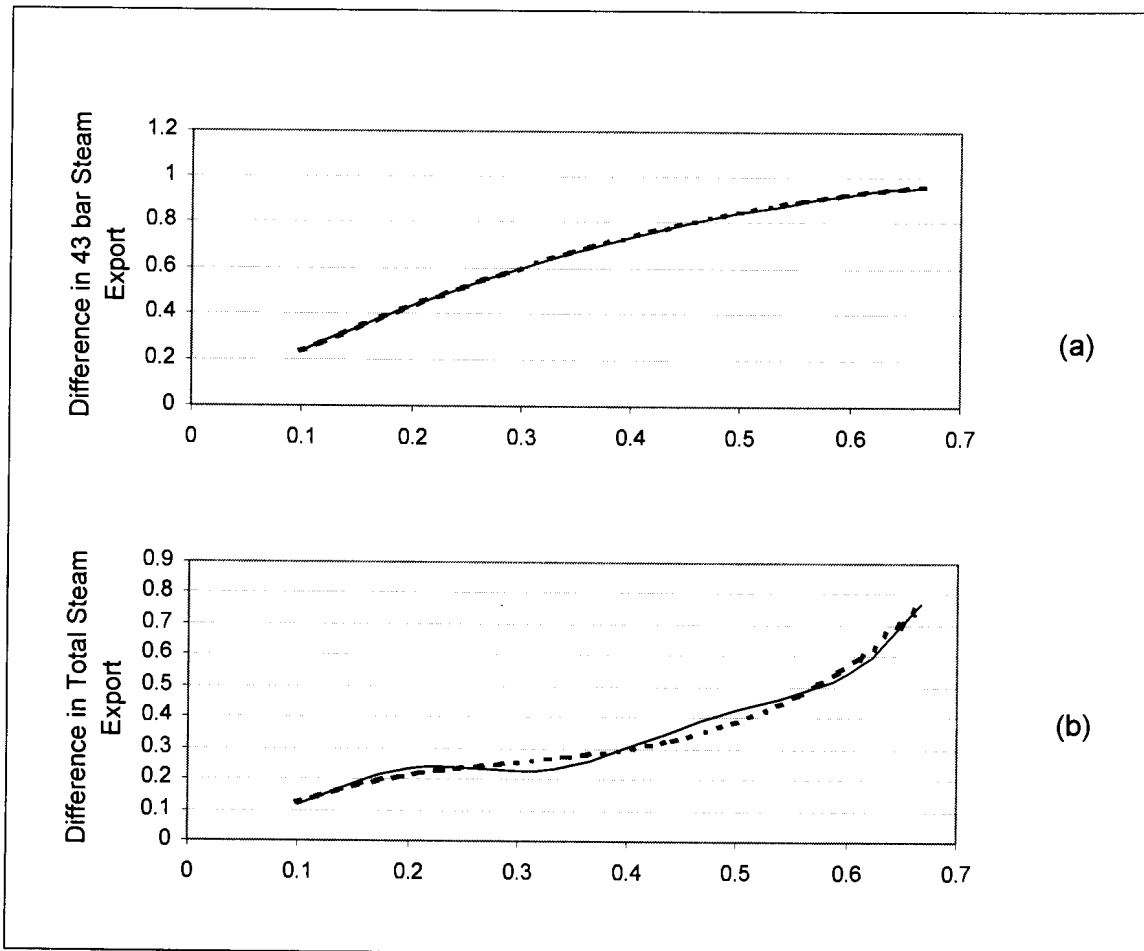


Figure 3.5. Differential relationships between PG feed and steam export



### 4.1. Introduction

---

The software used in this research was developed using Matlab. The full functionality of Matlab was implemented in order to create a toolbox that provides as much user-friendliness as is currently possible with Matlab. The creation of a separate complete NLMSPCA toolbox can be justified due to the following reasons:

- There is currently no software available for NLMSPCA since it is a new concept.
- Matlab does include a Wavelet Analysis Toolbox, but it does not contain any boundary corrected wavelet filters, does not incorporate the necessary threshold methods and cannot be used online.
- Matlab also includes a Neural Network Toolbox, but this toolbox does not allow input training or modification of algorithms.
- Using Matlab's toolbox functions makes it difficult to understand the mathematics and concepts.
- The structure of the toolboxes is such that it is very difficult to make alterations to the current software.
- It also would have been difficult to link the different toolboxes in order to form a complete functional step-by-step procedure.

The significance of the toolbox lies in the fact that:

- It operates independently from other toolboxes;
- It is understandable so that modifications or alternative ideas can easily be incorporated or linked to the current software;
- It is user-friendly;
- It automates the whole process allowing a step-by-step procedure for NLMSPCA;
- The complete toolbox can be used and accessed via user-interface;
- The flowcharting method used makes the various steps easy to follow;
- Help and background information are provided for quick reference.

The documentation provided here gives a thorough description of the toolbox and Appendix B provides extra information on the setup of the programs for someone who wishes to make alterations or use some of the applied methods in their own software development.

## 4.2. To Get Started

---

The toolbox is very easy to set up. Execute the following steps:

1. Simply copy the \Monitor directory located on the supplied cd to an appropriate directory on your hard drive for example c:\. For this example the directory c:\Monitor will then exist on you hard drive.

2. Start Matlab

3. At the Matlab command prompt set the path to the \Monitor directory for example

» cd c:\monitor

4. At the command prompt execute the following command

» startupm

5. The following message will appear

=====

This will set the search path to include all

the essential Monitor subdirectories

=====

Specify the path to the ...\monitor directory :

6. Enter the path to the \monitor directory for example

c:\monitor

and press enter. The setup may take a few seconds. While the setup is in progress the message in Figure 4.1 will appear.

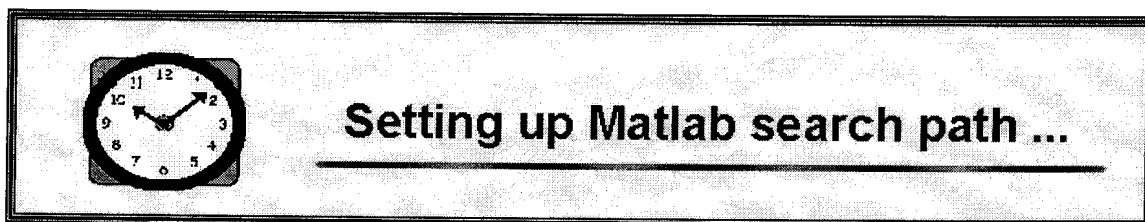


Figure 4.1 Setup progress display

7. The path setup will be acknowledged by displaying Figure 4.2.
-

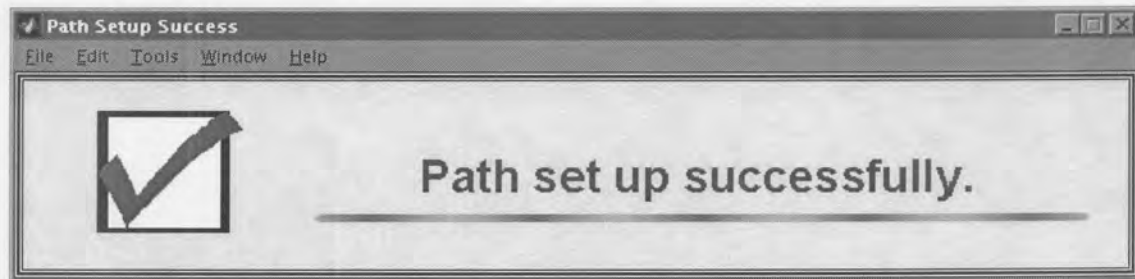


Figure 4.2. Path setup success display

8. Start the Matlab Path Browser from the Matlab workspace menu and save the path as illustrated in Figure 4.3.

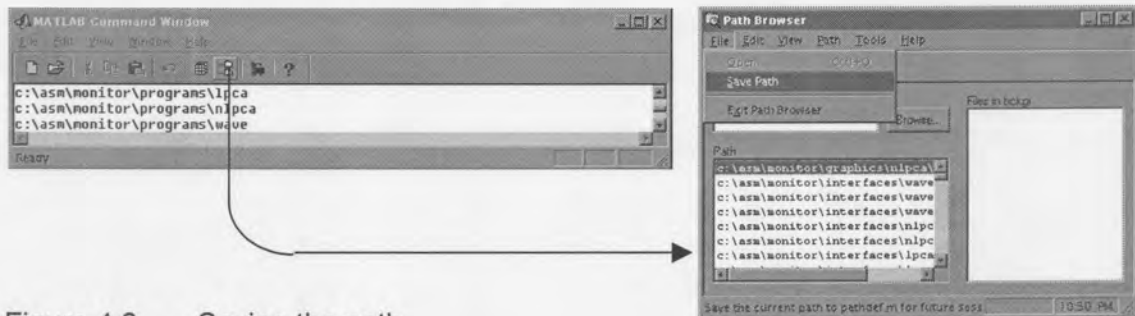


Figure 4.3. Saving the path

9. The setup is now complete. The setup will advance to the Database Setup, which is discussed in the next section.

## 4.3. The Program

### 4.3.1. DATABASE

The application makes use of a database in which all the necessary variables are stored. Each time the application is run, the variables in the database is updated. The database :

1. ensures that data is not lost while the training phase is in progress since it can take up to a few hours to generate this data,
2. saves the information generated during the training phase so that it is available for the application phase,
3. ensures that the data is available for further processing, comparisons and independent plotting.

The database is a Matlab mat-file. The system contains a default database called *data\_base.mat*. This database resides in the ...\monitor\database directory. Appendix C contains a list of all the variables that are contained in the database. As discussed in the previous section, after completion of the Path Setup, the setup will advance to the Database Setup in by displaying the Database Setup Interface in Figure 4.4. It contains the name of the default database. However, a new database can be created by changing the name of the specified database.



Figure 4.4. Database Setup display

The default database can be selected by clicking on the OK-button. If the default database is used the current variables that reside in the database can be used as default or can be overwritten by choosing the *Retain* option where available. If a new database is created no default options exist for the first time this database is used. The results in different databases can be compared with each other. Note however that the variables in the different databases will be the same so that if variables from two different databases need to be compared with each other the variables in the first database first need to be renamed before loading the second database. If not, the variables from the second database will overwrite the variables in the workspace loaded from the first database. After clicking the OK-button, creation of the database will be acknowledged by displaying Figure 4.5.

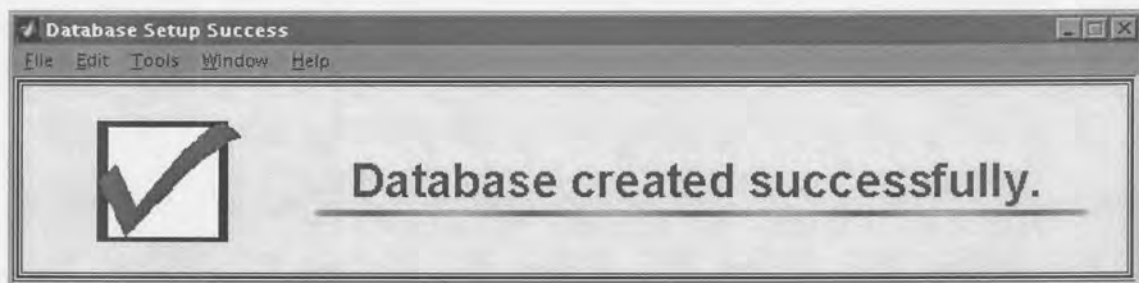


Figure 4.5. Database creation success display

### 4.3.2. INTRODUCTION DISPLAY

After creation of the database has been acknowledged the introductory window in Figure 4.6 will be displayed:

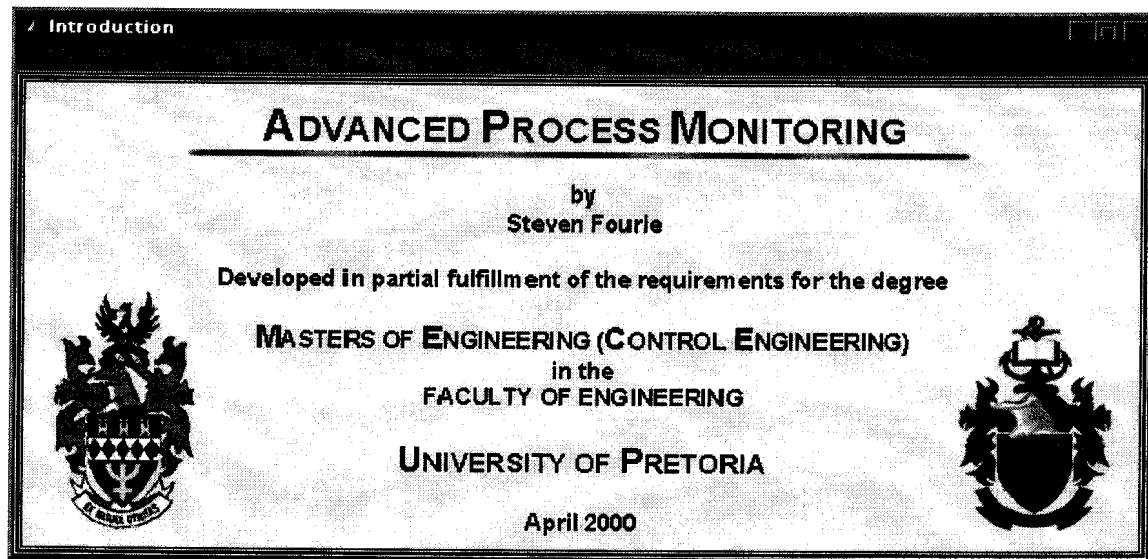


Figure 4.6. Introductory display

### 4.3.3. MAIN INTERFACE

Figure 4.7 Tags:

1. Go back to previous page.
2. Provides help on the current window.
3. Provides background information on the topic addressed in the current window.
4. Exit the program.
5. Exit the current window and advance to the following window.
6. Data selection and setup.

After displaying the introductory interface the setup will advance to the main interface in Figure 4.7. The main interface is a shortcut interface to all the main processing steps in the process monitoring setup which includes the following:

- a. Data selection and setup
- b. Wavelet analysis
- c. Linear principal component analysis

- d. Nonlinear principal component analysis
- e. Demapping
- f. Bivariate plot setup
- g. SPE setup
- h. Monitoring

(a) to (f) form part of the setup process which uses the normal operating process data. (h) is the actual monitoring process with new data. Selecting button 6 will take you to the first step in the process monitoring setup sequence. If you want to use current data from the database, you can jump to any other step in the setup sequence by selecting the appropriate button from the main menu. Thus, it is not necessary to start the whole process all over again if you were unable to complete the whole NLMSPCA setup process.

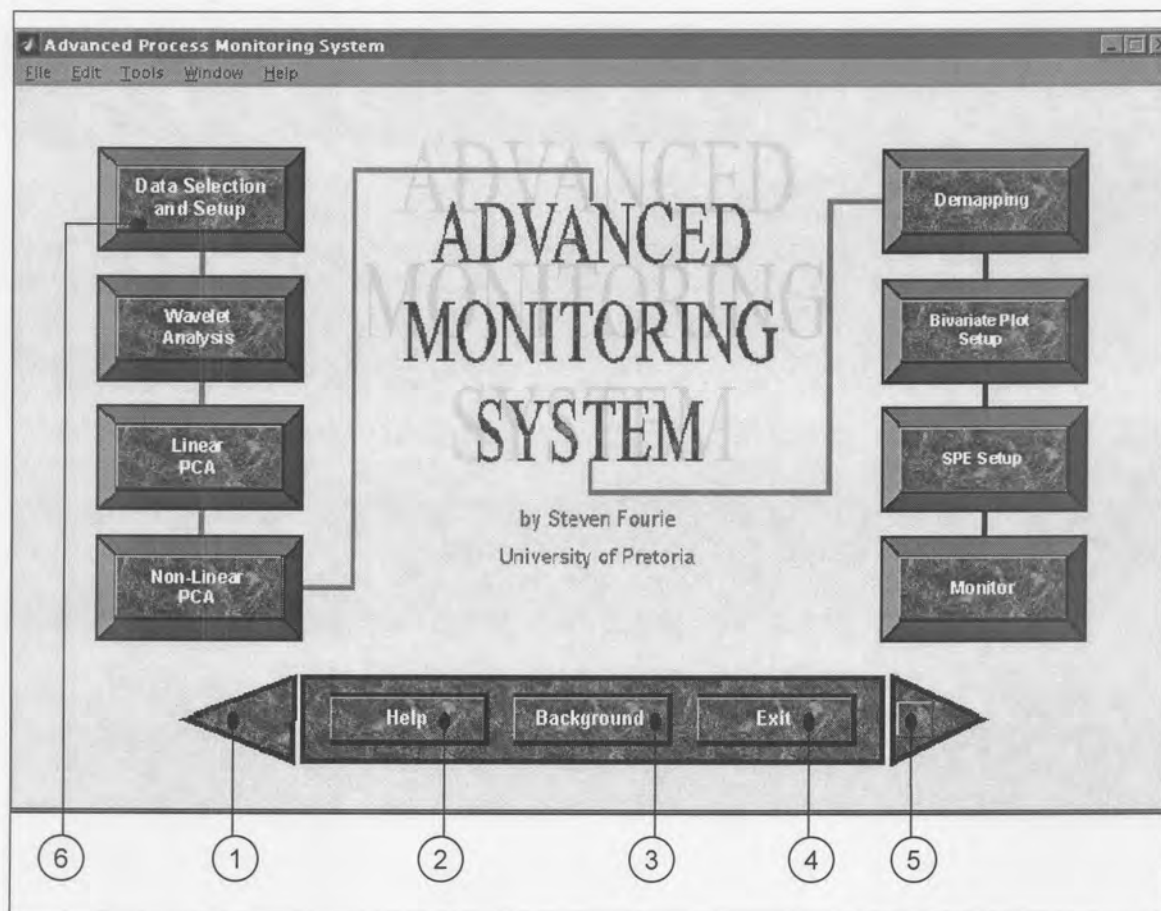


Figure 4.7. Main Interface

## 5.1. Introduction

---

The joint implementation of multiresolution analysis, wavelet filtering and non-linear PCA for process performance monitoring and fault detection is illustrated by application to a nonlinear industrial process which, in this case, is applied to the process data from case study 1 in Chapter 3 which is representative of that widely seen in the chemicals manufacturing industries.. Details of the process, except the background information provided in Chapter 3, are withheld for commercial confidentiality reasons. For the same reason the data was standardized prior to illustration. The data setup procedure is the first step in the process monitoring setup sequence and is accessed from the main menu as discussed in chapter 4.

