CHAPTER

8

Extended Process Control, Operator Situation Awareness, Alarm Management

Simplicity is the extreme degree of sophistication.
Leonardo da Vinci, 1452–1519

SUMMARY

At the heart of keeping a process within its safe operating window is the control room operator and the technology at his or her disposal. The latter technology went through an intense evolutionary and sometimes even revolutionary development. In the 1970s, George Stephanopoulos (MIT, Cambridge, Massachusetts) and others laid the descriptive mathematical foundation for process control. In his book, Stephanopoulos[1] lists the following objectives of process control. Foremost is maintaining a safe operation, followed by keeping product specification, staying within environmental regulation, taking account of operational constraints given by the physics and chemistry of the processes, and finally satisfying market conditions by optimizing process economics, hence overall performance. Given a set of optimal conditions, the influence of external disturbances shall be suppressed and the stability of the process ensured. Over the years, the economic consideration received more emphasis, and as we shall see, economic optimization within safety and environmental constraints is now the foremost objective of control. This is in line with the shift toward an overall system approach enabling a prediction of optimum output in a noisy environment making use of probabilistic tools.

Achieving process control objectives is easier said than done because of various kinds of mutual interactive effects of physical and chemical processes on each other's rates. Heat transfer rates and concentration equalization rates of reactive chemicals in solution, as well as the velocity of reactions and therefore rates of heat production or consumption, all interact. This introduces dynamics in various forms with different time constants of change, possibly resulting in oscillatory response behavior and growing instability. The basics of control, given a disturbance, exist therefore in the first place in measuring process output variables such as pressure, temperature, and flow. Next, control exists in analyzing the values of these variables and, based on the analysis, in deciding on actions to manipulate

Risk Analysis and Control for Industrial Processes - Gas, Oil and Chemicals. http://dx.doi.org/10.1016/B978-0-12-800057-1.00008-0

certain inputs to keep the process conditions at a set point level of steady state, or at least within desired limits. Hence, control is effectuated in a closed loop in which there is a *sensor* (measuring, e.g., temperature, pressure flow, or concentration), a (logic) *controller*, and an *actuator*, for example, a control valve.

The controller implements the control law. Today, these are programmable and are called programmable logic controllers. The two main approaches to serve the purpose are feedback and feed-forward control. The first just adjusts variables to their set point following observation of a change in output; the second is more of an anticipating nature, as it tries to measure a disturbance early on and taking into account time needed for changes (system time constants). It means that models are needed of the process, controlled by the loop, to make such prediction. This is also required in inferential control in case key variables cannot be measured directly. If in a system only one variable must be controlled, the control configuration is called single-input, single-output; if however, as is often the case in process industry, more than one output variable must be controlled simultaneously, the configuration is called multi-input, multi-output.

Response by a controller in feedback control to a disturbance, which can consist of a step function in change and felt only after a delay—the dead time—can be in several degrees of accuracy and speed. As control laws, one distinguishes only proportional response with gain proportional to the change (P-controller) with the risk of overshoot. Adding integration of the effect, smoothing it out, and bringing offset to zero improves this (PI-controller). Adding a derivative function proportional to the rate of change (PID-controller) is perfecting this further. If a process model can be described by a first-order differential equation, one calls the system first-order, while the more complex systems follow a second-order differential equation. The mathematics have been elaborated over the years, and sensor technology and instrumentation has become much more refined. The original analog controller hardware became completely digitized, and with all of that, the extent of automation rose tremendously.

Oscillatory behavior following a change in conditions or in input is analyzed by spectral methods, such as Fourier analysis determining eigenfrequency and power spectrum, and characterized by Bode plots of the ratio of output/input wave amplitude and also the wave phase lag, both as a function of frequency. Nyquist plots show the same in an alternative fashion. For linear, first-, and second-order control systems, analytical solutions of system descriptive equations are still possible and helpful in control design. A complication is that measurement of the variables is subject to uncertainty and noise. One therefore obtains a time series of measurement values shaped as wavy signals with superposed spikey "hair." Algorithms have been developed for processing and filtering the stochastic signals. One algorithm is the Kalman filter that assumes a linear dynamic system law and normally distributed (Gaussian) noise. The filter calculates in real time a best estimate of the measured variable using only covariances of current measurement and the previous calculated state. Covariances are derived by autoregression with the aid of least squares techniques. In case of high filter gain, more weight is placed on measurement of the current state with less weight on the previous state. Adoption of this in process industry began in the late 1980s.

The relatively reliable PID single-loop controllers became abundant in the ever more complex installations. But, because of the interactions among various plant sections and subsystems, overall control became a weak point. In case of a process upset at some location, interaction can lead to propagation of disturbances throughout the plant possibly presenting a hazard and certainly diminishing performance. Much effort has been spent to develop methods to mitigate this disadvantage with centralized model predictive control (MPC).

Development of MPC algorithms began in the late 1970s with application in refineries. Basically, MPC consists of a dynamic process model that is obtained by performing experiments with the plant measuring effects in outputs following changes in inputs. Initially, the models were linearized, but later nonlinear models were applied to achieve a better fit. Full digitization of signals and signal processing was introduced. Equations are solved by an iterative process making a prediction for a short-term time and a continually receding horizon when time progresses. Since the early 1990s, performance of an MPC algorithm was tested against the so-called Tennessee Eastman exothermic chemical model process.

Centralization inherits limitations. We shall make a jump in this chapter to the advanced distributive MPC with real-time optimization (RTO) that has become possible by the extensive use of computing in process control. Christofides et al.[2] recently reviewed and commented upon this approach and provided a future perspective. The newest network technology, wired or even wireless, will enable decentralized or distributed model predictive control (DMPC) for different plant sections, yet still together in optimizing the plant as a whole. Therefore, an operator is assigned the role of supervising the automated systems and managing abnormal situations.

Despite the sophistication of the control, there are problem areas potentially affecting safety. Besides failing or malfunctioning components, such as sticking pressure control valves, there are less explicit problems. In 2003, Venkatasubramanian et al.[3] made an extensive review of the problem field of operator process fault detection and diagnosis enabling intervention. If something is wrong, process pressure and or temperature quickly run up or start oscillating. Finding a fault and redressing a situation requires quick action, so that operators come easily under time pressure. This pressure will be exacerbated by alarm floods. Early warning and quickly understanding signals is important. Recent achievements in this respect will be briefly described.

A condensed description will be given of a most modern Supervisory Control and Data Acquisition (SCADA) architecture. Besides the above described control features and fault diagnosis, human factors also are taken into account. Apart from physiological ergonomic factors such as lighting, colors on screens, and information presentation on displays, operator situation awareness will be larger in case someone is actively involved in the control, but human performance will suffer if this workload becomes too high. Higher time pressure induces higher error rates.

Therefore, effective alarm management is an important topic. We shall review this type of problem and possible ways of improvement and optimization.

As already introduced in Chapter 5, at the end of Section 5.1, in 2011 Venkat Venkatasubramanian,[4] then still at Purdue University, wrote a vision article for the AIChE journal on systemic failures occurring and ending in disaster such as the Enron scandal, the financial crisis, but also the Deepwater Horizon demise and various process plant mishaps. These disasters can be ascribed to severe weaknesses in risk management, not able to produce convincing foresight of approaching collapse due to system complexity and lack of prognostic tools. He endorses Leveson's system approach and identifies the "need of real-time intelligent decision support systems that can effectively monitor various aspects of process operations, and detect, diagnose and advise operators and engineers about incipient abnormal events." He recommends pursuing multiperspective modeling, which means looking at a system from different perspectives: structure, behavior, and function. The combination with complexity science will then yield the support systems mentioned. In this chapter, we shall encounter a number of building blocks, which together with ones seen in earlier chapters, such as the blended Hazid method, will bring us closer to the objective he formulated.

8.1 PROBLEM ANALYSIS

As we have seen before in Chapter 3, from a safety point of view, distinction is made between a basic process control system (BPCS) and an instrumented protection safety system. The latter consists of safety instrumented systems letting off pressure or quenching the system and/or an emergency shut-down (ESD) system stopping flows and energy addition and bringing the installation to a full stop. The BPCS is physically separated from the safety systems and its components distributed over the plant (distributed control system or DCS). By the separation, the safety systems can function independently of a failing BPCS. Also, utilities for the safety system, such as instrumentation air and power, shall be independent of those for the BPCS. As a measure of last resort, ESD must be able to be activated manually.

Although the control architecture can provide a high level of reliability, there are still problems associated with the control as a whole. The first part of this has to do with unexpected and unforeseen events, which can be initiated externally disturbing the process or its instrumentation. Alternatively, the cause is internal by (partially) failing components, which are often not in the mainstream and have escaped attention for a while. A second part is due to system instabilities, which may be initiated by small disturbances or irregularities but grow and propagate through the plant and of which the cause is difficult to trace. Causes can be malfunctioning of components, such as valves or sensors, but they can also be changes in the chemical or physical processes. All this can be due to contaminated feedstock or deviating compositions, or even a plain error somewhere, for example, as a result of bad maintenance. Next, in Section 8.2 we shall explore in more depth the fundamentals and theory of model predictive process control, its strengths and weaknesses and future outlook. Section 8.3 will discuss present-day process control hardware and software.

A third problem category is associated with human factors. Operators usually work as a team, partly as control room or panel operators and partly as field or outside operators. Mostly on instigation of panel operators, the latter are closing or opening valves, checking particulars, and making close-in observations at equipment in the field in abnormal situations, or performing small, urgent repairs. Because of the progress in control technology in the newest systems, operators have only a supervisory role but may at the same time be responsible for more than one plant. After start-up, they have to come into action if some upset is developing or has already happened. In such case, recovery depends on their decisions, which will be prone to human error. In 1992, an abnormal situation management (ASM) team was formed by Honeywell with Amoco, Chevron, Exxon, and Shell oil companies as members. In 1994, the ASM team was converted to a consortium with a number of additional members. The ASM Consortium is still active and chaired by Honeywell. In Section 8.4, the human factor in control will be further explained and some of the achievements of the ASM Consortium will be summarized.

8.2 DEVELOPMENTS IN CONTROL THEORY

The drive to be best and most efficient fueled the wish to integrate processes with respect to energy (exergy concept), and exchange and subsequent conversion of materials of neighboring plants. In addition, product quality requirements and the sharpening of environmental and safety regulation placed tighter constraints. Changes in feedstock or desired product specification require increased flexibility. These developments demanded enhanced process control, which was already subject to difficulties. For example, exponential dependence of reaction rates on temperature and radiant heat transfer dependence on the fourth power of temperature causes nonlinearity of change. Hence, to correct an observed deviation from the right conditions requires a disproportional incremental alteration in input. Other limitations are imposed by the process model used for predictive control. Slight deviations of the fit from encountered process conditions under all circumstances will result in mismatches. Inherent delays in response, signal noise, and inaccuracy in quantitative responses of sensors, or design limitations of actuator reach and capacity (e.g., of control valves) add to the problem. Hence, besides nonlinearity, process control may also be plagued by time-varying uncertainty, all of which threaten stability.

The past 20 years have seen huge developments in mathematical signal processing and computing power to follow processes real time, also enabling automation. But there is still ample room for further improvements, as described by Christofides et al.[2] and by Christofides and El-Farra,[5] while there is not a single route to be explored. This all pivots around seeking stability via Lyapunov functions for the solution of differential equations describing dynamic, nonlinear systems. Around 1900 the Russian mathematician Lyapunov thought of modifications to a nonlinear system to obtain a linearly achievable stability near an equilibrium point. To understand broadly what it all means, and how it will lead to MPC serves the (over)simplified description below.

The continuous-time model describing the state-space of a *linear* process system (output roughly proportional to input) contains a state vector $\mathbf{x}(t)$ and its time derivative, an input vector $\mathbf{u}(t)$, an output vector $\mathbf{y}(t)$, all from sets of real numbers, and time variant matrices $A(t)$, $B(t)$, $C(t)$, and $D(t)$. Vector lengths are measured in space as Euclidian norm or metrics. Most generally, the state-space relations are formulated in the following so-called control dynamics and observer tracking equations:

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t)$$

$$\mathbf{y}(t) = C(t)\mathbf{x}(t) + D(t)\mathbf{u}(t)$$

The system is usually considered only with time-invariant matrices. For a general *nonlinear* system, the equations can be written as just a function of time t, state x, and input u: $\dot{\mathbf{x}}(t) = \mathbf{f}(t, x(t), u(t))$ and $\mathbf{y}(t) = \mathbf{h}(t, x(t), u(t))$. If the functions are a linear combination of states and inputs, the equations can again be described in matrix form as above and also usually with time-invariant matrices.

As the system here will be steered at least on the short term to a desired, fixed output, for the control function the first equation is relevant. The values of x are affected by disturbances and uncertainty, which also can be described explicitly as a term separate from u. Therefore, following Christofides et al.,[2] for the various subsystems of the process coupled through the states (and not through the inputs), the control dynamics equation can be aggregated to

$$\dot{x}(t) = f(x) + \sum_{i=1}^{m} g_i(x)u_i(t) + k(x)w(t)$$

where f is a function aggregated over the various subsystems $i = 1\ldots m$; g_i is a matrix relevant to subsystem i; w represents disturbances, assumed to be bounded; and k a matrix. It is assumed that these are locally all Lipschitz vector functions, which means that the ratio of the metrics or space distances between the vector elements of the functions of the state variable x, and the ones between the state variable elements, is equal or smaller than a constant, the Lipschitz constant. A further assumption is that the origin is a system equilibrium point unforced and undisturbed, so that $f(0) = 0$. By linearization and converting to the discrete time variant, approximated solutions of this equation for MPC have been proposed. However, Christofides et al.[2] propose a Lyapunov functions—based control solution, indicated as LMPC. This is by assuming a local Lipschitz control law making the origin of the closed-loop asymptotically stable and satisfying all state input constraints. Such control law covers the conditions of Lyapunov function based stability.

Overall control in a plant is hierarchical as it is realized in layers. The lowest layer is the PI and PID controllers for the local valves and other actuators. Above that is an MPC layer keeping main control variables at set-point values, while on top is an economic optimization layer for RTO. The latter is in high demand for coping with continual changes in market prices of energy and other resources. For the economics, optimization via MPC is required besides a model of the system, a performance index over a finite horizon, and a scheme that will have the horizon stepwise receding. The accuracy of the model to predict future trajectory without correction (open loop) determines the efficiency of the control, the performance index shall be minimized against the various constraints, and the receding scheme will provide the feedback to make up for model errors and disturbances. There is however the issue of stability of the scheme. Stability is achieved by combining here a nonlinear control law with the Lyapunov functions approach, although there are limitations to the length of the sampling time and the upper bound of the disturbances.

As mentioned, centralized MPC has the drawback of enlarged complexity, so that MPC decentralized to the subsystems, as shown in Figure 8.1, would enhance efficiency and flexibility. Depending on rates of change in the processes within a plant, separated MPCs can be on different timescales. If, however, the MPC units are not communicating, part of the advantage is lost. Making use of modern information technology (IT) networks (Ethernet, eventually at least partially wireless) enables communication between MPCs of different units, creating a distributed MPC or a DMPC. In each local MPC the cost function has to be optimized. This is realized by iteratively calculating the effect of its inputs on the plant as a whole assuming that the inputs of other MPCs are fixed at the optimum values of the previous iteration step. In a final step, a weighted sum of the individual MPC outcomes is determined and agreed upon.

Further, Christofides et al.[2] discuss asynchronous and delayed feedback effects in iterative DMPC. Future directions of research have the objective of covering other weaknesses such as optimal distribution of systems and optimal communication strategy between controllers. Another problem is coping with incompleteness of measurement data, while economic DMPC needs more attention. DMPC for controlling in part continuous time systems and for another part discrete event systems, the hybrids, is also a research topic deserving attention. Finally, DMPC provides the

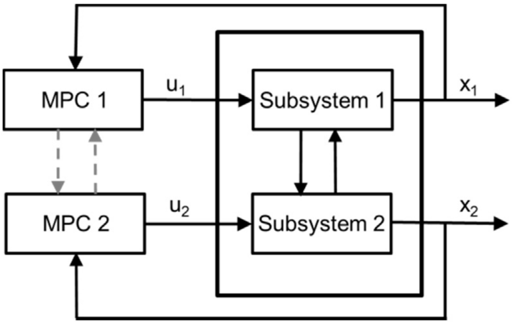

FIGURE 8.1

Decentralized or rather distributed but in parallel coordinating DMPC according to Christofides et al.[2] Only two units (LMPCs) are depicted.

possibility of monitoring and reconfiguring the system in case faults in the system are detected, so that the process can continue, while defect sensors or other components are being replaced. This kind of flexibility also offers the possibility to reduce data communication intensity in case of steady-state operation. For further reading on the background of MPC control see Christofides and El-Farra[5], and for integration of dynamic economic optimization (D-RTO) and MPC, see Ellis and Christofides.[6]

The latter authors describe a further development step. Where around 2000 RTOs were performed on an optimized steady-state model of the process, more recently a dynamic process model in the upper layer made it a dynamic RTO or D-RTO. Further, the simply weighted quadratic cost functions in the MPC were replaced by general economic ones, leading to feedback control economically optimized and integrated in one layer, the economic MPC or EMPC. Although this ambitious step proved possible, in the opinion of Ellis and Christofides[6] it cannot be absorbed easily due to necessary redesign of the control architecture and the requirement of fast response. Therefore, they propose a two-layered system of an optimizing EMPC on top, with Lyapunov stabilized MPC (LMPC) for controlling the process variables below.