## 5.2. Data Features

---

It has been pointed out several times in the recent literature that chemical processes are becoming more heavily instrumented and the data is recorded more frequently (Wise et. al, 1990; Kresta et. al, 1991). This is creating a data overload, and the result is that a good deal of the data is 'wasted', i.e. no useful information is obtained from it. The problem is one of both compression and extraction. Generally, there is a great deal of correlated or redundant information in process measurements. This information must be compressed in a manner that retains the essential information and is more easily displayed than each of the process variables individually. Also, often essential information lies not in any individual process variable but in how the variables change with respect to one another, i.e. how they co-vary. In this case the information must be extracted from the data. Furthermore, in the presence of large amounts of noise, it would be desirable to take advantage of some sort of signal denoising. These concepts will be discussed in more detail in subsequent chapters.

## 5.3. Data Setup Interface

---

This interface is used to load the various data sets into the workspace and save it to the database for further processing.

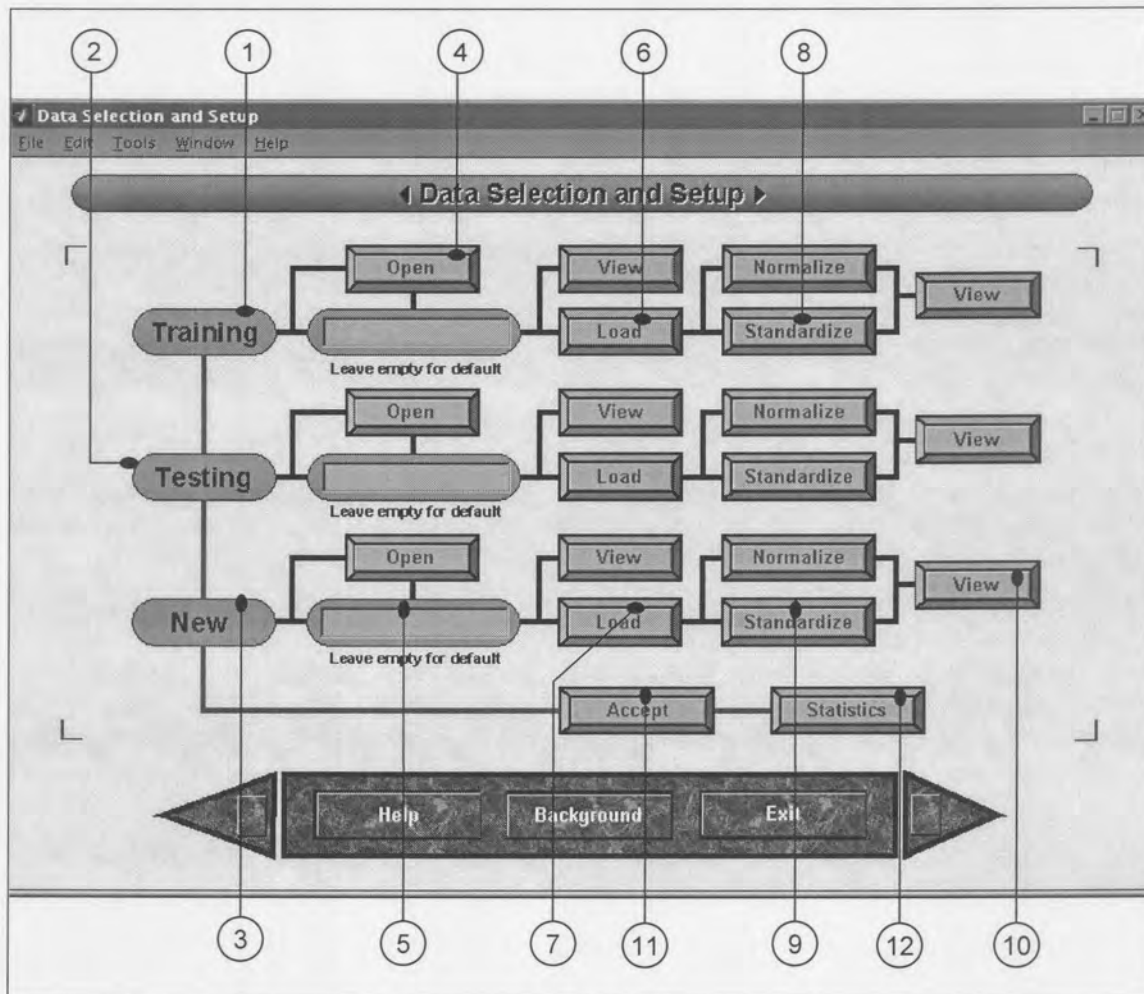


Figure 5.1. Data setup interface

Figure 5.1 Tags:

1. Training data set - this data is used for the actual training of the monitor system.
2. Testing data set - this data can be used for validation purposes when working with neural networks.
3. New data set - this data can be used to test the models in every section
4. Open existing mat data file
5. Display name of mat file or workspace variable
6. View original data
7. Load data into workspace
8. Normalise data
9. Standardise data



10. View normalised or standardised data
11. Write data to database
12. View some data statistics

The option exists to load the data from a mat-file by using button 4, a workspace variable or from the database. After the data has been chosen it can be loaded using button 7.

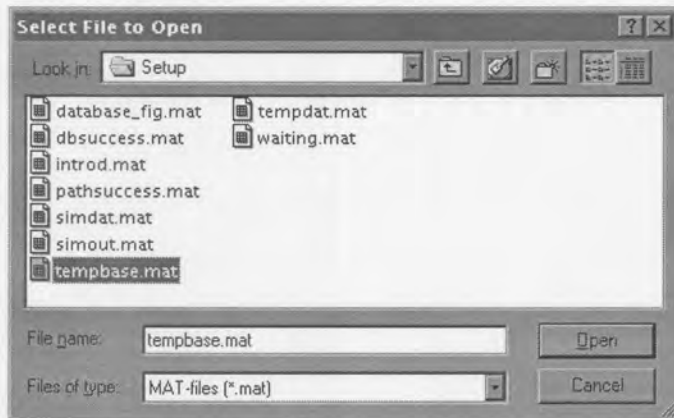


Figure 5.2. Open File interface

If no option is chosen (edit box 5 is left empty) then the data is loaded from the default database.

### 5.3. Data Viewer

The original data can then be viewed prior to normalization or standardization using button 6. This interface plots each individual variable separately as illustrated in Figure 5.3 and can be used to plot other variables in the workspace by changing the variable name.

Figure 5.3 Tags:

13. Variable number slider
14. Variable number display
15. Toggle between adding and removing the grid from the plot.
16. Toggle between hold and unhold. Use this if you need to plot more than one variable on the same graph.

17. Variable name. When the display window is opened it displays the default variable, in this case the variable `traindata`.

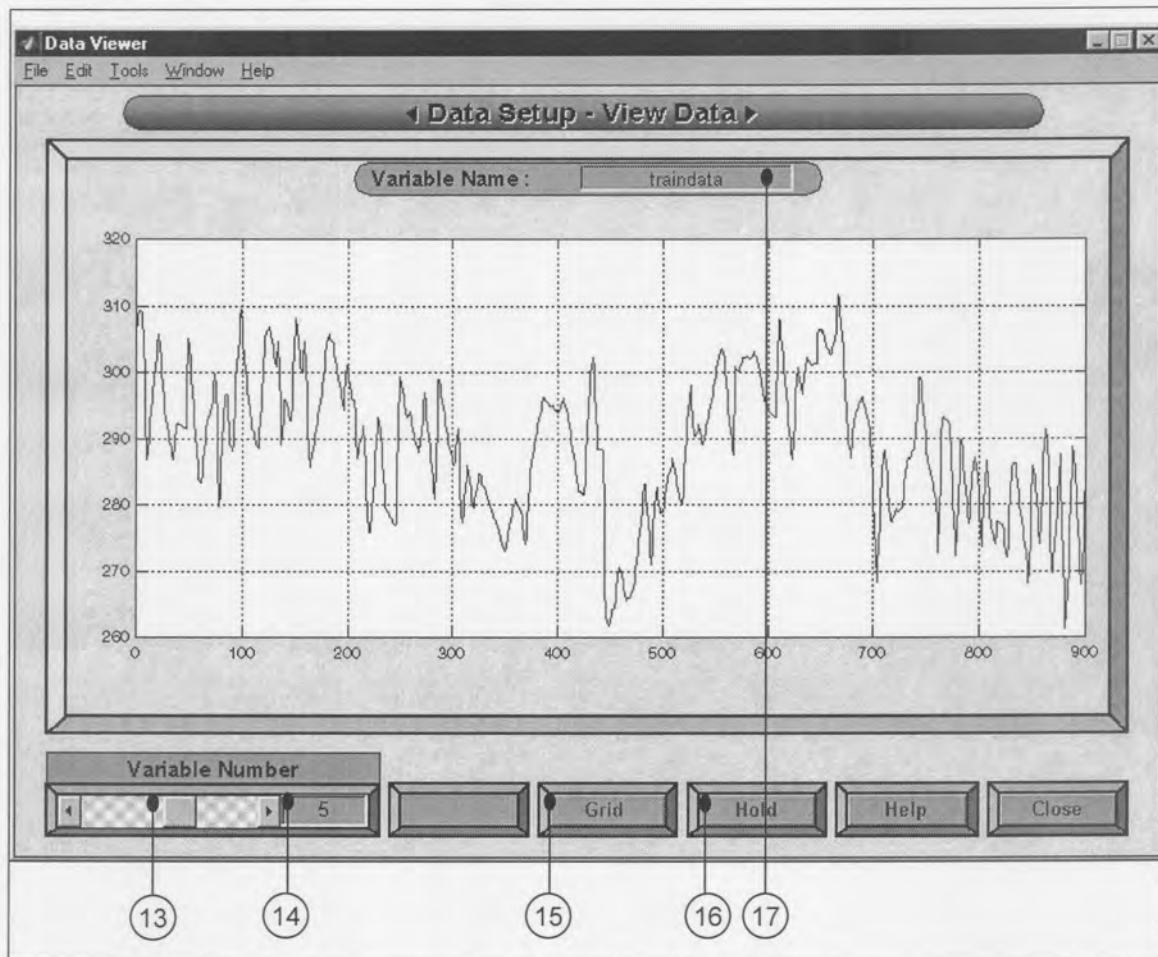


Figure 5.3. Data Viewer interface

## 5.5. Normalisation and Standardisation

The option exists to either normalise or standardise the data. The necessity for this becomes more apparent when the issue of principal component analysis is addressed. The normalised or standardised data can be viewed in a similar way as the original data using button 10.

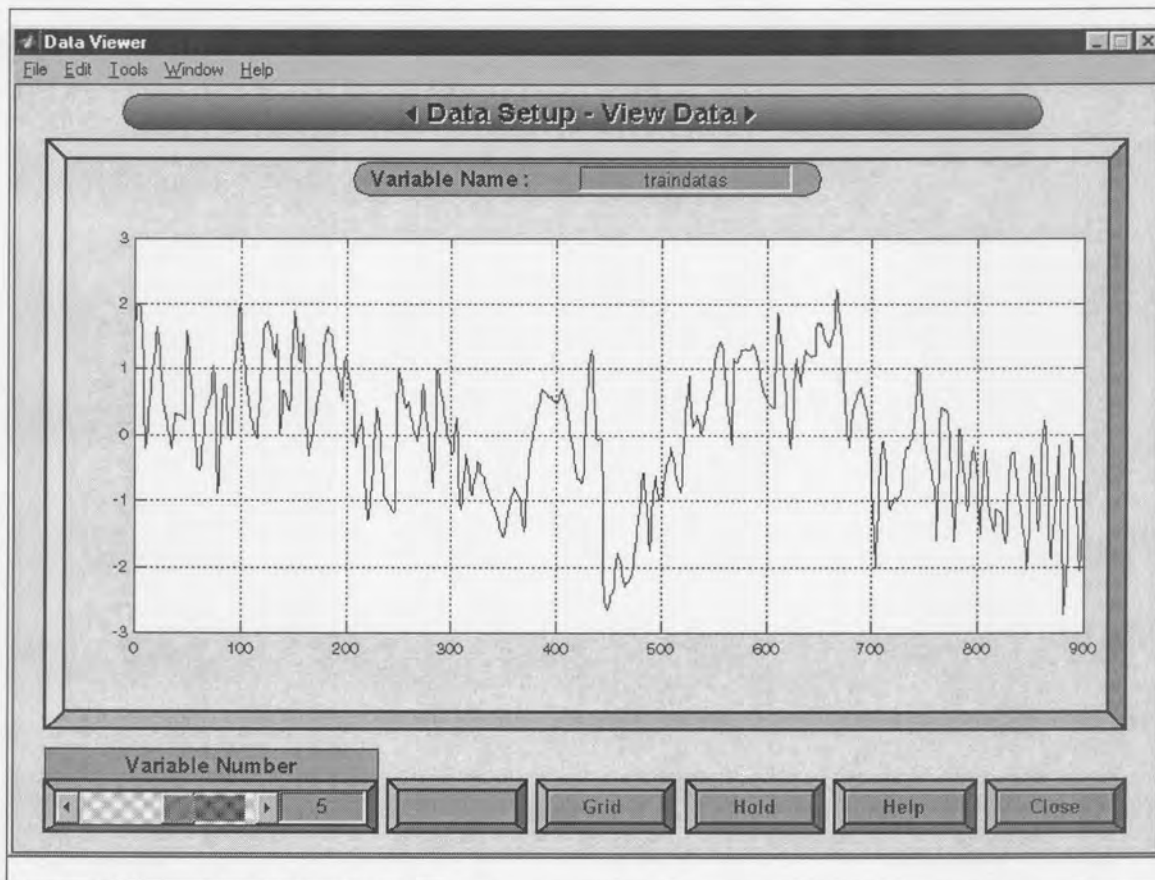


Figure 5.4. Standardized data

### 5.3. Statistics Viewer

The statistics viewer is accessed using button 12. Currently only the correlation coefficient can be calculated. However, other empty buttons are provided should the necessity of more statistic calculations be required.

Figure 5.5 Tags:

18. Choose from which data set the statistics are required.
19. Calculate the correlation coefficient of the specific data set.



Figure 5.5. Statistics Viewer: Correlation coefficients for training data set

## 5.5. Application

The first stage of the analysis was to manually carry out data pre-screening to identify and handle outliers, in-fill missing data, etc. Time-series plots of the process variables indicated that many of the measurements were corrupted by noise with some variables exhibiting sharp spikes. The sharp spikes were treated as outliers and were assumed to be due to missing data and faulty measurements. They were removed in Excel and replaced with the average of the five preceding values and five values following the outliers. Without appropriate pre-treatment of the data, the construction of a robust nominal process model for process performance monitoring is problematical and potentially worthless. Figure 5.7 is a plot of all eight variables on the same axes in order to show their relative values after removal of the outliers. Figure 5.8 is a plot of the same variables, but standardized. Both plots represent normal operation.

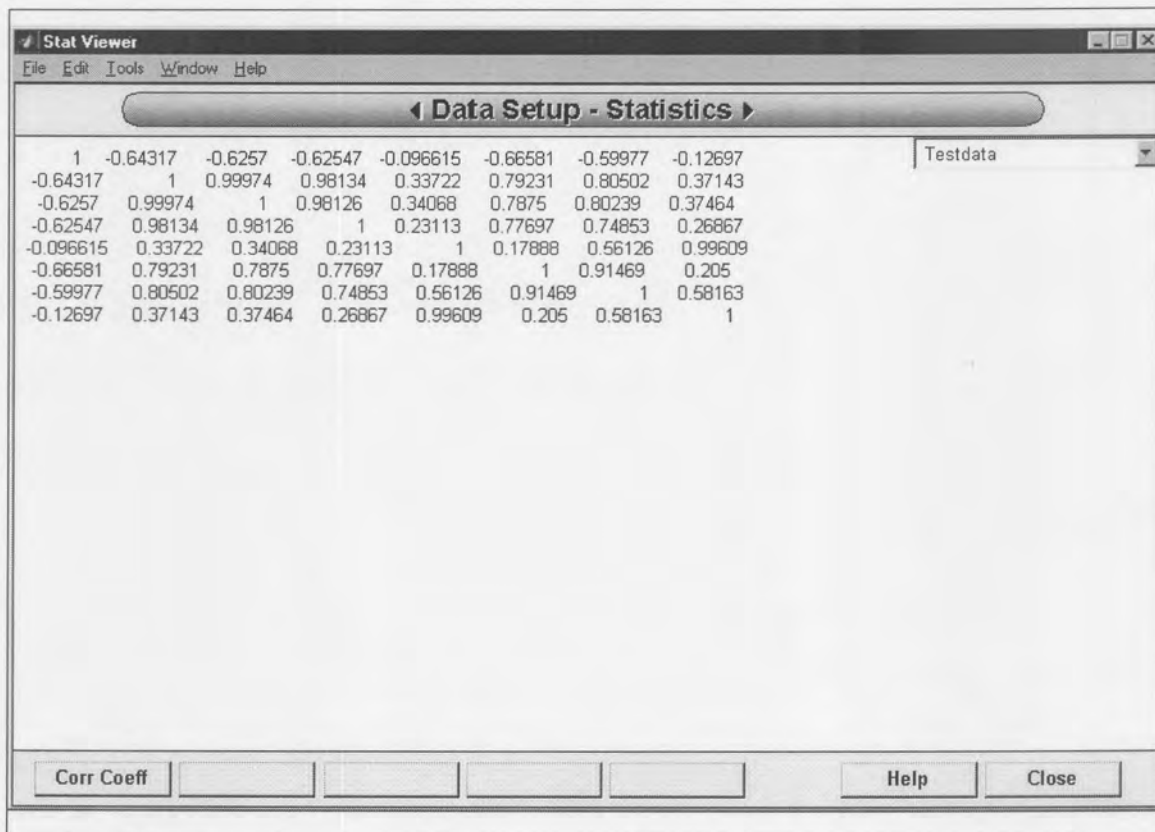


Figure 5.6. Statistics Viewer: Correlation coefficients for testing/validation data set

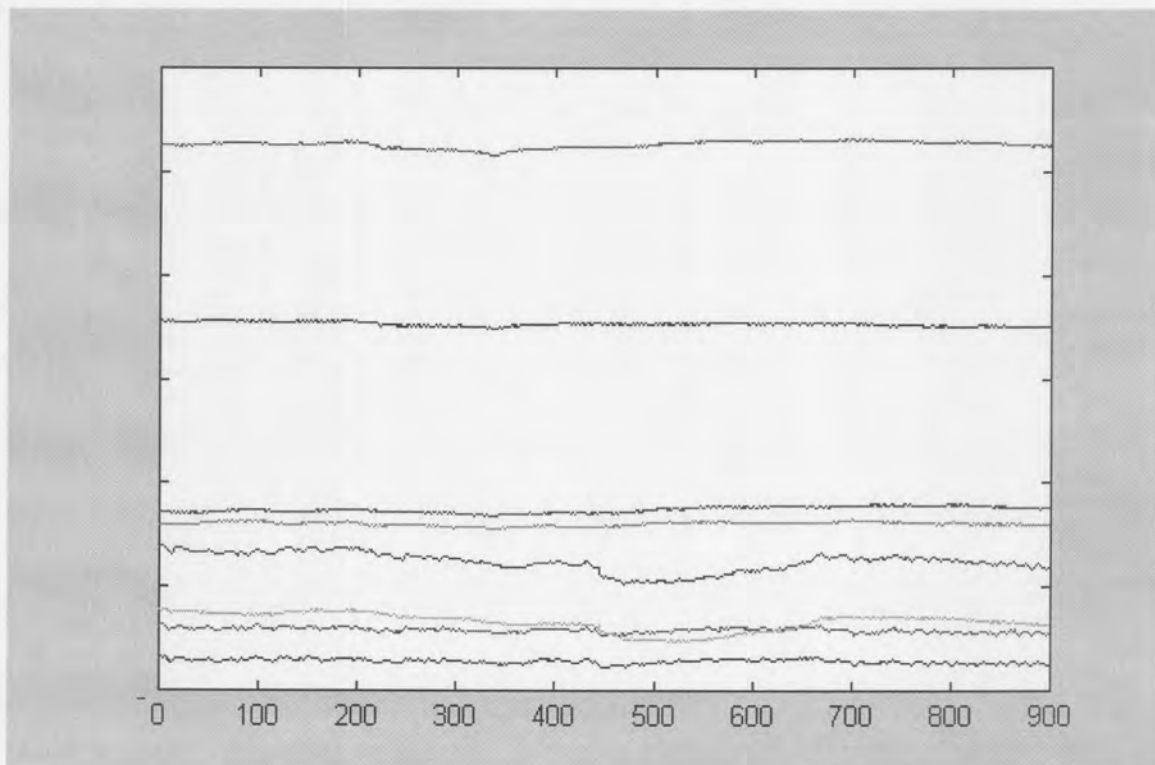


Figure 5.7. Plot of variables representing normal operation on the same axes.

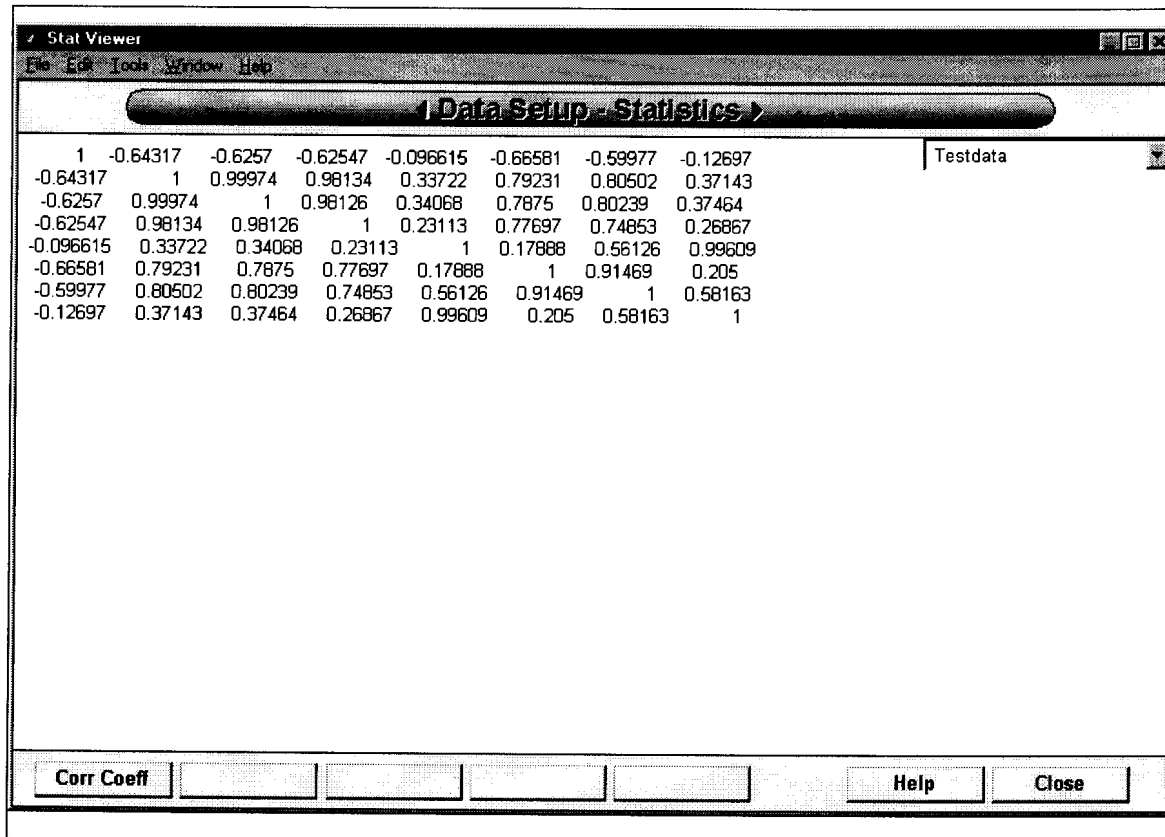


Figure 5.6. Statistics Viewer: Correlation coefficients for testing/validation data set

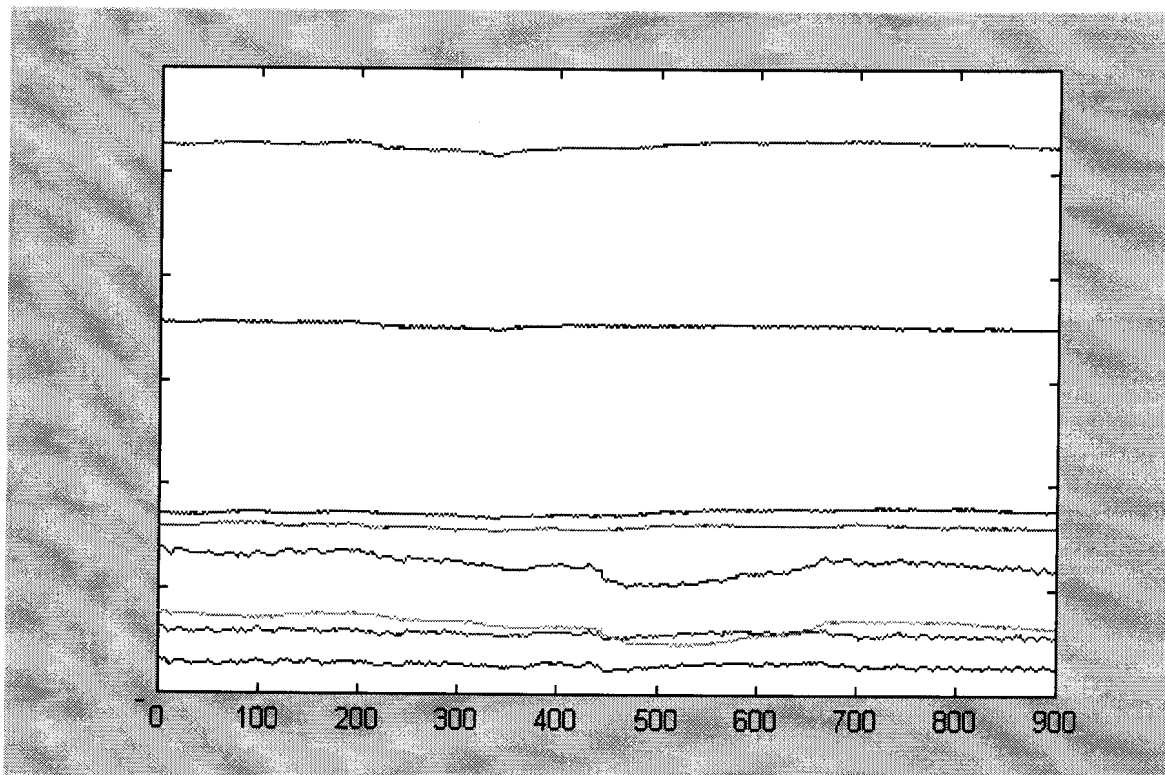


Figure 5.7. Plot of variables representing normal operation on the same axes.

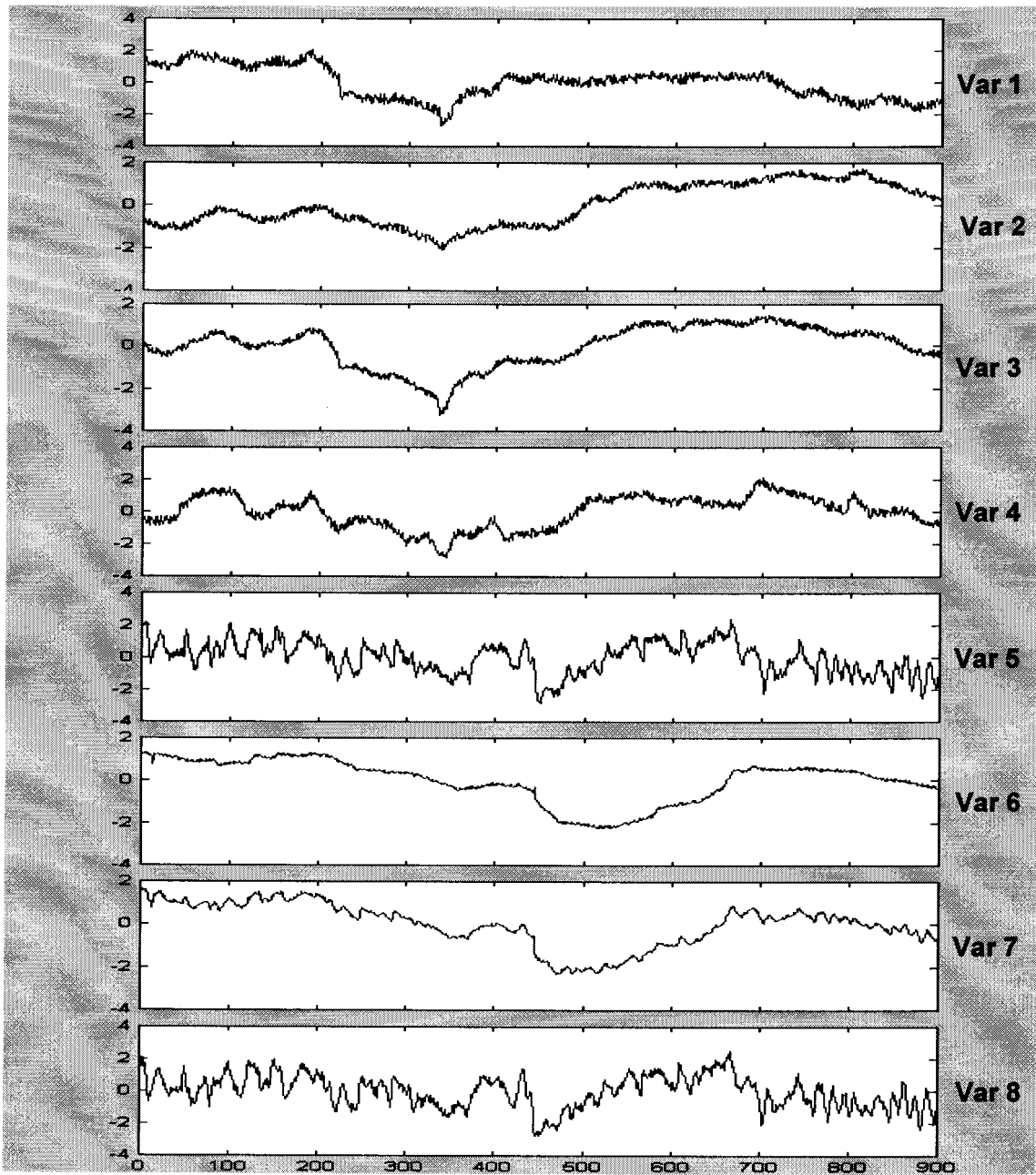


Figure 5.8. Standardized plot of variables representing normal operation

## 6.1. Introduction

---

Industrial data is synonymous with process measurement “noise”. Noise associated with the process measurements is known to have impact upon the robustness of the process model. It is therefore desirable to extract the “true” signal from the noise-corrupted data prior to carrying out any detailed statistical analysis. The most widely used forms of filtering algorithm found in the process industries include exponential and polynomial filters and the median filter. For data exhibiting small signal-to-noise ratios, heavy filtering can result in significant phase-shifts in the signal. A further limitation of some filters is that they cannot handle signal spikes efficiently or effectively. Finally, to implement some filtering algorithms it is necessary to have future values, e.g. in the median filter. In this respect they are unsuitable for on-line application. The wavelet transform addresses some of these limitations. In particular, through the application of wavelet de-noising, high-frequency noise as well as sharp spikes in the data can be removed without smoothing out the important features in the process data. The discrete wavelet transform is also an effective tool for reducing the amount of data.

## 6.2. Previous work on feature extraction of dynamic transients

---

This section briefly reviews some of the previous work on feature extraction. Feature extraction is basically a transformation of the data composing a dynamic trend to a lower dimensionality. An important property of such a transformation is that it is information preserving, that is, data is reduced by removing redundant components while preserving, in some optimal sense, information which is crucial for pattern discrimination (Chen et al., 1999).