8.3 FAULT DETECTION AND DIAGNOSIS AND FAULT-TOLERANT CONTROL

Fault detection and diagnosis in control and fault tolerance in control are important in general for running a process and maintaining the right product quality but also with respect to safety. The earlier detection and diagnosis, the earlier intervention will be possible. In 2003, Venkatasubramanian et al.[3] published an extensive review of methods of fault detection and diagnosis. As also becomes clear from the previous Section 8.2, computing led to much progress in control and enabled refined automation but added also to the complexity. In case of upset, which can arrive propagating from a remote part of the plant, the operator supervising the system must come in action to diagnose the fault and to take measures. Hence, abnormal event or ASM has not become an easy task. On the contrary, certainly younger operators with less experience in the mechanisms playing a role in the process may become overwhelmed by the number of causes that could be at the origin of the upset. Their mental image of the process may be incomplete or even wrong, leading to wrong decision making. Venkatasubramanian et al.[3] reviewed all methods that can help in such situations.

First, the authors specified the terminology. Fault is a deviation from the normal value of an observed variable or calculated parameter beyond the acceptable limit; underlying it is a (hidden) root cause or basic event, for example, a malfunctioning or failed component, but it can also be incompleteness in the model. A reliable diagnostic classifier with sufficient resolution but also speed would therefore be desirable. Such a tool should also be robust, not sensitive to noise, while known abnormalities

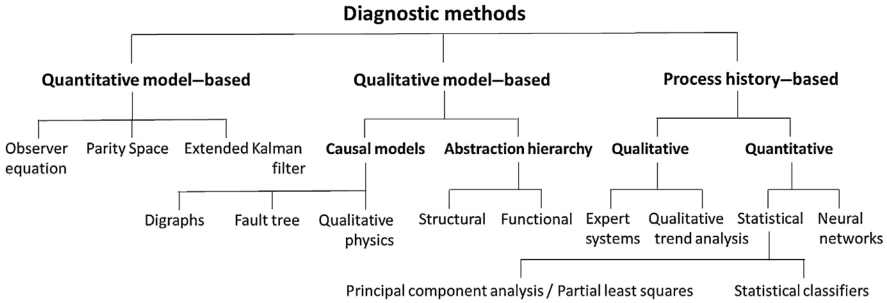

FIGURE 8.2

Diagnostic methods for fault classification.

Adapted from Venkatasubramanian et al.[3]

and faults shall be distinguished from unknown ones. Adaptability to changes in the process should be possible and an online decision support system should be able to explain the cause and the propagation with minimal modeling and minimal computational effort. Finally, the system should be able to identify multiple faults, which coincidentally appear, although this will be a difficult requirement.

Obviously, inputs to a diagnostic system come from measurements/observations. In a next stage, from the measurements features are extracted that can help in the diagnosis. Features are then fit into a decision space usually provided with an objective function to minimize error. This is performed by applying a statistical discriminant function analysis or by using simple threshold values. A discriminant function analysis differs from variance analysis in that it is tested whether a set of parameters determining a class fits the observation. Finally, the class of fault is derived. An overview of diagnostic methods is given in Figure 8.2.

8.3.1 QUANTITATIVE METHODS

Most of the quantitative model—based fault detection and isolation (FDI) occurs by the type of observer equations we have seen in the previous section. The methods are two-step ones: the first step generates residuals, that is, squared differences between expected and observed values, yielding possible inconsistency; the second step serves to decide what is at fault. In principle this can be achieved by hardware redundancy and comparison. However, because of cost reasons it is mostly realized by analytical, also called functional, inherent, or artificial redundancy. This is accomplished by comparing expected values of model-related process variables with observed ones. Direct redundancy is finding a faulty sensor by comparing sensor signals related via the model. Temporal redundancy analysis means relating outputs of different sensors and inputs of actuators to find out which component is failing. Venkatasubramanian et al.[3] provide equations for discrete and continuous, linear and nonlinear systems. As an example of the latter, Kazantzis et al.[7] developed a nonlinear observer typically suited for a chemical batch reactor or a continuously stirred tank reactor with multiple stable steady states.

Parity space relations consist of checks of the output of sensors against known process inputs. Residuals are never fully zero because of various disturbances: noise, model errors, bias, etc. Because disturbances have mostly a dominant random component, a third possibility is to apply a Kalman filter. Such a filter is an optimum predictor of the state parameters for a linear stochastic system in a noisy environment. (Note: today a Kalman filter can be substituted by a Dynamic Bayesian Network which will not limit application to linear Gaussian systems.)

In case of redundant hardware with voting as in high integrity systems, differences in signals may lead to fault isolation. Also, enhanced residual methods have been developed such as directional and structural residuals. Here, the essence is that a set of residuals is determined and that not only the existence of a fault is shown but also what fault, making the approach not only fault sensitive but also fault selective. “Directional” indicates the direction in which the fault should be found and “structural” links to a subclass of faults.

Although in principle there are possibilities by making use of the MPC gear already present, Venkatasubramanian et al.[3] report in their review that in practice several problems are encountered. Complexity of the system with associated nonlinearity, the high dimensionality of the vectors involved, and lack of good data complicate the development of a sufficiently accurate model for relying on residuals. There have been trials applying statistical treatment on the correlated diagnostic residuals by determining, for example, distributions, a moving average, and performing statistical tests. But the control on the basis of residuals, in particular in case of chemical processes which are by nature nonlinear, proved in general to be problematic, although some successful examples have been described.

A decade after the review mentioned above, Mhaskar et al.[8] assuming time-invariant, nonlinear systems proposed a fault-tolerant MPC stabilized by Lyapunov functions (LMPC), capable of FDI. After explaining in more detail the mathematics behind LMPC, first a single input nonlinear algorithm was derived capable of fault detection and fault tolerant control of input constraints due to failing actuators. Subsequently, a case was worked out of multi-input, multi-output subject to multi-failing actuators. The scheme introduces a fault detection filter with an identical state equation as the closed loop that is tuned first to a fault-free closed-loop state measuring a minimal residual. The filter runs parallel to the process and actual states are compared with the ideal ones. If the difference in residuals surpasses a certain threshold for a certain input variable, there is a fault in the actuator controlling this input. However, also below the threshold a fault can be present. By trimming the threshold as far as possible, the probability of missing a fault will be as low as possible. The magnitude of the difference of residual on- versus off-line is representative of the position in which the actuator failed.

Fault tolerance can be obtained by making use of a switching policy governed by a switching rule. Switching is executed by the plant supervisor, who can determine whether a “fall-back” control configuration can be established, which still can provide a stable closed-loop control around the nominal equilibrium point desired. However, in the case where this is no longer possible, it is proposed to obtain a "safe-park" point. This can be realized by establishing for the system a number of safe-park points off-line for the various positions of the potentially failing actuator, so that in the case it fails, the actuator can be repaired/replaced while the process need not be shut down. The fault information about which actuator is failing, and to what extent, is thus crucial for the selection of the safe-park point.

An even more challenging problem for nonlinear systems is locating a faulty sensor. To that end, a high-gain controller design is proposed and again residuals determined. Now, the difference between the state estimate of the high-gain observer and a sufficiently accurate predicted value is monitored. The information on which the latter is derived is again established initially when all sensors function. The residual will reveal the sensor configuration of which the faulty sensor is part. After isolation, redundancy will enable to switch off the faulty sensor. A second approach in case of sensor data losses or asynchronous measurements with loss of feedback is to have actuators that can store the last optimal input trajectory instead of going to zero or the last given value. For this, the Lyapunov-based control has been modified to deliver that trajectory information. The use of the methods was illustrated in example processes.