Some researchers have adapted the episode representation technique originated by William (1986) to qualitative interpretation of transient signals. Janusz and Venkatasubramanian (1991) developed an episode approach that uses nine primitives to represent any plots of a function. Each primitive consists of the signs and the first and second derivatives of the function. Therefore, each primitive possesses the information about whether the function is positive or negative, increasing, decreasing, or not changing and the concavity. An episode is an interval described by only one primitive and the time interval the episode spans. A trend is a series of episodes that when grouped together can completely describe the dynamic feature. The approach automatically converts on-line sensor data to qualitative classification trees. Cheung



and Stephanopoulos (1990) developed the triangular-episode that uses seven triangle components to describe a dynamic trend. Bakshi and Stephanopoulos (1994, 1996) used wavelet decomposition of functions in different scales and zero-crossing of wavelet derivatives to find the inflections of decomposition. In this way, episodes can be identified automatically by computers. Based on episode analysis, dynamic trends can be interpreted as symbolic representations. The main idea of dynamic trend interpretation using episode approaches is to classify a trend such as increasing or decreasing pieces. This interpretation is sometimes not enough and inadequate in process analysis. Furthermore, there is no noise filtering in any of the episode based approaches, which significantly limits the trend representation and identification capability.

Whiteley and Davis (1992) applied back-propagation neural networks (BPNN) to convert numerical sensor data into symbolic abstractions. The major limitation of this approach is that it requires training data to train the model first.

The best known technique for signal analysis is probably the Fourier transform and it is therefore necessary to mention it here.

For a continuous function of period  $2P$ , the Fourier series is given by;

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left\{ a_n \cos\left(\frac{n\pi x}{P}\right) + b_n \sin\left(\frac{n\pi x}{P}\right) \right\} \quad (6.1)$$

where the Fourier coefficients are calculated by,

$$a_n = \frac{1}{P} \int_C^{C+2P} F(x) \cos\left(\frac{n\pi x}{P}\right) dx \quad (6.2)$$

$$b_n = \frac{1}{P} \int_C^{C+2P} F(x) \sin\left(\frac{n\pi x}{P}\right) dx \quad (6.3)$$

Fourier transform uses sine and cosine functions as its building blocks to decompose a function into a sum of frequency components. However, Fourier transform does not show how frequency varies with time, therefore it is not able to detect when a particular event took place. It means that the non-stationary feature of the signal is not captured. The short-time Fourier transform is able to overcome this limitation by sliding a window over the signal in time. However in time-frequency analysis of a non-stationary signal, there are two conflicting requirements. The window width must be long enough to give the desired frequency resolution but must also be short enough to lose track of time dependent events. While it is possible to optimise the design of window shapes, or

trade-off time and frequency resolution, there is a fundamental limitation on what can be achieved, for a given fixed window width (Dai, Joseph & Motard, 1994).

### 6.3. What is a wavelet?

---

Only a very brief introduction to wavelet transformation for signal processing will be presented. Only the main mathematical issues will be addressed to give some background to its calculation since it is too broad to cover here and won't facilitate a better understanding for this purpose.

According to Chen et al. (1999), wavelets can be viewed as an extension to Fourier analysis that is well-suited and designed to address the problem of non-stationary signals. Such signals are not well represented in time and frequency by the Fourier transform methods. One major advantage afforded by wavelets is the ability to perform local analysis — that is, to analyze a localized area of a larger signal. Wavelet analysis is capable of revealing aspects of data that other signal analysis techniques miss, aspects like trends, breakdown points, discontinuities in higher derivatives, and self-similarity. Further, because it affords a different view of data than those presented by traditional techniques, wavelet analysis can often compress or de-noise a signal without appreciable degradation. Wavelets offer a technique to localise events in both time and frequency and they can be applied to continuous and discrete-time problems and to two-dimensional, and in principle, to higher-dimensional data.

Another useful property of wavelets is that although they are not known to be exact eigenfunctions or principal components of any operators, they are approximate eigenfunctions of a large variety of operators (Wornell, 1990; Dickerman and Majumdar, 1994). Consequently, the wavelet coefficients of most stochastic processes are approximately decorrelated. The variance of the wavelet coefficients at different scales represents the energy of the stochastic process in the corresponding range of frequencies, and corresponds to its power spectrum. Thus, for an uncorrelated Gaussian stochastic process or white noise, the variance of the wavelet coefficients is constant at all scales, whereas for coloured noise, the variance decreases at finer scales.

A wavelet is a waveform of effectively limited duration that has an average value of zero. Compare wavelets with sine waves, which are the basis of Fourier analysis. Sinusoids do not have limited duration — they extend from minus to plus infinity. And where sinusoids are smooth and predictable, wavelets tend to be irregular and asymmetric as illustrated in Figure 6.1.

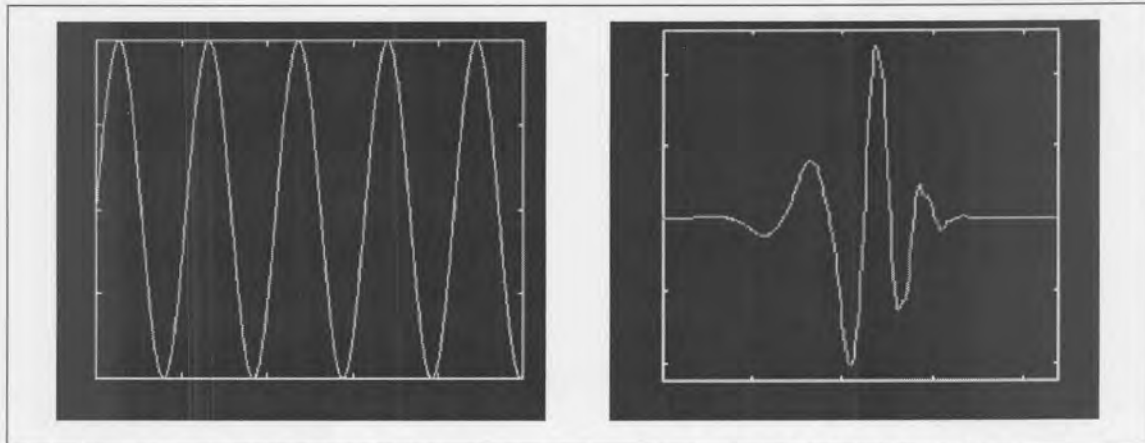


Figure 6.1. A comparison of the sine wave and the daubechies 5 wavelet

Fourier analysis consists of breaking up a signal into sine waves of various frequencies. Similarly, wavelet analysis involves the breaking up of a signal or time function into simple, fixed building blocks, termed wavelets (Rioul & Vetterli, 1991; Motard & Joseph, 1994; Chui, 1992). These building blocks are actually a family of functions which are derived from a single generating function called the mother wavelet by translation and dilation operations. Dilation, also known as scaling, compresses or stretches the mother wavelet and translation shifts it along the time axis. That is, the signal is mapped to a time-scale plane, as illustrated in Figure 6.2, that is analogous to the time-frequency plane used in the short-time Fourier transform.

The mother wavelet satisfies

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (6.4)$$

and the translation and scaling operations on  $\psi(t)$  create a family of functions,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (6.5)$$

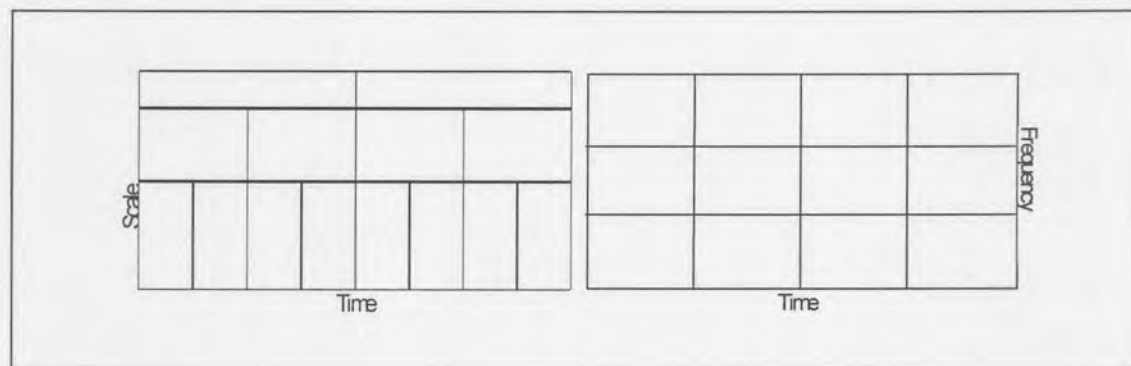


Figure 6.2. Time-scale characteristic of Wavelets (left) and STFT (right)

The parameter  $a$  is a scaling factor and stretches (or compresses) the mother wavelet. The parameter  $b$  is a translation along the time axis and simply shifts a wavelet and so delays or advances the time at which it is activated. Mathematically delaying a function  $f(t)$  by  $t_d$  is represented by  $f(t-t_d)$ . The factor  $1/\sqrt{a}$  is used to ensure the energy of the scaled and translated versions are the same as the mother wavelet.

The stretched and compressed wavelets through scaling operation are used to capture the different frequency components of the function being analysed. The translation operation, on the other hand, involves shifting of the mother wavelet along the time axis to capture the time information of the function to be analysed at a different position. In this way, a family of scaled and translated wavelets can be created using scaling and translation parameters  $a$  and  $b$ . This allows signals occurring at different times and having different frequencies to be analysed. In contrast to the short-time Fourier transform, which uses a single analysis window function, the wavelet transform can use short windows at high frequencies or long windows at low frequencies. Thus wavelet transform is capable of zooming in on short-lived high frequency phenomena and zooming-out on sustained low frequency phenomena. This is the main advantage of the wavelet over the short-time Fourier transform.

#### 6.4. Wavelet Analysis Methodology

---

In the introduction the effect of noise was mentioned. Noise is a phenomenon that affects all frequencies and appears in different forms such as high-frequency measurement noise and spikes due to process filters being purged and other process operations. However, the “true” signal tends to dominate the low-frequency area, especially in chemical processes. The traditional approach to filtering is to remove the high-frequency components above a certain level since they are associated with noise. Small wavelet coefficients at low scales (high-frequency area) are usually expected to be mainly due to noise components. The procedure for wavelet de-noising is as follows:

- Apply the wavelet transform to a noisy signal and obtain the noisy wavelet coefficients,
- Threshold those elements in the wavelet coefficients that are believed to be attributed to noise,
- Apply the inverse wavelet transform to the thresholded wavelet coefficients to obtain a de-noised signal.

Each of these issues will be addressed in more detail in subsequent sections.

## 6.5. The Discrete Wavelet Transform

### 6.5.1. INTRODUCTION

Wavelet transforms can be categorized into continuous and discrete. Continuous, in the context of wavelet transform, implies that the scaling and translation parameters  $a$  and  $b$  change continuously. However, calculating wavelet coefficients for every possible scale can represent a considerable effort and result in a vast amount of data. Therefore a discrete parameter wavelet transform is often used where we choose only a subset of scales and positions at which to make our calculations. The discrete parameter wavelet transform (DWT) uses scale and position values based on powers of two (so-called dyadic scales and positions) and makes the analysis much more efficient, whilst remaining accurate. To do this, the scale and time parameters are discretised as follows,

$$a = a_0^{-m/2}, \quad b = nb_0 a_0^n \quad (6.6)$$

The family of wavelets  $\{\psi_{m,n}(t)\}$  is given by

$$\psi_{m,n}(t) = a_0^{-m/2} \psi(a_0^{-m} t - nb_0) \quad (6.7)$$

resulting in a discrete wavelet transform (DWT) having the form

$$\begin{aligned} DWT_f(m,n) &= \langle f, \psi_{m,n} \rangle \\ &= a_0^{-m/2} \int_{-\infty}^{+\infty} f(t) \psi(a_0^{-m} t - nb_0) \end{aligned} \quad (6.8)$$

An efficient way to implement this scheme using filters was developed in Mallat (1989). This very practical filtering algorithm yields a fast wavelet transform — a box into which a signal passes, and out of which wavelet coefficients quickly emerge.

### 6.5.2. ONE-STAGE FILTERING: APPROXIMATIONS AND DETAILS

For many signals, the low frequency content is the most important part that gives a signal its identity. The high frequency content, on the other hand provides flavour or nuance. In wavelet analysis the high-scale, low frequency content is called the *approximation* and the low-scale, high frequency content is called the *detail*. The filtering process uses *lowpass* and *highpass* filters to decompose an original signal into the *approximation* and *detail* parts. The filtering process at its most basic level, which is a single-level decomposition, is illustrated in Figure 6.3 where the original signal,

$s = f(t)$ , passes through two complementary high- and lowpass filters and emerges as two signals.

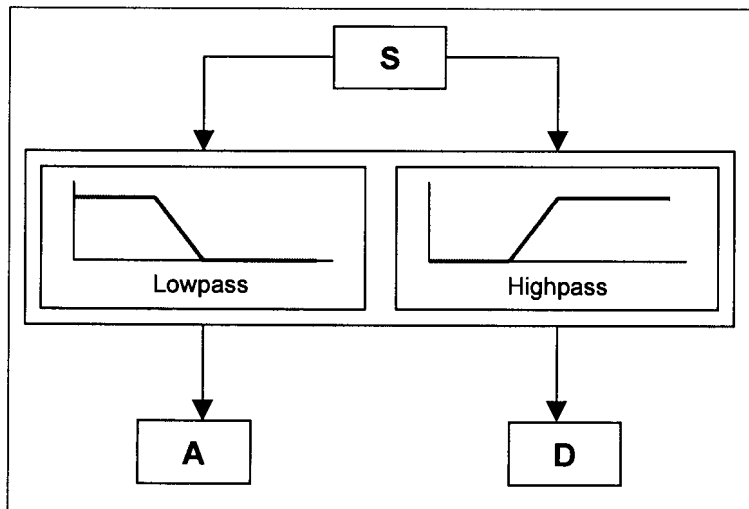


Figure 6.3. Basic discrete wavelet filtering

Unfortunately, if we actually perform this operation on a real digital signal, we end up with twice as much data as we started with. Suppose, for instance, that the original signal  $s$  consists of 1000 samples of data. Then the approximation and the detail will each have 1000 samples, for a total of 2000. However, it is not necessary to preserve all the outputs from the filters and therefore, to correct this problem, we introduce the notion of downsampling where we keep only the even components of the *lowpass* and *highpass* filter outputs and throw away every second data point. While doing this introduces aliasing, which is a type of error (Strang and Nguyen, 1995), in the signal components, it turns out we can account for this later on in the process. This procedure is illustrated in Figure 6.4(a) and (b).

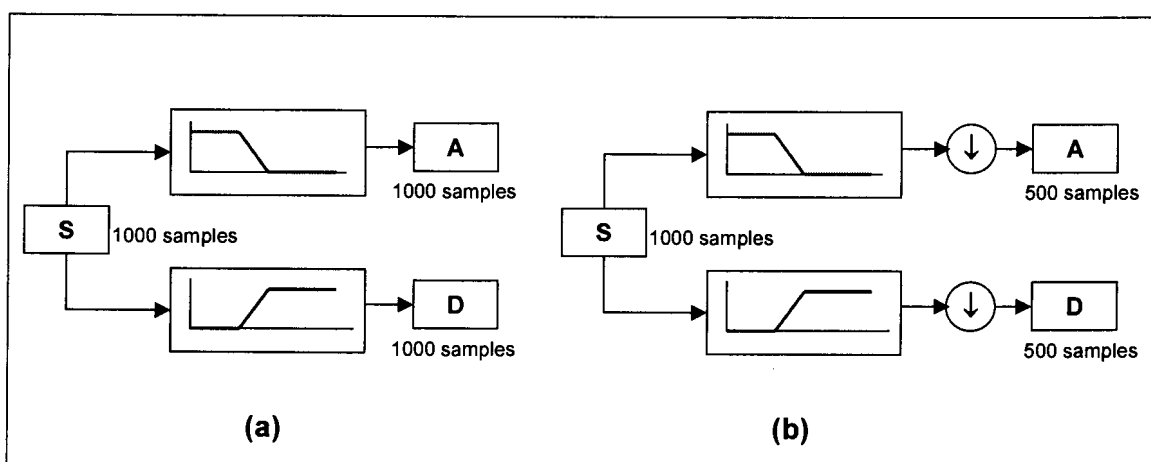


Figure 6.4(a) Wavelet decomposition without downsampling, and (b) with downsampling.

The process in Figure 6.4(b), which includes downsampling, produces discrete wavelet transform (DWT) coefficients. The detail coefficients will consist mainly of the high-frequency noise, while the approximation coefficients will contain much less noise than does the original signal.

The actual lengths of the detail as well as the approximation coefficient vectors will be slightly more than half the length of the original signal. This has to do with the filtering process, which is implemented by convolving the signal with a filter. The convolution “smears” the signal, introducing several extra samples into the result.