Because of the complexity of the method—the mathematics involved have not been detailed here—and the fact that one controls the process with the same system as one diagnoses, a second, independent method of diagnosis would be desirable, also to reduce the chance of false alarms. Failures can be other than failures of sensors and actuators, and it will not be simple to distinguish signals caused by true, maybe unknown instabilities from those caused by failure. Below are some candidate methods.

8.3.2 QUALITATIVE METHODS

The two structured methods for diagnosis, summarized by Venkatasubramanian et al.[3] in part 2 of their review, are signed digraph (SDG) and fault tree (FT) to link cause–effect chains. We could add here Bayesian network (BN), which became available more recently, adding also a quantitative element. We have described BNs in Chapter 7. In addition, there is the common sense reasoning of qualitative physics.

An SDG is a directed graph, which is a qualitative "cousin" of the BN. It consists of nodes connected by arcs directed toward the effect. The nodes represent variables or events; in an SDG the arcs have positive or negative signs attached to them depending on the direction of change they indicate. SDGs are efficient in showing graphically a cause–effect model, hence a causal graph. In contrast to Bayesian nets, SDGs can be in part cyclic representing self-amplifying or attenuating processes. In this respect, SDGs resemble system dynamic graphics. Venkatasubramanian et al.[3] present a simple example of a digraph of a tank being filled by flow F_1, while at the same time at the bottom flow F_2 exits. Rate of F_2 depends on the height Z of the liquid level in the tank. Hence, $F_1 - F_2 = \frac{dZ}{dt}$ and $F_2 = Z/R$, where R is flow resistance for F_2. The corresponding digraph is shown in Figure 8.3.

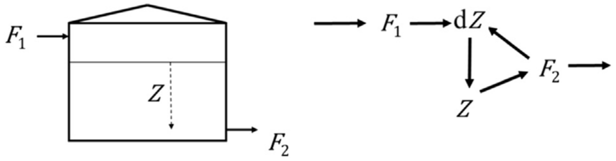

FIGURE 8.3

A simple digraph of a tank being filled, while at the bottom a flow exits.

A process can be modeled in an SDG with all material and information flows initiating or undergoing changes. Application in process control fault diagnostics started in the late 1970s and some quantification describing dynamics by means of ordinary and partial differential equations (as in system dynamics models) entered in the late 1980s. Combination with fuzzy logic and fuzzy reasoning (briefly described in Section 7.6.2) followed in the late 1990s.

FT analysis has been briefly described in Chapter 3 (Section 3.5.2.2). Given a failure event (top event), one asks continuously going further down in detail what underlying possible causes can yield the event, until one reaches basic failure events. So, in reverse one can see how a fault propagates to the top-event. The logic structure has an advantage over SDG as it can explicitly show that two or more underlying causes only together can cause a higher failure (AND-gate), or whether one or the other (OR) suffices, or one or the other but not simultaneously (exclusive or: XOR), or one but not in combination with the other (NAND). But like BN, also FT cannot represent self-amplification or attenuation as SDG can. FT is used quantitatively by attributing failure probability values to the nodes. Modeling temporal effects and partial failure in FT is, however, tiresome. On the contrary, SDG represents conditional processes easily. As we have seen in Chapter 7, Bayesian nets are more flexible than FT in, for example, modeling confounding causes and is able to handle probability distributions rather than only discrete values, and even to a certain degree able to model dynamics. But the rigidity of the FT structure often helps to obtain a quick overview.

SDG, FT, and BN are just methods to structure cause—consequence chains of which causes have been identified by different methods, while FT and BN can calculate top-event probability given quantified underlying causes. Methods to *identify* what can go wrong have also been treated in Chapter 3. The main ones are hazard and operability analysis (HAZOP) and failure mode and effect analysis (FMEA). HAZOP will be supported by common sense qualitative physics (CSQP) as mentioned by Venkatasubramanian et al.[3]

"Over-the-thumb" CSQP can be structured in two ways. The first is to derive qualitative, so-called "confluence equations" in terms of increase or decrease, and high or

[3] Note: As signed digraph in case of increasing flow F_1 all arrows will have a + sign attached, only the one between F_2 and dZ a - sign.

low, or order of magnitude change. The second is called “precedence ordering,” which is ordering of variables in the direction of the information flow through them. CSQP can be elaborated into qualitative simulation or QSIM of the physics of a process.

The last qualitative methods according to the scheme in Figure 8.2 are the *structural* and *functional* abstraction hierarchy of process knowledge. These consider the hierarchical decomposition of the system, its subsystems and further, and how effects can be explained by their interaction. Structural decomposition looks top-down and represents connectivity (as we have seen in the system approach of Rasmussen and Leveson in Chapter 5), and functional decomposition concerns the means-to-ends relations bottom-up (as does HAZOP). In structural approach, pleas have been made to consider control loops for each subsystem (as Leveson with STPA advocates), while for functional analysis consideration of mass and energy flows at various functional levels has been suggested (multilevel flow models). This again is covered by HAZOP and even better by the recently developed Blended Hazid method, explained in Chapter 7 (Section 7.3), which is based on a functional systems framework.

Venkatasubramanian et al.[3] conclude discussion of qualitative methods by considering search strategies and mention the two types: *topographical* and *symptomatic*. The first is either by structural or functional search. Structural is by identifying the path information flows through a unit. In case of a fault found on this path, the fault is further localized in successive refinements. Functional search is by finding differences/mismatches between normal and actual functioning of all components connected. Once a subsystem is identified as suspect in a functional sense, the functional search is continued through further decomposition. In practice, structural and functional search are often combined. Symptomatic search consists of systematically applying input variations and observing outputs, while comparing with how these should be in the normal situation.

As we have seen here and before in Chapter 3, HAZOP and FMEA are main contributors for identifying faults and failures and their effects. The problem is that in general a process installation is large and complex and contains many components allowing many different ways in which control can be lost. Structuring identified mishap mechanisms and computerization is desirable for reliable storage and quick retrieval needed for diagnosis. Obviously, FMEA is much associated with FT for which software packages are commercially available. Digraph models have been applied by Vaidhyanathan and Venkatasubramanian[9] in an effort to automate HAZOP analysis (Chapter 7, Section 7.3.3).

However, in this connection the applicability of the new Blended Hazid method by Cameron (Chapter 7, Section 7.3) should be stressed. Computer storage of the results of the combined HAZOP and FMEA in Blended Hazid enables quick retrieval. The option of generating causal graphs by just entering in a computer running the software for an observed deviation will make this technique even more successful for fault diagnosis (in the sense of a functional search). As Bayesian network (BN) finds applications for medical diagnosis, also quantified Blended Hazid could fulfill that role by producing a probability ranking of potential causes. For example, in a BN portraying various modes of a disease of an organ with its causes and indications, based on historical data probabilities can be specified of possible different disease causes, conditional on patient history, gender, age, and many other preconditions. Over the years all these data may be known as means or distributions for a population as a whole. The quantified network then relates the hidden causes to indicators (e.g., blood constituents, temperature, and appearance). By performing an observation on an indicator for a particular patient, the most probable cause in that instance can be inferred. Similarly, this kind of symptomatic hidden cause search approach based on (deviating) observables could work for a process plant. In Section 8.5.5 we will return to this possibility.

8.3.3 PROCESS HISTORY—BASED METHODS

The third part of the fault diagnosis methods reviewed by Venkatasubramanian et al.[3] as shown in Figure 8.2 consists of five methods: The first two are qualitative: expert systems and qualitative trend analysis (QTA); the other three are quantitative: principal component analysis/partial least squares (PCA/PLS), statistical classifiers, and neural networks.

Expert systems have been around and are useful in a rather specialized field. The technique is rule based: if…then…else. Venkatasubramanian and coworkers have referenced a number of papers on applications of this method of feature extraction in the process industry.

Qualitative trend analysis can be considered as part of process monitoring and supervisory control; recognizing a trend will enable prediction. As process variable signals are blurred by noise, QTA uses filters, for example, autoregressive ones, to smooth the signal to detect a trend. Filtering introduces the risk that a true significant peak may not be observed, although patterns may be recognized. By using filters on different time-frequency domains, some relief of this drawback can be obtained. Most references concerning this topic stem from the 1980s to the early 1990s. In a follow-on article to the reviews, Dash, Rengaswamy, and Venkatasubramanian[10] discussed the application of fuzzy logic for a two-stage trend analysis. The first stage is only analysis in a qualitative sense but the second applies fuzzy set for a rough quantitative estimate that enabled identification of a faulty sensor, given knowledge of fault signatures.