### 6.5.3. MULTIPLE-LEVEL DECOMPOSITION

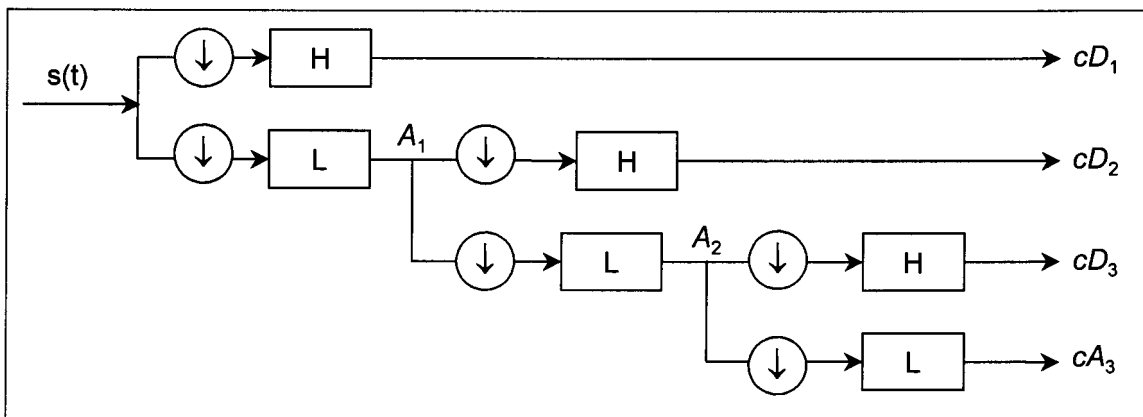


Figure 6.5 Multilevel decomposition tree (An octave band non-subsampled filter bank.)

The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower-resolution components. This is called the wavelet decomposition tree, illustrated in Figure 6.5, which can yield valuable information.

Since the analysis process is iterative, in theory it can be continued indefinitely. In reality, the decomposition can proceed only until the individual details consist of a single sample or pixel. In practice, you’ll select a suitable number of levels based on the nature of the signal, or on a suitable criterion such as entropy.

After calculating the wavelet coefficients, these coefficients can be thresholded to remove noise prior to reconstruction. Wavelet thresholding is discussed in more detail in Section 6.6, but it is worth noting that this step is applied after calculating the wavelet coefficients.

#### 6.5.4. WAVELET RECONSTRUCTION

The process of assembling the components back into the original signal with no loss of information is called reconstruction, or synthesis. The mathematical manipulation that affects synthesis is called the inverse discrete wavelet transform (IDWT). Where wavelet analysis involves filtering and downsampling, the wavelet reconstruction process consists of upsampling and filtering. Upsampling is the process of lengthening a signal component by inserting zeros between samples.

The filtering part of the reconstruction process is crucial since achieving perfect reconstruction of the original signal depends on the choice of filters. In the case of a discrete wavelet transform, reconstruction of the original signal is not guaranteed. Recall that the downsampling of the signal components performed during the decomposition phase introduces a distortion called aliasing. It turns out that by carefully choosing filters for the decomposition and reconstruction phases that are closely related (but not identical), we can “cancel out” the effects of aliasing. This was the breakthrough made possible by the work of Daubechies (1992) who developed conditions under which  $\{\psi_{m,n}\}$  forms an orthonormal basis. A technical discussion of how to design these filters can be found in p. 347 of the book *Wavelets and Filter Banks*, by Strang and Nguyen (1995). Usually,  $a_0 = 2$  and  $b_0 = 1$  are used, although any values can be used. In this case, both the transform and reconstruction are complete because the family of wavelets form an orthonormal basis. The low- and highpass decomposition filters (L and H), together with their associated reconstruction filters (L' and H'), form a system of what is called quadrature mirror filters.

#### 6.5.5. RECONSTRUCTING APPROXIMATIONS AND DETAILS

So it is possible to reconstruct the original signal from the coefficients of the approximations and details. It is also possible to reconstruct the approximations and details themselves from their coefficient vectors. As an example, let's consider how we would reconstruct the first-level approximation  $A_1$  from the coefficient vector  $cA_1$ . We pass the coefficient vector  $cA_1$  through the same process we used to reconstruct the original signal. However, instead of combining it with the level-one detail  $cD_1$ , we feed in a vector of zeros in place of the details as in Figure 6.6.

The process yields a reconstructed approximation  $A_1$ , which has the same length as the original signal  $s$  and which is a real approximation of it. Similarly, we can reconstruct the first-level detail  $D_1$ , using the analogous process illustrated in Figure 6.7.



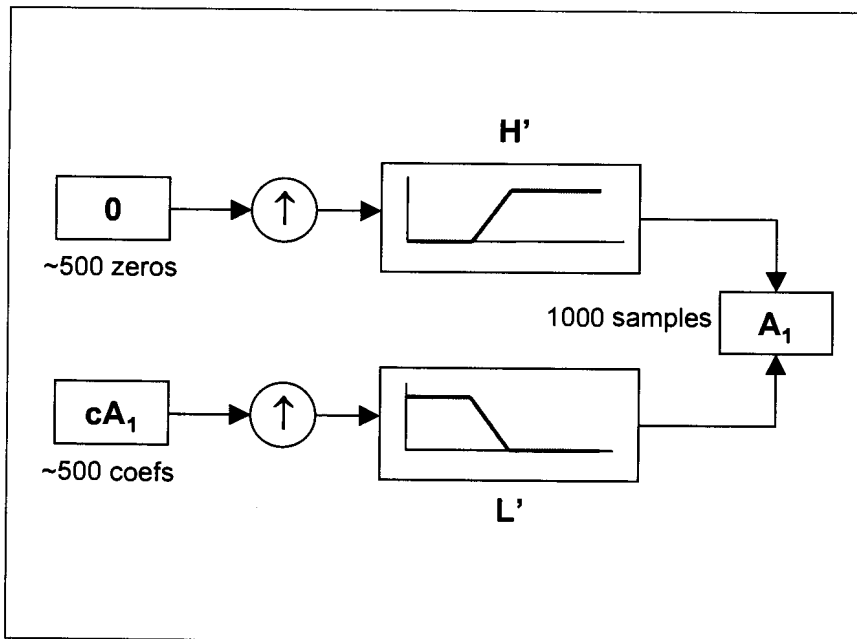


Figure 6.6. First-level reconstruction of approximation

The reconstructed details and approximations are true constituents of the original signal. In fact, we find when we combine them that:

$$A_1 + D_1 = S$$

Note that the coefficient vectors  $cA_1$  and  $cD_1$  — because they were produced by downsampling, contain aliasing distortion, and are only half the length of the original signal — cannot directly be combined to reproduce the signal. It is necessary to reconstruct the approximations and details before combining them.

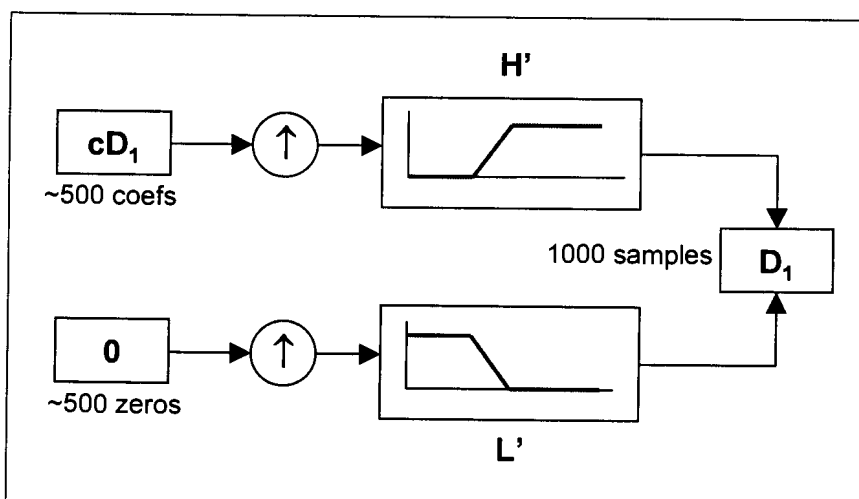


Figure 6.7. First-level reconstruction of detail

Extending this technique to the components of a multi-level analysis, we find that similar relationships hold for all the reconstructed signal constituents. That is, there are several ways to reassemble the original signal:

$$S = A_1 + D_1$$

$$S = A_2 + D_2 + D_1$$

$$S = A_3 + D_3 + D_2 + D_1$$

### 6.5.6. FILTERS USED TO CALCULATE THE DWT AND IDWT

For an orthogonal wavelet, in the multiresolution framework, we start with the scaling function  $\phi$  and the wavelet function  $\psi$ . One of the fundamental relations is the twin-scale relation (dilation equation or refinement equation):

$$\frac{1}{2}\phi\left(\frac{x}{2}\right) = \sum_{n \in \mathbb{Z}} w_n \phi(x-n) \quad (6.9)$$

All the filters used in DWT and IDWT are intimately related to the sequence  $(w_n)_{n \in \mathbb{Z}}$ . Clearly if  $\phi$  is compactly supported, the sequence  $(w_n)$  is finite and can be viewed as a filter. The filter  $W$ , which is called the scaling filter (non-normalized), is:

- Finite Impulse Response (FIR)
- of length  $2N$
- of sum 1
- of norm  $\frac{1}{\sqrt{2}}$
- a low-pass filter

From filter  $W$ , we define four FIR filters, of length  $2N$  and of norm 1, organized as in Table 6.1.

The four filters are computed using the scheme in Figure 6.8 where  $qmf$  is such that  $H'$  and  $L'$  are quadrature mirror filters (i.e.,  $H'(k) = (-1)^k L'(2N-1-k)$ ). Note that  $wrev$  flips the filter coefficients so  $H$  and  $L$  are also quadrature mirror filters.

Table 6.1. Filter representation

Filters	Low-pass	High-pass
Decomposition	$L$	$H$
Reconstruction	$L'$	$H'$

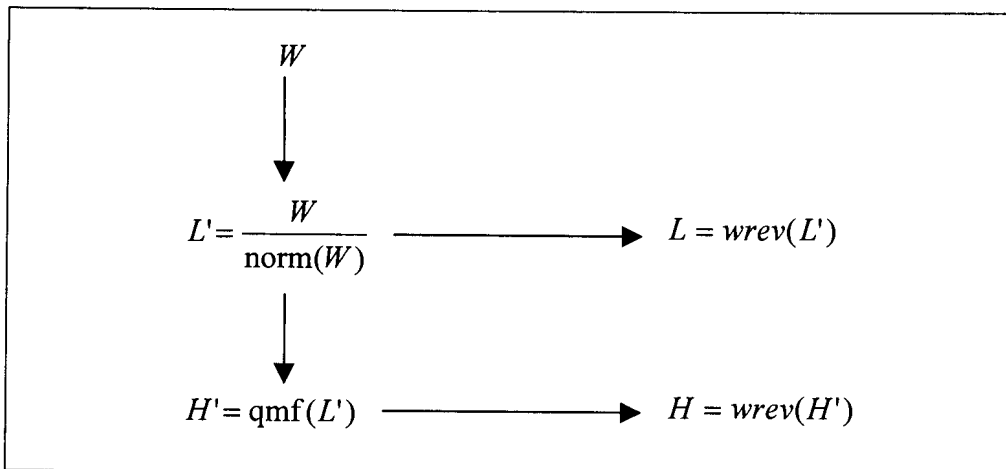


Figure 6.8. Wavelet filter computing scheme

## 6.6. Wavelet Denoising through Thresholding

To address the issue of noise, wavelet de-noising can be applied where the wavelet coefficients  $cA_i$  and  $cD_i$  are thresholded in order to remove noise components contained in the signal and thus also in the wavelet coefficients.

Multiscale rectification using wavelets is based on the observation that random errors in a signal are present over all the coefficients, while deterministic changes get captured in a small number of relatively large coefficients. Thus, stationary Gaussian noise may be removed by suppressing coefficients smaller than a selected value (Donoho et al., 1995).

Donoho and coworkers have studied the statistical properties of wavelet thresholding and have shown that for a noisy signal of length  $n$ , the rectified signal will have an error of order  $\log n$  of the error between the error-free signal and the signal rectified with a-priori knowledge about the smoothness of the underlying signal (Donoho and Johnstone, 1994).

Generally speaking, wavelet thresholding can be divided into two categories: *global thresholding* and *level-dependent thresholding*. If the threshold value is denoted as  $\lambda$ , then in global thresholding a single value of  $\lambda$  is selected and is applied globally to all empirical wavelet coefficients above a certain frequency level. For *leveldependent thresholding*, a different threshold value  $\lambda_j$  can be selected for the wavelet coefficient at level  $j$ . This approach is necessary when the noise in the data is non-stationary and/or correlated and is the approach used in this study.

Selecting the proper value of the threshold is a critical step in the rectification process and a number of different methods for selecting appropriate threshold values for wavelet denoising have been proposed in the literature (e.g. Donoho and Johnstone, 1994,1995; Donoho, 1995; Donoho et al., 1996; Hall et al., 1996; Hall and Patil, 1996; Nason, 1996).

Generally, wavelet denoising methods are based on either a hard or a soft thresholding approach. If the threshold value is denoted as  $\lambda$ , hard thresholding is given by Equation 6.10, whilst soft thresholding is given by Equation 6.11. Soft thresholding shrinks the value of the wavelet coefficients towards zero (eliminates coefficients) if they are above a certain threshold and hard thresholding if they are smaller.

$$\delta_{\lambda}^H(x) = \begin{cases} x & \text{if } |x| > \lambda \\ 0 & \text{otherwise} \end{cases} \quad (6.10)$$

$$\delta_{\lambda}^S(x) = \begin{cases} x - \lambda & \text{if } x > \lambda \\ 0 & \text{if } |x| \leq \lambda \\ x + \lambda & \text{if } x < -\lambda \end{cases} \quad (6.11)$$

Hard thresholding can lead to better reproduction of peak heights and discontinuities, but at the price of occasional artifacts that can roughen the appearance of the rectified signal, while soft thresholding usually gives better visual quality of rectification and less artifacts. An artifact, which is not present in the original signal, is created in the reconstructed signal when the wavelet function used to represent a feature in the signal and the feature itself does not align. Such artifacts are due to a localized Gibbs phenomenon which is caused by the lack of translational invariance in orthonormal wavelet decomposition.

Two factors that can influence the performance of wavelet thresholding are considered in the selection of the threshold values, these are the sample size  $N$  and the noise level  $\sigma$ . For good visual quality of the rectified signal, the VisuShrink method determines the threshold as

$$t_j = \sigma_j \sqrt{2 \log N} \quad (6.12)$$

where  $N$  is the signal length and  $\sigma_j$  is the standard deviation of the errors at scale  $j$ .

In practice, the value of the standard deviation of the noise in the data,  $\sigma$ , is unknown and is replaced by an estimate  $\hat{\sigma}$ . Donoho and Johnstone (1995) proposed the use of the median of the absolute deviation (MAD) of the wavelet coefficients at the finest level (level=1):

$$\hat{\sigma} = \text{median}(|w_{1,i}|) / 0.6745 \quad (6.13)$$

where  $i = 0, \dots, 2^{J-1} - 1$ ,  $J = \log_2(N)$ . The median absolute deviation of the coefficients is a robust estimate of  $\sigma$ . When coloured noise is suspected, the noise level  $\sigma$  needs to be estimated level-by-level using a similar kind of strategy and the threshold values also need to be modified according to the level-dependent estimation of the noise.

Wavelet de-noising is able to remove as much noise as required but not at the expense of smoothing out any real fine-scale features (Ogden, 1997). The advantage of spatially adaptive methods such as wavelet de-noising is that they perform close to the optimum across the whole range of noise levels, no matter the smoothness of the signal. On the other hand, the best performing median filter is almost as efficient as the wavelet de-noising methods at relatively high signal to noise ratios, if the window size is selected appropriately. However, for low signal-to-noise ratios, phase-shift may result. Moreover, future values are needed to apply the median filtering algorithm, thus making it unsuitable for on-line application and therefore wavelet de-noising remains a better alternative.

Wavelet-based multiscale rectification is a very effective approach for denoising signals contaminated by white, as well as correlated Gaussian noise. If the traditional wavelet decomposition algorithm is applied to a signal with non-Gaussian errors, outliers will be present at multiple scales in both the scaled and detailed signals, and large coefficients corresponding to outliers get confused with those corresponding to important features. Thus, wavelet thresholding is not effective in eliminating non-Gaussian errors. This limitation may be overcome by combining wavelet thresholding with multiscale median filtering as in the robust multiscale rectification technique (Bruce et al., 1994).

## 6.7. Algorithms

This section takes you through the most important steps of the wavelet analysis and de-noising algorithms in view of the actual implementation. It considers in more detail the magnitude and nature of the different calculated values and signals.

Starting out with a signal  $s$  of length  $N$ , the DWT consists of  $\log_2 N$  stages at most. The first step produces, starting from  $s$ , two sets of coefficients: approximation coefficients  $cA_1$  and detail coefficients  $cD_1$ . These vectors are obtained by convolving  $s$  with the low-pass filter  $L$  for approximation, and with the high-pass filter  $H$  for detail, followed by dyadic decimation. The first step is illustrated by Figure 6.9.

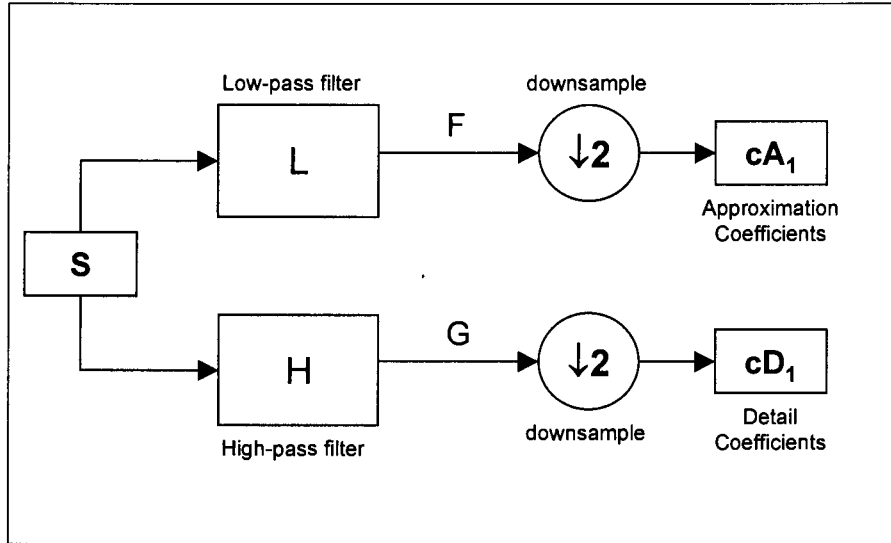


Figure 6.9. First step in the wavelet analysis algorithm

The length of each filter is equal to  $2N$ . If  $n = \text{length}(s)$ , the signals  $F$  and  $G$ , are of length  $N + 2N - 1$  and then the coefficients  $cA_1$  and  $cD_1$  are of length

$$\text{floor}\left(\frac{n-1}{2}\right) + N.$$

The next step splits the approximation coefficients  $cA_1$  in two parts using the same scheme, replacing  $s$  by  $cA_1$ , and producing  $cA_2$  and  $cD_2$ , and so on as illustrated in Figure 6.5. So the wavelet decomposition of the signal  $s$  analyzed at level  $j$  has the following structure:  $[cA_j, cD_j, \dots, cD_1]$ .

The next step involves applying level-dependent thresholding to the coefficients so that the wavelet decomposition of the signal  $s$  analyzed at level  $j$  now has the following structure:  $[cA'_j, cD'_j, \dots, cD'_1]$ , where  $cA'_j$  and  $cD'_j$  are the thresholded approximation and detail wavelet coefficients.

Conversely, starting from  $cA'_j$  and  $cD'_j$ , the IDWT reconstructs  $A'_{j-1}$ , the reconstructed approximation signal, inverting the decomposition step by inserting zeros and

convolving the results with the reconstruction filters as depicted in Figure 6.10 where *wkeep* means taking the central part of *U* with the convenient length.

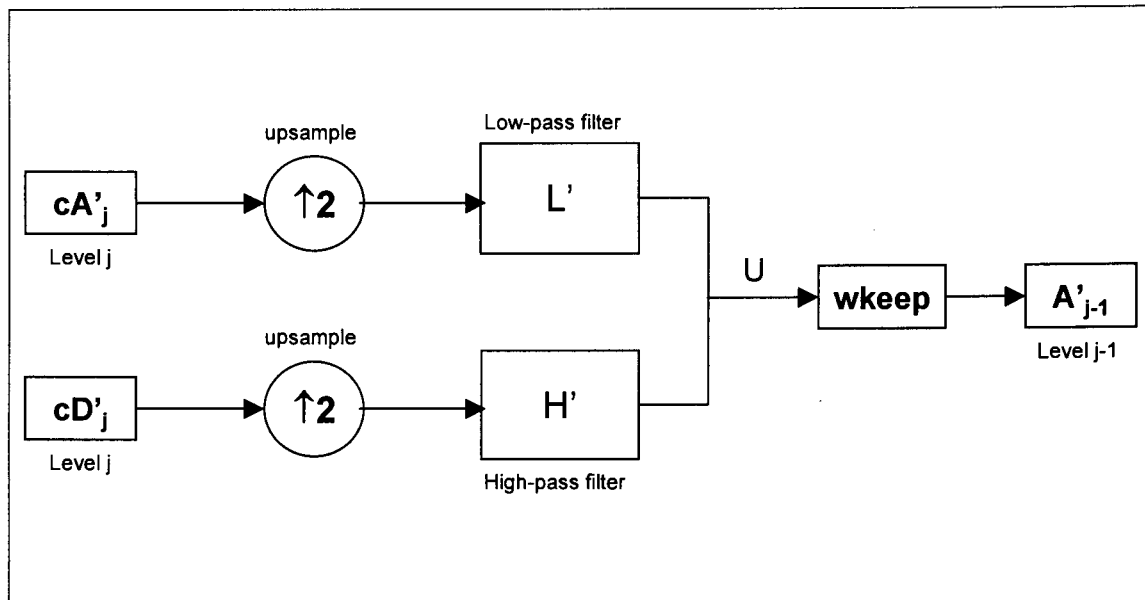


Figure 6.10. Reconstruction of the wavelet coefficients using reconstruction filters

## 6.8. On-Line Multiscale Rectification

Existing nonlinear rectification techniques do perform better than linear filters for a broad variety of signals. However, a significant disadvantage of these nonlinear multiscale methods is that they cannot be implemented online. In general wavelet filters are noncausal in nature and require future measured data for calculating the current wavelet coefficient. This introduces a time delay in the computation that increases at coarser scales and smoother filters. This time delay may be overcome in a rigorous manner by using special wavelets at edges that eliminate boundary errors while being orthonormal to the other wavelets (Cohen, et al., 1993). These boundary corrected filters are causal and require no information about the future to compute wavelet coefficients at the signal end points. Another reason for restricting the wavelet-based methods to off-line use is the dyadic discretization of the wavelet parameters, which requires a signal of dyadic length for the wavelet decomposition.

A signal containing a dyadic number of measurements can be decomposed as shown in Figure 6.11(a). In contrast, if the number of measurements is odd, the last point cannot be decomposed without a time delay as shown in Figure 6.11(b). In many applications such a time delay is unacceptable. Consequently, this section describes an online method for multiscale rectification (OLMS), where absolutely no time delay is allowed.

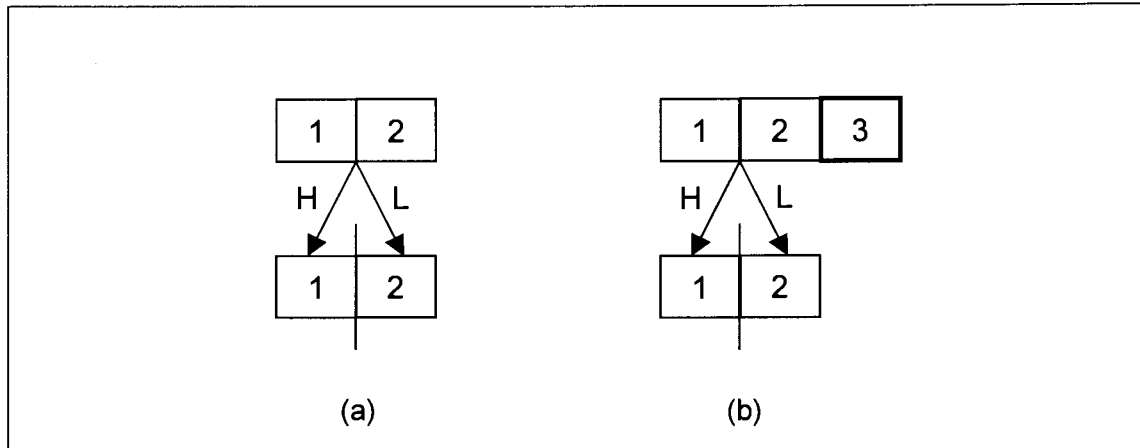


Figure 6.11. Time delay introduced due to dyadic length requirement in wavelet decomposition

On-line multiscale rectification is based on multiscale rectification of data in a moving window of dyadic length, as shown in Figure 6.12. The OLMS methodology can be summarized as follows:

- (1) Decompose the measured data within a window of dyadic length using a causal boundary corrected wavelet filter.
- (2) Threshold the wavelet coefficients and reconstruct the rectified signal.
- (3) Retain only the last data point of the reconstructed signal for on-line use.
- (4) When new measured data are available, move the window in time to include the most recent measurement while maintaining the maximum dyadic window length.

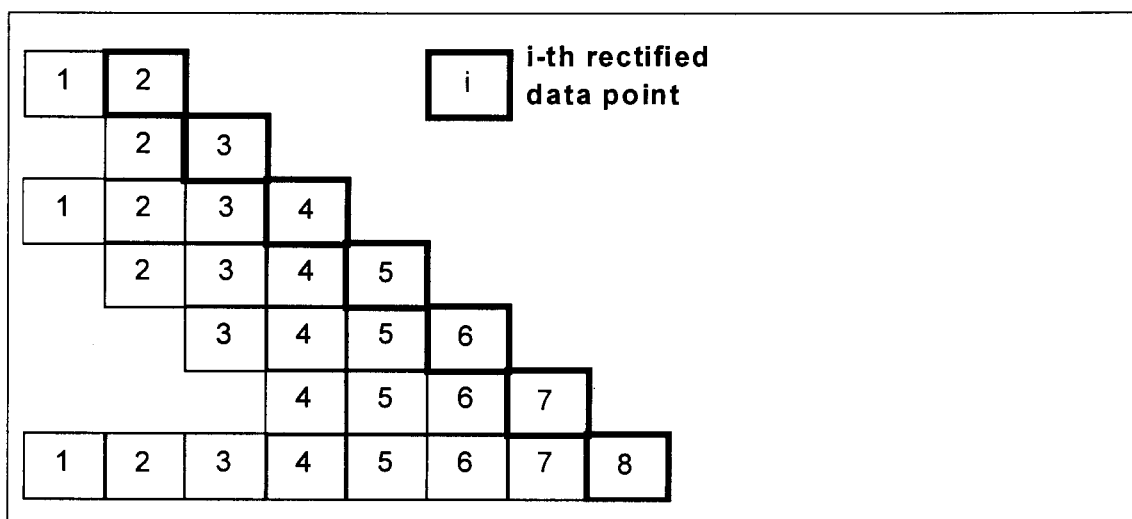


Figure 6.12. OLMS rectification



The measurements in each window are rectified by the wavelet thresholding approach of Donoho et al. (1995) described in the previous section. This simple approach is very effective compared to the single-scale techniques and retains the benefits of the wavelet decomposition in each moving window, while allowing each measurement to be rectified on-line.

## 6.9. Practical issues of OLMS

---

Any filtering method requires typical data or information about the underlying signal and noise for selecting the filter parameters. In OLMS rectification the filter tuning parameters are the value of the threshold and the maximum depth of the wavelet decomposition. Other practical issues include selecting a wavelet and the maximum length of the moving window.

### 6.9.1. VALUE OF THRESHOLD

The threshold may be estimated by applying the Visushrink method (Donoho et al., 1995; Nason, 1996; Nounou and Bakshi, 1999) to the available measurements. For data corrupted by stationary errors, the threshold value stops changing much after an adequate number of measurements are available. Consequently, for stationary noise, the threshold may be estimated from the measurements until the change is below a user-specified value. This approach for estimating the threshold cannot be performed recursively due to the median operator used in Equation 6.9 and will require storage of a large number of measurements.

### 6.9.2. DEPTH OF DECOMPOSITION

Thresholding wavelet coefficients at very coarse scales may result in the elimination of important features, whereas thresholding only at very fine scales may not eliminate enough noise. Therefore, the depth of wavelet decomposition needs to be selected to optimize the quality of the rectified signal. Empirical evidence suggests that a good initial guess for the decomposition depth is about half of the maximum possible depth, that is  $(\log_2(n))/2$  where  $n$  is the moving window length. However, a smaller depth might be more appropriate in OLMS rectification if a long boundary corrected filter with a large support is used in the decomposition since the filters at the two edges might overlap at very coarse scales. The depth may also be determined by cross-validation.

## 6.9.3. SELECTED WAVELET FILTER

The type, length, and nature of the wavelet filter used in OLMS affect the quality of the rectification. Since the OLMS rectification uses only the last rectified data point from each translated signal, it is crucial that only boundary corrected causal wavelet filters are used. If boundary corrected filters are not used, then the last point is among the least accurate ones due to the end effect errors. OLMS rectification using Daubechies second-order boundary corrected filters was used and results in smaller mean-square error than OLMS rectification using other simpler wavelet filters like the Haar wavelet.

## 6.10. Application

## 6.10.1. SOFTWARE SETUP

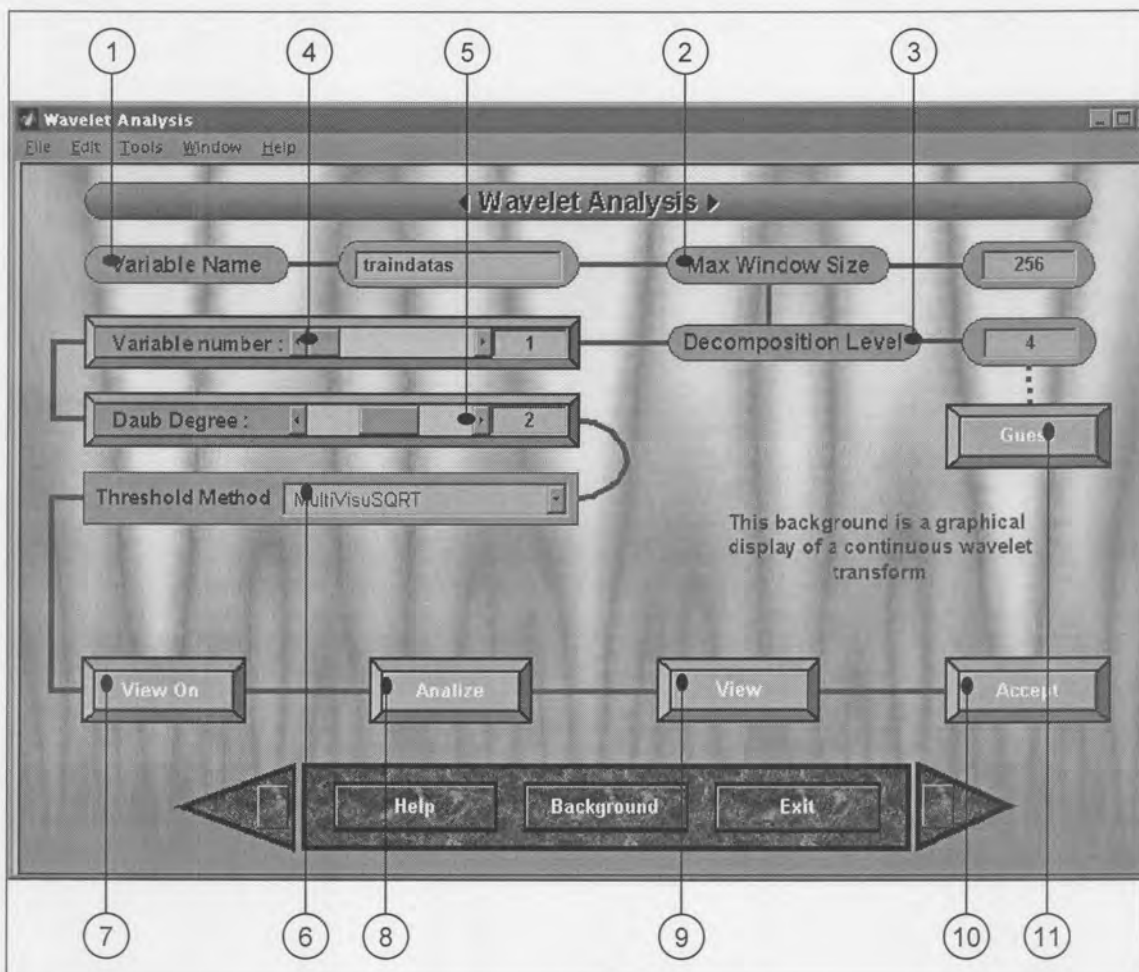


Figure 6.13. Wavelet analysis main interface

The wavelet analysis user interface is displayed by using the *Next* button on the data setup interface or it can be accessed via the main user interface. The options are related to the

theory discussed in the previous sections. Normally one would have to play around with different combinations of the parameters in order to select the best combination since there are no definite rules.