In the quantitative, process history—based methods as PCA/PLS, the Gaussian time series of process signals (oscillatory with superposed white noise) are sampled online and analyzed much more rigorously. Multivariate analytical techniques (Pearson covariance matrix methods and correlation coefficient calculation) are applied to extract main features and separate correlated variables from uncorrelated ones. This technique differs principally from the quantitative fault-finding methods applying a process model (MPC). Numerous references on applications of PCA/PLS quoted are typically from the 1990s. Faults of, for example, sensors and valves should be detected as early as possible. Therefore, online sampling and statistical signal processing must be accompanied by threshold definitions for alarming. A design challenge is the development of a stopping rule to avoid false alarms as much as possible.

More recently, further analysis of wavy and noisy process variable signals has been undertaken. In these cases patterns resulting from nonlinearity can be recognized. As in many plants an operator has tens and even hundreds of control valves "under his wings," it is not only of interest to know whether one of those is malfunctioning but also which one. Shoukat Choudhury et al.[11] performed a thorough analysis in relation to valve stiction (sticking-friction). Nonlinear behavior can be induced by backlash, hysteresis, dead-band, and dead-zone of control valves caused, for example, by wear, corrosion, or design faults. Oscillations may show plant-wide propagation. The methods have been developed and tested in detection and diagnosis of these oscillations.

The signals may contain Gaussian noise but other distributions are possible. The signals may also possess an autocovariance property, which means a time-delayed autocorrelation of a stochastic system. (Autocovariance is not to be confused with autocorrelation. If the autocovariance is normalized by the product of the standard deviations at the two points of time—time and time delayed—the autocorrelation coefficient is obtained). Linear autocovariance is second-order moment related, but for nonlinear processes a third moment is needed for characterization, deriving the cumulant.

Analysis requires higher-order statistics. Fourier transform analysis is applied to reveal the dominant frequencies of a signal characterized by its power spectrum. This spectrum is the Fourier transform of the autocovariance function. The bispectrum is the frequency domain representation of the cumulant and derived as a double discrete Fourier transform. Bicoherence is the normalized bispectrum and can be plotted as an ordinate in a 3-D diagram versus two horizontal frequency domain axes, revealing nonlinearity (see Figure 8.4). Wiener filtering is applied to signal parts relevant in contributing to the nonlinearity. In addition, stiction is quantified by applying the fuzzy C-means clustering technique[13] on scattered groups of data points to derive an elliptic plot of just the controlled process variable output (pv) versus controller output (op). These plots have been made for various types of valve faults producing typical fault signatures, so that the method can be interpreted as a statistical classifier.

Although these methods are quite intricate, some striking results are shown. Valve stiction can easily be ascertained by singling out the valve and applying an invasive test. However, for an operating plant such a test is expensive. The afore-described fuzzy C-means clustering technique method was however successful in detecting stiction of a particular valve in a running plant and quantifying its extent by the developed algorithm independent of type of valve or loop, only requiring a set point, controlled process variable output, and controller output. The method was patented and is applied in industry. The methods described by Shoukat Choudhury et al.[11] were also effective in diagnosing plant-wide oscillations.

The last history-based method of the review of Venkatasubramanian et al.[3] for fault classification by feature extraction is neural network. Considerable work in this direction was done in the late 1980s and first half of the 1990s. One of the results is the fuzzy clustering algorithm mentioned above.

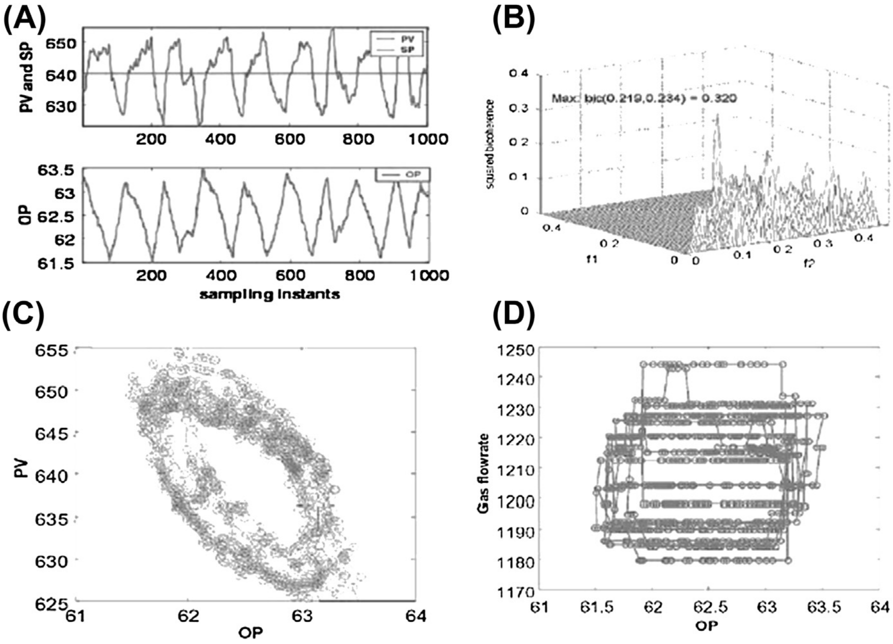

FIGURE 8.4

(A) Time series signal samples of the temperature control loop of a natural gas—fed combustion dryer system, (B) the bicoherence plot, (C) the derived *pv-op* (controlled process variable output-controller output) ellipse, and (D) the manipulated gas flow rate variable versus controller output after the 2004 work of Shoukat Choudhury et al.,[12] showing valve backlash and stiction.

8.3.4 COMPARISON OF THE VARIOUS METHODS

In comparing, we shall only highlight some salient methods. The recent quantitative method applying LMPC looks feasible but is too young to assess. According to Venkatasubramanian et al.,[3] industrial applications are mostly history-based statistical approaches with some emphasis on detecting faulty sensors. Fast online detection of abnormal situations with relatively easy-to-build statistical means is in demand. As we have seen in the previous section, recently, detection of valve stiction, applying higher-order statistics, has also become quite feasible. More refinement and depth of processing, however, requires more computational capacity. Older methods such as neural network and qualitative trend analysis have also been in use.

For the future, the instantaneous generation of causal graphs also seems promising to diagnose various kinds of defect and malfunction problems that can cause process upsets, as will be confirmed in Section 8.5.5.

8.4 TRENDS IN SCADA SYSTEM INFRASTRUCTURE

This section will be about SCADA, which integrates in its architecture the all-modular control equipment of a process distributed over the plant area (DCS). SCADA further centralizes all plant data and presents the information to the operator in the control room, at the so-called human–machine interface (HMI). Signal and logging data are given time stamps and stored for later use. Over the last decade security concerns with respect to the systems have been rising. This security is not only with respect to physical security of premises but in particular to cyber security.

The newest equipment, such as the Honeywell Experion® with universal channel technology, is enabling analog or digital, input or output on any channel and software configuration. As before, basic process control and emergency safety shut-down channels are fully physically separated and visibly distinguishable with blue and orange stickers on channels and a solver. Besides the in- and outputs, the safety modules also contain the safety logic solver, and they can be placed at the site of the equipment to be safeguarded. All devices are SIL3 TÜV certified, have a temperature range of −40 °C to +70 °C, and are hazardous area class Zone 2 (or Division 2) safe.

Field connections are by fiberoptic cable able to transport many different signals simultaneously, which simplifies the cable network considerably. The rest is within a fault-tolerant Ethernet. Wireless connection is in principle also possible (wide area network, WAN, and large WAN). Wireless device managers (WDMs) are for monitoring and are connected to servers. These WDMs are for actual functioning in a control loop not yet (summer 2013) sufficiently robust due to the battery capacity. Configuring field devices is realized securely with the control builder, while DCS system settings are protected by different types of interlocks, for example, for pre-shut-down conditioning related to safety systems. Control of the safety systems is via the safety manager.