Figure 6.13 Tags:

1. Name of variable to which the wavelet transform should be applied. By default *traindatas* from the database is used. Any other variable name may be specified. However, it is important to know that the data should be normalized or standardized prior to applying the wavelet analysis.
2. Maximum dyadic (power of two) window size. For this application a maximum window size of 256 ( $2^8$ ) was used.
3. Multiresolution decomposition level.
4. Variable number. The wavelet analysis is applied to one variable at a time. The specific variable is specified via its column number in the data matrix.
5. Daubechies degree. In most cases the second (2) degree works the best.
6. Type of threshold to apply to the wavelet coefficients. Those methods with *multi* as prefix refer to level dependent thresholding.
7. Toggle between real-time viewer on an off. If the viewer is on, one is able to view the coefficients and reconstructed approximations and details as they are calculated as illustrated in Figure 6.14 and Figure 6.15.
8. Apply wavelet analysis. This starts the wavelet analysis process.
9. View reconstructed multiresolution data.
10. Write results to database.
11. Initial guess for decomposition level.

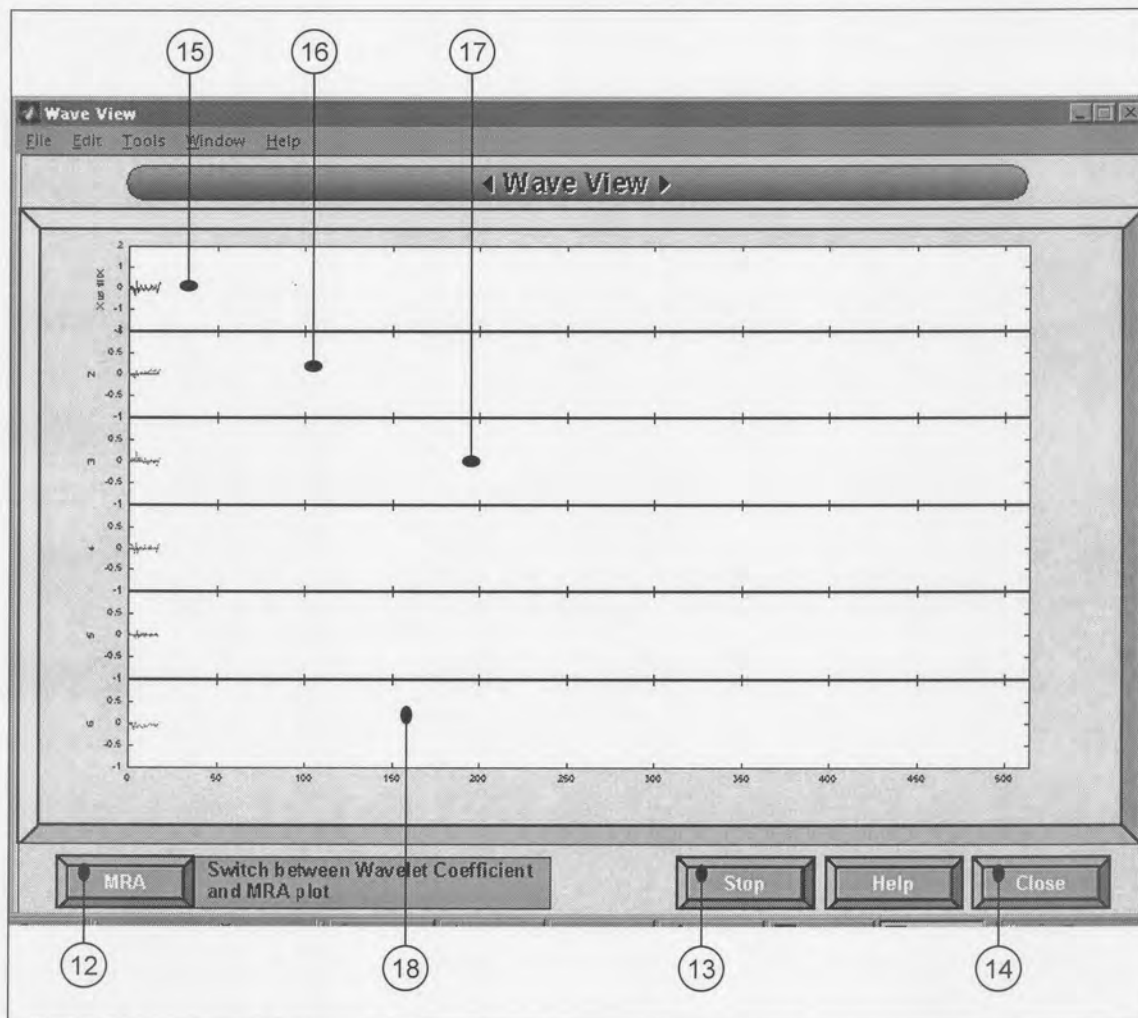


Figure 6.14. Wavelet analysis viewer – Reconstructed details and approximation

Figure 6.14 Tags:

12. Switch between multiresolution analysis (Figure 6.14) and wavelet coefficient (Figure 6.15) plot.
13. Stop multiresolution analysis.
14. Close the plotting window.
15. Original (black) and reconstructed approximation/filtered (red) data plot.
16. Reconstructed detail signal at finest scale/level before (cyan) and after (blue) thresholding
17. Reconstructed detail signal at second scale/level before (cyan) and after (blue) thresholding
18. Reconstructed detail signal at coarsest scale/level before (cyan) and after (blue) thresholding

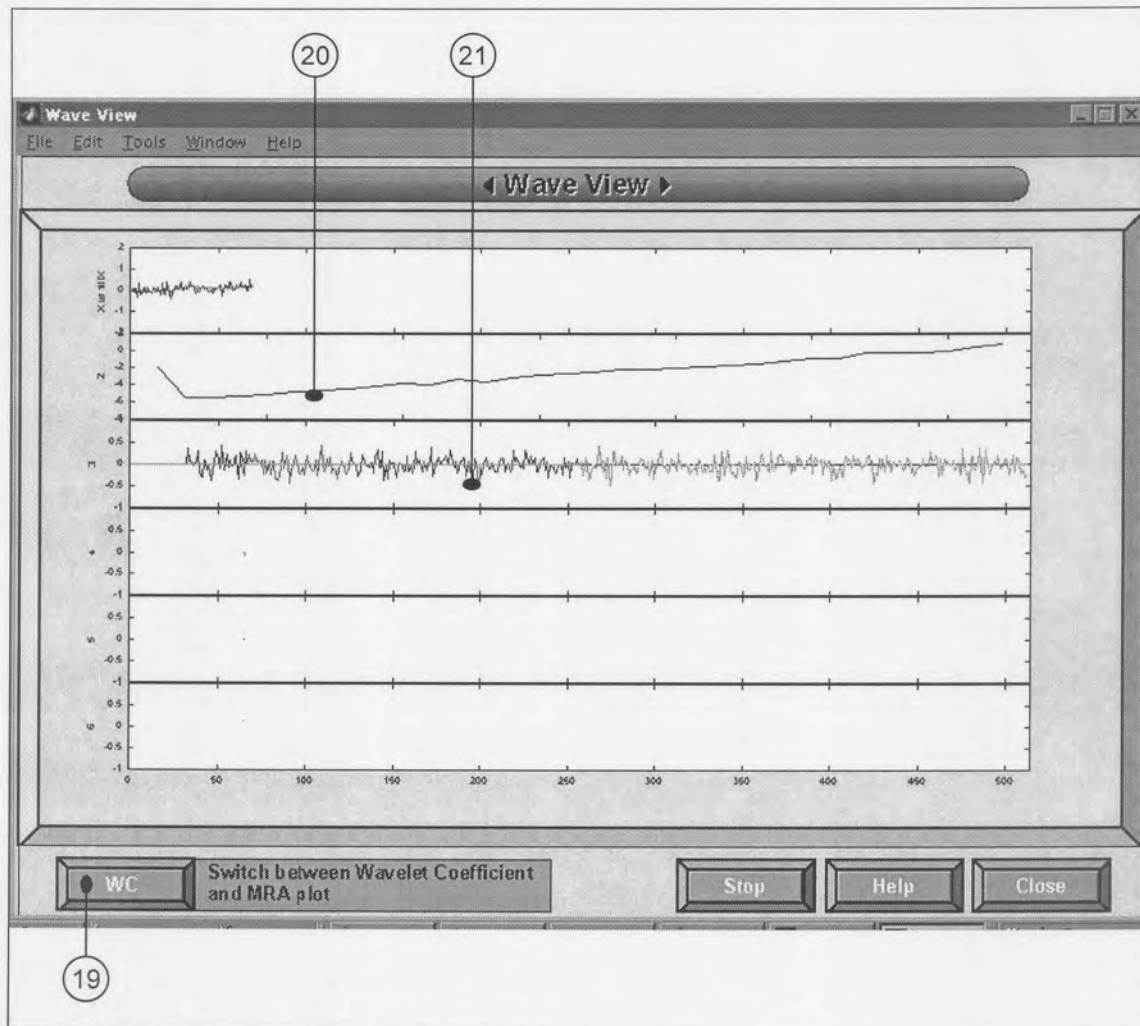


Figure 6.15. Wavelet analysis viewer – Wavelet coefficients

Figure 6.15 Tags:

- 19. Switch back to multiresolution analysis plot
- 20. Approximation coefficients of dyadic window after each time interval
- 21. Detail wavelet coefficient plot of dyadic window after each time interval.

Figure 6.16 is a completed version of Figure 6.14. This was the multiresolution wavelet analysis of variable one.

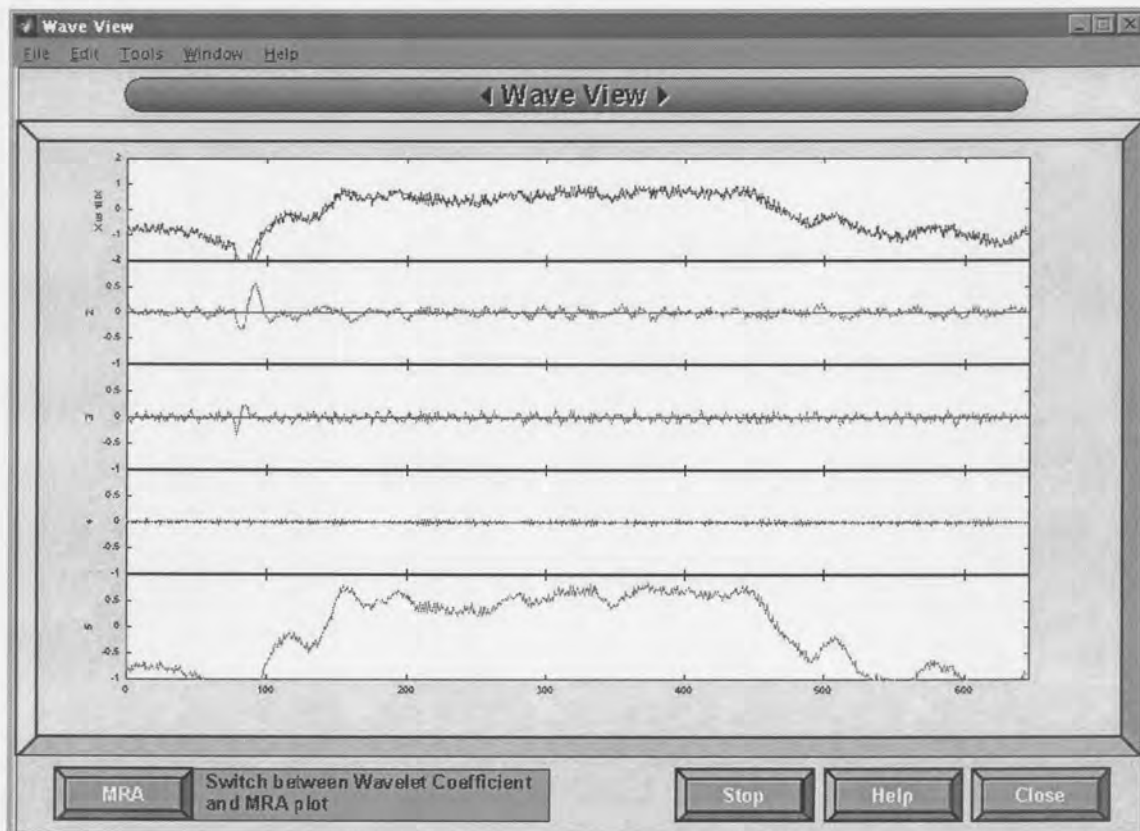


Figure 6.16. Completed multiresolution analysis of variable one.

Figure 6.17 Tags:

22. The name of the variable containing the thresholded multiresolution data for a specific level. Here *thmra\_level\_1* refers to the thresholded multiresolution data of level one which is a data matrix containing the first detail level of all the variables each in a separate column.
23. The nonthresholded (black) and thresholded (red) reconstructed detail level of the level specified by tag 22 and variable number specified by tag 24. Here the nonthresholded and thresholded detail of detail level one of variable one can be viewed.
24. Variable number of which the information is required.
25. Here the effect of removing or adding the specific nonthresholded detail level can be viewed. The black plot represents the original nonthresholded reconstructed signal and the red plot the thresholded signal with the added effect of removal or adding of a nonthresholded detail level. This is used if one wishes to override the thresholding of a specific level. During thresholding a specific detail level may be zeroed (removed) as in this example. The user may however decide that the specific level is significant and that it contains important information. In such a case the detail level may be replaced and the effect on the final signal can be viewed.

Thresholding may also retain some information in the detail levels that the user may decide is insignificant, in which case it can be removed.

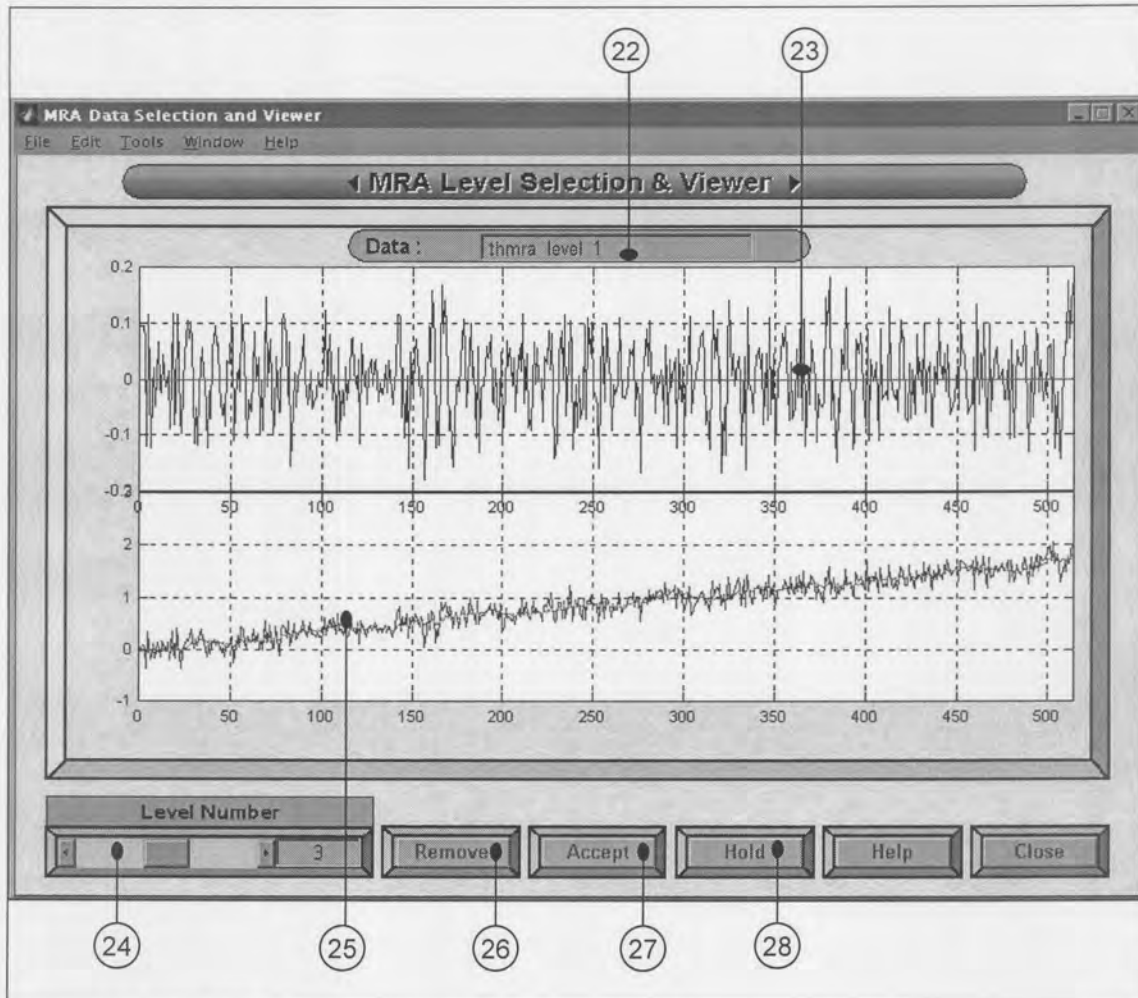


Figure 6.17. MRA data selection and viewer

26. Toggle between removing and adding a specific detail level.

27. Accept the configuration and save it to the database.

28. Use this if more than one detail level at a time needs to be viewed.

## 6.10.2. EXPERIMENTAL

### 6.10.2.1. Wavelet Analysis

The eight standardized variables from chapter 5 were decomposed into their contributions in different regions of the time-frequency space by projection on the corresponding wavelet basis function, as depicted in Figure 6.18 for variable one.