Also, the instrumentation has been improved in stability, accuracy, and response time. Level measurement and chemical seal detection are more reliable; detection of pipeline vibration and plugging, and leak detection, will be possible in the near future. Signal communication after digitization in remote terminal units is to the field device manager (FDM), which supports up to HART 7 (Highway Addressable Remote Transducer) communication protocol for Foundation Field Bus and Process Field Bus (Profibus) data communication standards. The FDM enables the common tasks, such as loop tests, range updates, and following calibration procedures, but also less common ones, such as control valve stroke tests, drift analysis, and flow diagnostics. The control performance monitor (CPM) enables checking the performance of valves (stiction), sensors (drift, bias, failures), and controls (interactions, disturbances, service factors, and saturation/tuning issues). CPM produces management production key performance indicator reports and feeds the history database.

By the much enlarged computing capacity, the new system offers possibilities for virtualization up to DCS level enabling, for example, off-process developments and pre-acceptance, installing a (remote) backup control center, design and training, and many other activities. The system will help to make better use of the equipment and will save on the number of PC computers because on one computer other, virtual ones can be simulated. Large-screen technology (collaboration station) enables easy communication independent of distance. Another feature is a management of change application designed for preparing and keeping track of changes to the automation.

Much attention is given to cyber security measures, which continually will be updated. For a number of reasons, SCADA systems were believed to be relatively secure because of their specificity, but with the huge spread of IT knowledge this is not true anymore as the 2010 Stuxnet virus attacks have shown. Cyber security will be of increasing importance because of the evolution toward Internet use (TCP/IP-based equipment), cloud computing, and satellite communication. Drivers are cost savings and the enabling of remote unmanned operations. SCADA systems are crucial in keeping going processes that are the “blood, oxygen, and nutrients” for the economy, and for society cyber security is the contrary of a luxury. The International Society of Automation (ISA) is working on standards.

HMI including alarm management of SCADA will be discussed in the next section. The ASM Consortium, as mentioned earlier led by Honeywell, published several ASM Guidelines[14] on how to optimize HMI.

8.5 HUMAN FACTORS IN CONTROL, CONTROL ROOM DESIGN, ALARM MANAGEMENT

There is great pressure on companies to do more with fewer people. However, there is a limit to what people, given the best of intentions and motivation, can do. On the other hand, that limit shifts continually to higher level of capability as a result of improved and innovative technology. But even so, there are limits. This is also the case with the console operator taking process observations from his displays and making decisions about whether and how to make interventions. A crucial aspect in an operator’s vigilance is their situation awareness, which we shall also discuss below. In acting, the operator must follow procedures, in some of which an error can evoke a large risk. Not only can it be tough to follow a procedure error free but writing a procedure that is interpreted unambiguously is not an easy task either.

8.5.1 PROCEDURES

The issue of automation of procedures was brought up at the 2013 MKOPSC Symposium by Thomas Williams,[15] chairman of the ASM consortium. For operating a plant there may be hundreds, even thousands of different procedures. Certainly when it concerns executing procedures in abnormal situations, apart from increasing the operator’s work load, errors will hardly be unavoidable. Therefore, there is a trend to partly or fully automate procedures or at least to guide the operator executing those, although the investment in time to develop such automated procedures is not insignificant. For batch processes this has already being done for a while, and for which the ANSI/ISA-88 standard is available; for continuous processes, the ISA established committee ISA-106 to develop a standard. Benefits are particularly high when the procedure has to be used in an abnormal situation, when there may be time pressure and the operator does not have routine practice.

8.5.2 SITUATION AWARENESS AND EXTENT OF AUTOMATION

Mica Endsley[16] published a series of papers on the topic of operator situation awareness[a] culminating in two well-known papers in the mid-1990s. The key points from these have been summarized recently by Chris Wickens.[17] The construct of *situation awareness* (SA) can be defined pithily as "knowing, what's going on" and, more formally, as "the *perception* of the elements in the environment within a volume of time and space, the *comprehension* of their meaning and the *projection* of their status in the near future." In the latter, one can recognize the three main elements, also called levels of SA: perception, comprehension, and projection, distinction of which has consequences for system design and training. To a certain extent it resembles Rasmussen's operator knowledge-based behavior path shown in Figure 6.4.

SA is not the same as performance. At a high level of operation, performance may be good but the operator's SA may be low. Secondly, the time constant of SA is in seconds to at maximum some hours, while a *mental model* may be built in hours to years. Thirdly, the product of SA is not the same as the process of updating SA, but the distinction is fuzzy, certainly at the first level. Endsley enlightened SA by describing the effect of factors such as time, space, and team SA and also the links to decision making and performance. She further characterized the relation with other psychological concepts, for example, pre-attentive processing, attention, perception, working memory capacity, long-term memory, development of schemata and mental models, confidence level, automaticity, goals, plans, and scripts (outcomes).

Important to designers is the level of automation. In principle one can automate much, but it is known that the presence of a human supervisor is crucial in case of abnormal, not foreseen, or rare situations. If that occurs, he/she shall be alert and anticipate as much as possible. Decision time can be critical. However, with higher levels of automation, alertness decreases. So, to find an optimum is a dilemma as shown graphically in Figure 8.5. Quantitative information would be useful. Measurement attempts by various researchers have not produced clear answers. There are indications supporting case (b) but in some instances (c).

Another side of the same question of how far the ever more capable automation shall go has been analyzed by Flemisch et al.[18] This was accomplished in the light

[a] One finds perhaps even more often the expression *situational awareness*, but it concerns the awareness of the situation, so *situation awareness* is preferred here.

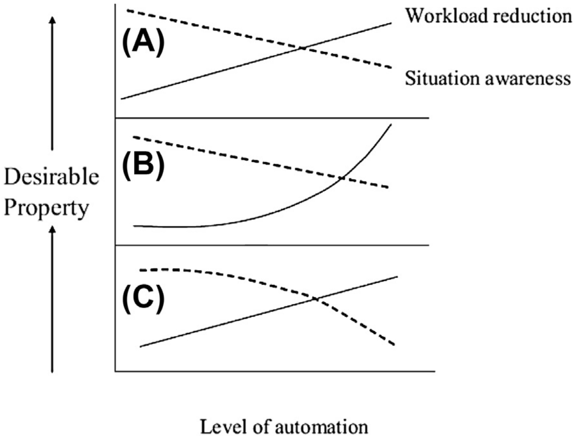

FIGURE 8.5

At higher levels of automation situation awareness (SA) tends to decrease, while at the same time workload reduces. The question for design is which of the three hypothetical curves is most real and where is the optimum? According to Wickens,[17] in case (a) it depends on the weighting of workload reduction versus SA; in (b) it will either be at the high or at the low end; and in case (c) it could be somewhere in the middle where the (absolute) values of the slopes are the same.

of maintaining consistency among four concepts: ability, authority, control, and responsibility, which are shared between the human and the machine in different ratios. Ability enables control, authority the right to decide, control the power to influence, and responsibility is being accountable for acting. The latter of course is relevant when something goes wrong and liability becomes an issue. Is the machine at fault or the operator? It turns out to be a delicate, even dynamic, balance which can be shifting in time, depending on the type of HMI and the work conditions. In full automation it will be all at the machine side, but in many situations the human keeps the authority and responsibility by having an essential part of the control. No clear, fixed rules can be given.

8.5.3 CONTROL ROOM ERGONOMICS

What does improve continually is the working environment from the point of view of ergonomics. Lighting, positions, console design, screens, the graphical user interface with (intuitive) display contents, colors, size are all optimized based on experience and experiment as indicated in the ASM Guidelines.[14] The way information is presented is also optimized with respect to decision making. The more is automated, the more complex are the questions remaining to be solved by the operator. According to Rasmussen's skill, rule, knowledge-rule (Chapter 6) this means a shift to higher cognition requiring more time for deciding to avoid error. On the other hand, the modern systems enable operators to call in data and communicate with each other via their screen regardless of distance, which facilitates decision making. But alarm and response of an operator is still considered to have on average a probability of failure on demand larger than 1 in 10. Hence, the combination cannot be classed as SIL 1. The above forms a compelling reason to prioritize alarms so that an operator can focus his mind on the problem requiring his attention most urgently.

8.5.4 TRAINING FOR OBTAINING THE MENTAL IMAGE OF THE PROCESS

All this increases the risk of operators losing physical contact with their plant. They don't smell it, and they don't hear it. This is certainly the case with plants remotely operated over large distance, underground, or out at sea. To build up a mental model of the process, necessary at the time of an abnormal situation, good knowledge, training, and physical experience are crucial. For training, simulation or "virtualization" with "human in the loop" shall be further developed but will not easily replace the real plant. A simulator for training shall be carefully developed to avoid being exposed to incomplete and simplified "reality" and learning a wrong procedure to solve a problem. In view of Leveson's system approach and operations management, not only a one-time training but also retraining shall be considered. The effect of training is fading out over a period of a year or so. Hence, training shall not be seen as static but as dynamic, in which personnel periodically combine, as with Bayesian updating. The goal of the retraining sessions is to attain progressively higher levels of effectiveness and reliability.

8.5.5 ALARM MANAGEMENT

As a result of the thousands of controlled variables in a plant, each equipped with Lo—Lo via Lo to Hi and Hi—Hi alerts or alarms, in case of a trip, breakdown of a component or other failure or abnormal situation or even achieving a designated target, alarms begin to annunciate on the screens or maybe alert loudly. So, an alarm's main task is to help keep the process within a safe and desired envelope and avoid ESD. Alarm annunciation is accompanied by data as time stamp, type of alarm, process variable tag, priority, and message. *False alarms* shall be avoided since they undermine safety culture. False-positive is just false; false-negative is when the alarm should have been activated. Because the many interacting and mutually influencing process variables (connectivity) propagate a disturbance, one alarm may be followed by many other alarms, up to dozens and higher. *Alarm flooding* increases operator's stress level, which may seriously decrease his ability to make the right decisions. This problem has been recognized since the 1994 accident at Milford Haven[19] in the UK, and there have been quite a few efforts to restrain the avalanches (some of these measures will be elucidated below).

Because of the seriousness of the problem, the Engineering Equipment and Materials Users' Association (EEMUA) published a guideline[20] and ANSI/ISA[21] published standard 18.2 in 2009. In addition, there is a host of literature available, amongst others a Guideline of the ASM Consortium[14] and by UK HSE.[22] The detailed EEMUA publication distinguishes types of alarms for a variety of purposes, how an alarm will be made effective, how it should be set, how alarms should be structured, how alarms can best appear on a display, priorities in handling alarms, et cetera. Highest priority shall be given to restraining alarm floods by rationalization/redesign. (Related) nuisance alarms and chattering (itself repeating) alarms should be suppressed. For prompt handling, it is recommended to reduce the alarm rate to *one per 10 min*. Much can be done to curb alarm flooding in the engineering stage by setting up an alarm database with settings and all other data. HAZOP-ing on P&ID and performing risk assessment will provide information for setting priorities, to identify groupings that can be automatically suppressed following the first alarm, which alarms should not be suppressed, et cetera. In an operational stage, the commonly too-large alarm frequency is addressed by "alarm management," that is, trying to identify clusters, taking account of connectivity, and analyzing the most frequent "bad actors."

In 2004, Brooks et al.[23] published an EPSC-awarded method applying a parallel coordinate transformation to display Hi—Hi and Lo—Lo alarm limits of all relevant process variables, together with an operating point for each of them and their operable ranges. By connecting the points in one display (reproduced in Figure 8.6) as zig-zag lines in outer red, center-blue, and a safe green envelope contour in between, respectively, an operator gets a more realistic overview. Shah and coworkers published several papers on graphical tools to trace related and redundant alarms in operating plant; a problem overview paper is by Izadi et al.[24] Kondaveeti et al.[25] developed based on the Jaccard similarity index applied to a high density alarm plot of, for example, one week of operation on an alarm similarity color map. Another one by Yang et al.[26] correlates points of time of specific tagged alarms by first superposing at each point a suitable Gaussian kernel function so that a wavy time series arises. Subsequently, the maximum cross-correlation coefficient is determined with the series shifted by a time lag. Other possible similarity analysis

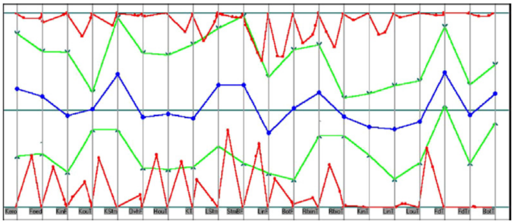

FIGURE 8.6

Display of process variables by Brooks et al.[23] for a distillation column example to keep a controlling wall temperature at a desired level: center plotted line; in an operating window: gray middle lines; within alarm dead-band limits: outer plots (on a screen the plots would have the colors blue, green, and red respectively).

methods are also described in the literature. A more direct alternative method to remove chattering alarms after detection is by introducing a suitable delay time. This is proposed by Wang and Chen.[27] The delay time is calculated by making use of previously observed patterns.

However, there remains the abnormal situation that needs to be detected and dealt with. In a technical sense if individual alarm settings can be made to avoid false and chattering alarms, the root of the remaining problem looks to be the connectivity. If in the cause—effect chain the right basic cause can be found quickly, in principle just one alarm can be activated and the connected cluster does not need to join. Signed digraph has been suggested to model connectivity. This could make a difference when combined with a quantifiable model digraph network in which time and probability are represented. At first sight, Petri net with its capability of simulation of timing seems better suited for this purpose than Bayesian net. Petri net was also the choice of Chao and Liu,[28] who proposed in 2004 a timed Petri net (named by them a Time Constraint Petri Net or TCPN). At closer look no PN would be needed, if a hidden cause of a safety critical upset can be established based on observables fast. Cause-effect with respect to alarm connectivity is predetermined by design and process models and can be called on when the cause is identified.

As we saw in Chapter 7, after a Blended Hazid once an upset shows up a causal graph can be called up by an operator without delay. However, for prioritizing repair actions historical data of probability values of different causes for a particular upset would be desirable. Meanwhile, Hu et al.[29] based on multi-level flow modeling supported HAZOP (see Chapter 7, Section 7.3.3), historical process upset data, a dynamic BN structure learned from the data and a two-step inference achieved the desired objective. Thus, the relatively easy-to-construct BNs will be helpful for both improved alarm management and situation awareness. The latter is shown by Naderpour et al.[30] for two examples of processes in which hazard can develop. The authors constructed dynamic BNs based on the process bow-ties and called it a situational network. Root nodes are fed by observable process variable values from the SCADA system producing risk indicators, while risks of abnormal conditions as operators would assess them are calculated by applying fuzzy logic. An alarm will sound when a risk threshold is passed that is still lower than critical, so that the operator can anticipate the way the abnormal situation may develop, to identify the cause from a number of different causes possible and prepare his decisions for recovery. This will relax the time pressure the operator will sustain otherwise. An alternative idea is to run a simulation of the process in parallel with the process. In the case of an incipient deviation of a known or presumed cause, the effects can be accelerated in the simulation. Here, though, one has to be sure about the cause—consequence scenario, otherwise confusion may result.

Adhitya et al.[31] performed a human factors study to investigate operator accuracy in establishing a diagnosis given an upcoming abnormal situation detected by a system called Early Warning. This system predicts a critical alarm shortly before it actually annunciates. The research applied third- to fifth-year chemical engineering students instead of (experienced) operators. To compensate, the process systems chosen were kept relatively simple. The students were given brief training while they were also given a task survey to find out whether their knowledge level was assessed sufficient. Time to diagnosis and accuracy of diagnosis were measured with and without Early Warning decision support. Although diagnosis with the support occurred faster, accuracy was not better and varied, depending on the scenario, roughly between 50% and 90% correct. Hence, not only early warning but also guidance to the right cause of upset is needed.

8.6 START-UP, SHUT-DOWN, AND TURN-AROUND

The "abnormal" situations of in particular start-up and turn-around, or unscheduled shut-down, are subject to a very substantial part of the incidents occurring in a plant. At start-up this is due to the rapidly changing conditions and possible lack of routine, while at turn-around there are many parallel maintenance activities, which are not all routine, and many exposed workers are present. Little systematic analysis has been done with respect to these process episodes. Only Ostrowski and Keim[32] addressed the issue in a very useful series of articles. The Transient Operation HAZOP that they developed focuses on operational tasks and procedures. They make a number of very practical recommendations. Guide words are *Who*, *What*, *When*, and *How long*. The exercise shall be preceded by a tour of the facility.

As already mentioned, errors in executing procedures in abnormal situations may lead to serious accidents. As mentioned in Section 8.5, Williams[15] reported on an effort to improve this situation by computer assistance to the operator or partly or fully automating procedures. However, a systemic error in an automated process shut-down may also be detrimental. Van Paassen and Wieringa[33] quote Lind's Multilevel Flow Modeling, which we briefly described as applied in HAZOP automation in Chapter 7, Section 7.3.3, as an excellent tool for reasoning when developing procedures for start-up and shut-down, whether or not automated.