Figure 6.18 represents a moving window width of 256 data samples at a given time instance.

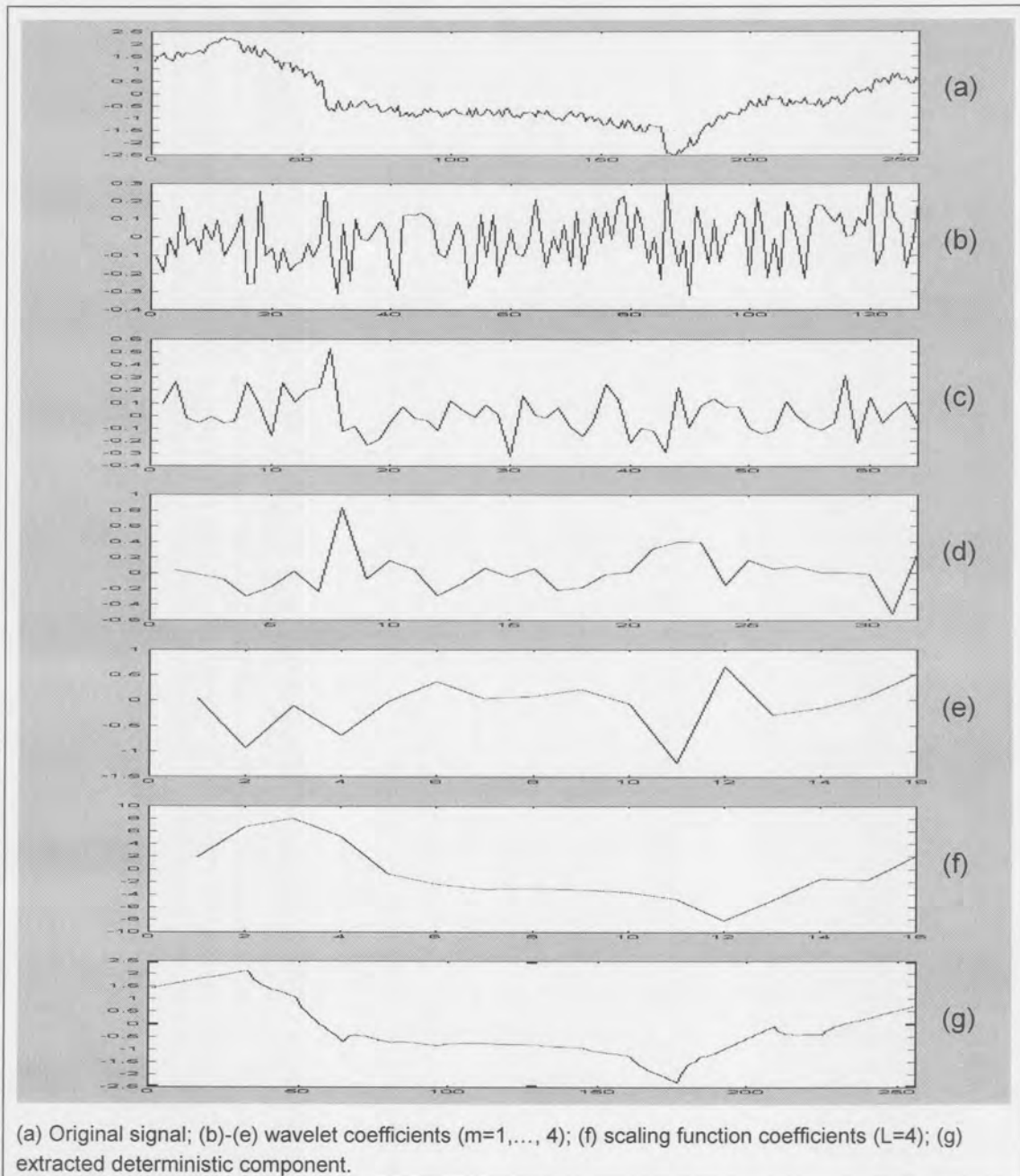


Figure 6.18. Wavelet decomposition and separation of stochastic and deterministic components.

The high-scale, low frequency content (approximation) of variable one is represented on a set of scaling functions, as depicted in Figure 6.18(f). The low-scale, high frequency content (detail) of variable one is illustrated by Figure 6.18(b)-(e).

The same process was repeated for each variable during each time interval.



## 6.10.2.2. Multiresolution decomposition

Multiresolution decomposition based on wavelets was carried out for each variable to observe both the general trend and the detailed features of the process data. The discrete fast wavelet transform using a boundary-corrected Daubechies second order filter at level  $L = 4$ , which is half the maximum length, was used.

Decomposition at level 4:  $s = A_4 + D_4 + D_3 + D_2 + D_1$

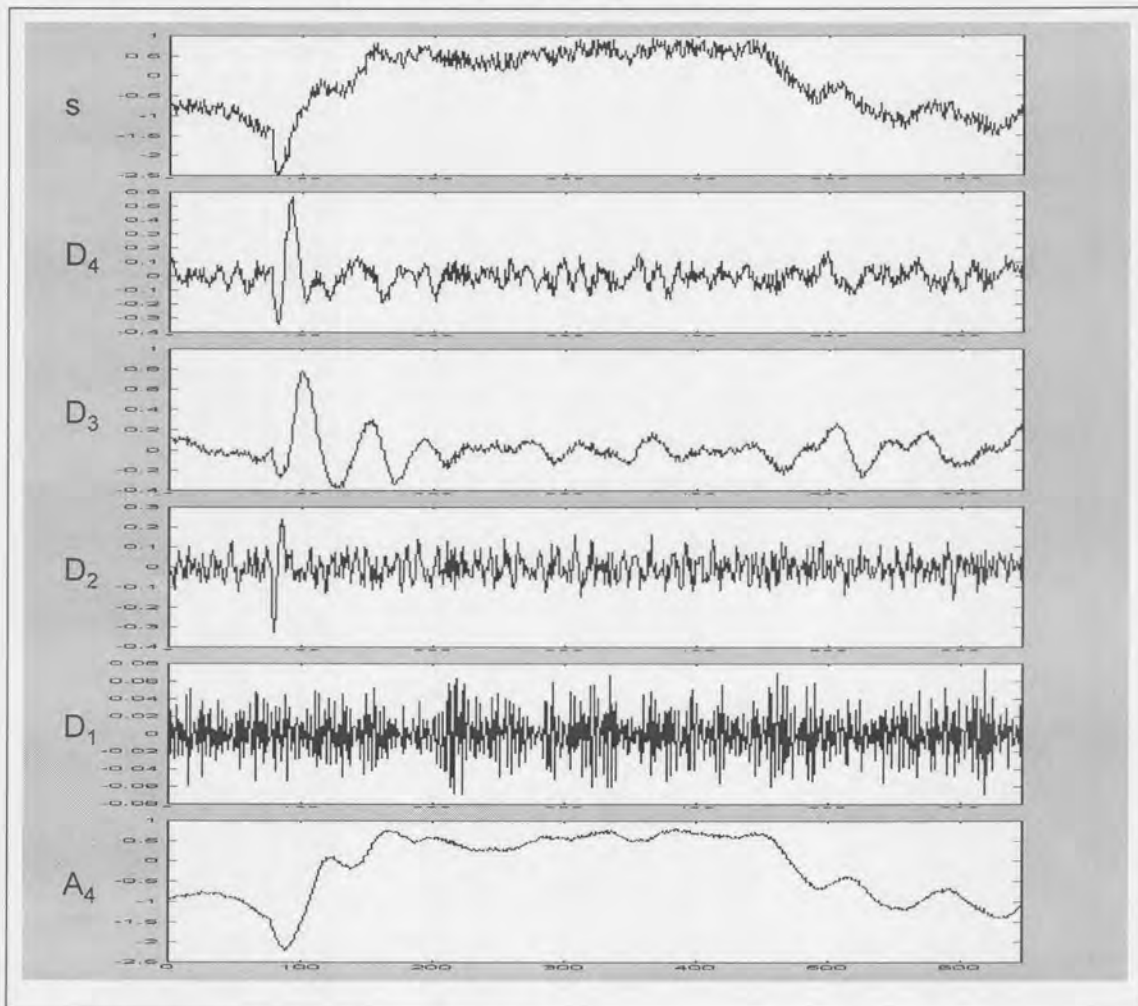


Figure 6.19. Multiresolution analysis plot

The length of the dyadic window was chosen as 256 which is a power of two ( $2^8$ ) since classically the discrete wavelet transform is defined for sequences with length of some power of two. Note that this initial window length is not a restriction. Theoretically, OLMS rectification can start with any dyadic set of measurements, starting at two. However, since the threshold is estimated from the data in the moving window, the threshold estimate improves as the moving window length increases. When the noise is assumed to be stationary, the threshold stops changing after a large set of measurements are collected, and, thus, the moving window length can be held constant. Figure 6.19 shows the multiresolution analysis plot for process variable one. Approximation coefficients at scale 4 ( $A_4$ ) represent the underlying trend of the signal(s)

whilst wavelet coefficients ( $D_4 \sim D_1$ ) show the high-frequency details. Examining the multiresolution analysis results for all eight process variables (see Paragraph 3.11 of Chapter 3), the level of noise corruption was found to be different for each variable necessitating level dependent thresholding.

This was repeated for each of the eight variables and provided similar results for each of the variables.

#### 6.10.2.3. Wavelet thresholding

Wavelet thresholding based on hard thresholding was then used to remove the high-frequency noise as well as the spikes known to be outliers. Level-dependent threshold values were derived from the VisuShrink threshold strategy. In this manner both noise and spikes were removed from the signal without affecting the underlying process trends. The thresholding zeroed all the detail coefficients indicating that all the detail could be attributed to noise. The approximation coefficients obtained in Figure 6.19 for variable one and all the other variables preserve the process trend well in a compact form since all high-frequency elements are omitted (Shimizu et al., 1997).

#### 6.9.2.4. Multilevel signal reconstruction

The thresholded and non-thresholded wavelet coefficients were used to construct thresholded and non-thresholded approximations and details. The non-thresholded details and approximations were combined to form dataset 2 from which the combined principal component model was derived. At each level the thresholded details were investigated to see if they contained any significant contributions. The significant contributions were combined according to the level from which the detail principal component model for each separate level was derived. The investigation revealed that the current thresholded details did not contain any significant contributions so that no need existed to derive detail principal component models. An approximation principal component model was derived from the combined approximations (dataset 1). By removing the undesirable high-frequency elements from the nominal data, the possibility of input-training network overfitting (Chapter 8) is greatly reduced. If by chance any desirable high-frequency elements were removed, it would be accounted for in the combined model.

Figure 6.20 gives dataset 1 which contains the approximations of all eight variables. Dataset 2 contains these approximations together with all the nonthresholded details of all eight variables. The methodology explained in this section is illustrated by Figure 6.21.

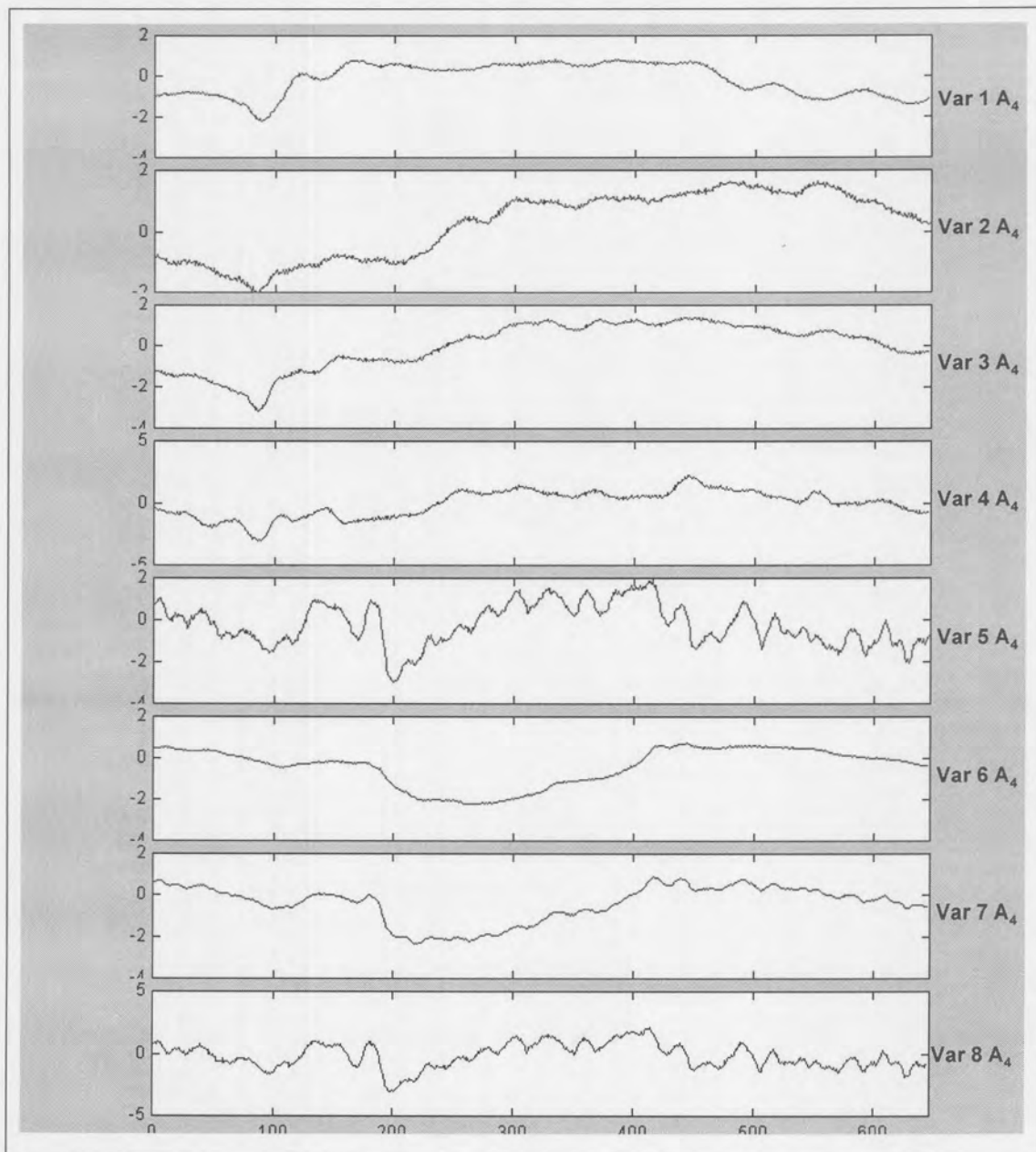


Figure 6.20. Dataset 1 containing the approximations of all eight variables.

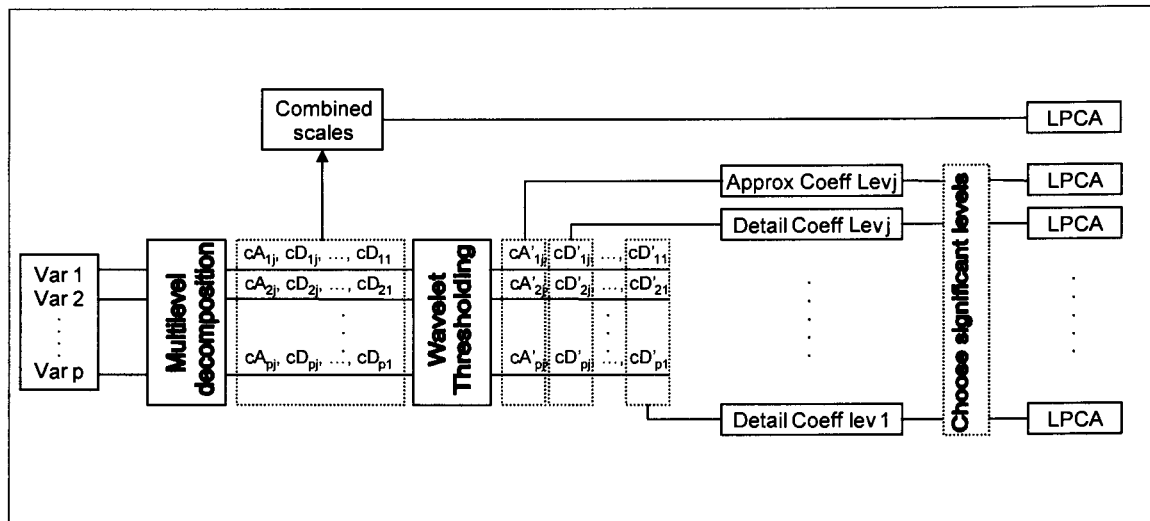


Figure 6.21. Comparison between the original signal, de-noised signal and the approximation coefficients used for model derivation