8.7 CONCLUSIONS

Process control development is toward Lyapunov function—based distributed model DMPC, which besides improved control effectiveness and economic optimization also to some degree enables fault diagnosis of valves and sensors. The latter topic has been the subject of an extensive review by Venkatasubramanian et al.[3] Over the years, many methods have been proposed. Lately, higher-order statistical treatment of time series determining bicoherence in nonlinear systems offers much perspective.

A brief description of the newest control equipment showed many improvements, also with respect to security. In addition, system and control room ergonomics have improved the operator's work environment, which also stimulates situation awareness. Alarm management and the avoidance of alarm floods is still a "hot" topic. In recent years, a great variety of approaches have been proposed to relieve the problem. More insight in the causal structure of an abnormal situation, obtained by applying methods such as Blended Hazid on the design of the plant, may help to unveil connectivity. Installing its result as computer software in the control room and reinforcing it in the future by a BN may enable penetrating via the causal structure into the core of the problem.

Together with the signals collected by the SCADA system, the above will provide a step toward the prognostic process control tool that Venkatasubramanian[4] foresees. This could be further supplemented with safety management system performance indicator value information. This would enable that not only the direct, short-term risk factor effects may be seen but also the risk factors with effects on culture over a longer term as suggested in the holistic approach by Pasman et al.[34] The prognostic tool would then become even more universal and could in a more popular way really be termed a "safety dashboard."

REFERENCES

1. Stephanopoulos G. *Chemical process control, an introduction to theory and practice*. Englewood Cliffs (N.J.): PRT Prentice Hall; 1984. ISBN 0-13-128629-3.
2. Christofides PD, Scattolini R, Muñoz de la Peñad D, Liu J. Distributed model predictive control: a tutorial review and future research directions. *Comput Chem Eng* 2013;**51**: 21−41.
3. Venkatasubramanian V, Rengaswamy R, Kavuri SN, Yin K. A review of process fault detection and diagnosis, part I: quantitative model based methods. *Comput Chem Eng* 2003;**27**:293−311. Part II: Qualitative models and search strategies. *Comput Chem Eng* 2003;**27**:312−26; Part III: Process history based methods. *Comput Chem Eng* 2003;**27**:327−46.
4. Venkatasubramanian V. Systemic failures: challenges and opportunities in risk management in complex systems. *AIChE J* 2011;**57**(1):2−9.
5. Christofides PD, El-Farra NH. *Control of nonlinear and hybrid process systems: designs for uncertainty, constraints and time delays*. Springer; 2005. Monograph.
6. Ellis M, Christofides PD. Integrating dynamic economic optimization and model predictive control for optimal operation of nonlinear process systems. *Control Eng Pract* 2014;**22**:242−51.
7. Kazantzis N, Kravaris C, Wright RA. Nonlinear observer design for process monitoring. *Ind Eng Chem Res* 2000;**39**:408−19.
8. Mhaskar P, Liu J, Christofides PD. *Fault-tolerant process control, methods and applications*. Springer; 2013. ISBN 978-1-4471-4807-4. http://dx.doi.org/10.1007/978-1-4471-4808-1. ISBN 978-1-4471-4808-1 (eBook).
9. Vaidhyanathan R, Venkatasubramanian V. Digraph-based models for automated HAZOP analysis. *Reliab Eng Syst Saf* 1995;**50**(1):33−49.
10. Dash S, Rengaswamy R, Venkatasubramanian V. Fuzzy-logic based trend classification for fault diagnosis of chemical processes. *Comput Chem Eng* 2003;**27**:347−62.
11. Shoukat Choudhury MAA, Shah SL, Thornhill N. Diagnosis of process nonlinearities and valve stiction, data driven approaches. In: *Advances of industrial control*. Springer; 2008. ISBN 978-3-540-79223-9.

12. Shoukat Choudhury MAA, Shah SL, Thornhill N. Diagnosis of poor control-loop performance using, higher-order statistics. *Automatica* 2004;**40**:1719–28.
13. Dulyakarn P, Rangsaneri Y. Fuzzy c-means clustering using spatial information with application to remote sensing. In: *Proceedings of the 22nd Asian conference on remote sensing, Singapore, 2001*; 2001.
14. ASM Guidelines from Abnormal Situation Management Consortium (www.asmconsortium.com) can be obtained from www.Amazon.com.
15. Williams Jr ThN. Procedural automation. In: *Proceedings 16th annual international symposium, Mary Kay O'Connor process safety center, October 22–24, 2013, College Station, Texas*; 2013. p. 126–33.
16. a. Endsley MR. Toward a theory of situation awareness in dynamic systems. *Hum Factors: J Hum Factors Ergon Soc* 1995;**37**:32–64;
    b. Measurement of situation awareness in dynamic systems. *Ibidem* 1995;**37**:65–84.
17. Wickens ChD. Situation awareness: review of Mica Endsley's 1995 articles on situation awareness, theory and measurement. *Hum Factors: J Hum Factors Ergon Soc* 2008;**50**: 397–403.
18. Flemisch F, Heesen M, Hesse T, Kelsch J, Schieben A, Beller J. Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. *Cognit, Technol Work* 2012;**14**:3–18.
19. Health and Safety Executive. *The explosion and fires at the Texaco Refinery, Milford Haven, 24 July 1994: a report of the investigation by the health and safety executive into the explosion and fires on the Pembroke Cracking Company Plant at the Texaco Refinery, Milford Haven on 24 July 1994*. ISBN 0-7176-1413-1.
20. EEMUA. *Alarm systems, a guide to design, management and procurement*. 2nd ed. London: Publication No. 191; 2007. ISBN 0-85931-155-4.
21. ISA, *Management of alarm systems for the process industries*. 2nd ed. Technical Report ANSI/ISA-18. 2-2009 International Society of Automation ISA, Research Triangle Park, NC; ISA, Alarm Management: A Comprehensive Guide.
22. Health and Safety Executive. The management of alarm systems, prepared by Bransby automation Ltd and Tekton engineering, Contract Research Report, 166/1998.
23. Brooks R, Thorpe R, Wilson J. A new method for defining and managing process alarms and for correcting process operation when an alarm occurs,. *J Hazard Mater* 2004;**115**: 169–74.
24. Izadi I, Shah SL, Chen T. Effective resource utilization for alarm management. In: *49th IEEE conference on decision and control, December 15–17, 2010*. Atlanta (GA, USA): Hilton Atlanta Hotel; 2010.
25. Kondaveeti SR, Izadi I, Shah SL, Black T, Chen T. Graphical tools for routine assessment of industrial alarm systems. *Comput Chem Eng* 2012;**46**:39–47.
26. Yang F, Shah SL, Xiao D, Chen T. Improved correlation analysis and visualization of industrial alarm data. *ISA Trans* 2012;**51**:499–506.
27. Wang J, Chen T. An online method to remove chattering and repeating alarms based on alarm durations and intervals. *Comput Chem Eng* 2014;**67**:43–52.
28. Chao C-S, Liu A-C. An alarm management framework for automated network fault identification. *Comput Commun* 2004;**27**:1341–53.
29. Hu J, Zhang L, Cai Zh, Wang Y, Wang A. Fault propagation behavior study and root cause reasoning with dynamic Bayesian network based framework. *Process Saf Environ Prot*, in press, accepted manuscript, available on line 7 April 2015.

30. a. Naderpour M, Liu J, Zhang G. A situation risk awareness approach for process systems safety. *Safety Science* 2014;**64**:173–89;
b. Naderpour M, Liu J, Zhang G. Supporting operator's situation awareness in safety-critical systems: an abnormal situation modeling method. *Reliab Eng Syst Saf* 2015;**133**:33–47.
31. Adhitya A, Cheng SF, Lee Z, Srinivasan R. Quantifying the effectiveness of an alarm management system through human factors studies. *Comput Chem Eng* 2014;**67**:1–12.
32. Ostrowski SW, Keim KK. *Tame your transient operations, use a special method to identify and address potential hazards, chemical processing*; June 2010. http://www.chemicalprocessing.com/articles/2010/123/?start=0.
33. Van Paassen MM, Wieringa PA. Reasoning with multilevel flow models. *Reliab Eng Syst Saf* 1999;**64**:151–65.
34. Pasman HJ, Knegtering B, Rogers WJ. A holistic approach to control process safety risks: possible ways forward. *Reliab Eng Syst Saf* 2013;**117**:21–9.