Applications of Object Oriented Bayesian Networks for Causal Analysis of Process Disturbances

G. Weidl
Galia.Weidl@iff.uni-stuttgart.de
Institute of Industrial Manufacturing and Management, University of Stuttgart
70 569 Stuttgart, Germany

A. L. Madsen
Anders.L.Madsen@hugin.com
Hugin Expert A/S,
Niels Jernes Vej 10, 9220 Aalborg, Denmark

E. Dahlquist
Erik.Dahlquist@mdh.se
Department of Public Technology
Mälardalen University,
S-721 78 Västerås, Sweden

Abstract. We discuss a hybrid approach for causal analysis of disturbances in industrial process operation. It represents a combination of OOBN with first level diagnostic packages and physical models serving as agents in the system design and providing evidence for automated reasoning on abnormality in process operation. The aim is causal analysis of non-measurable disturbances as a decision advice complement to the distributed control system (DCS). The approach includes prediction of signals' level-trend development, risk assessment for disturbance analysis and predictive maintenance on demand. The methodology has been applied on a screening process with a pressure-flow network in a Pulp Mil.

1. Introduction[1]

Industrial systems grow in their complexity. A sophisticated industrial process can generate output from hundreds or thousands of sensors, which should be monitored continuously. In the case of deviations from normal process conditions, the relevant information should be singled out, the cause of the failure should be found and appropriate corrective actions should be taken.

The sheer amount of data and the continuity of the process ask for a high level of automation of operation and maintenance control. But not all operations can be completely automated. Often it is necessary to let a human operator steer the process in critical situations. This poses a formidable challenge on the concentration and capability of the human being and on the efficiency of his decisions.

To support the operator in the task of disturbance analysis, an adaptive system for Root Cause Analysis (RCA) and Decision Support (DS) will collect the data from the sensors and transfer it into structured and relevant information. Simultaneously, the process overview should be maintained and relevant explanations provided with an advice on corrective sequence of actions. Then the operator can make educated decisions, based both on both artificial intelligence and human experience. This should help avoid unplanned production interruption or at least ensure that the lost production is minimal.

If only a classification of the failure type is required, neural networks or statistical classifiers may be more adequate. However, if decision support is needed, Bayesian networks (BN) for probabilistic reasoning in intelligent systems [1], [2], [3] can be used to calculate the posterior probabilities and have the ability to adapt to changes [4].

Troubleshooting based on decision theory was first proposed in [5], and further analyzed in [6]. In a pre-study, we have also considered neuro-fuzzy hybrid systems as an alternative approach [7]. The neuro-fuzzy approach would not provide causal interpretation of diagnostic conclusions, which was one of the main system requirements for explanatory decision support on demand or continuously. Models reusability, simple construction and modification of generic BN-fragments were other selection criteria in favor of object oriented BN (OOBN) [8].

1.1. A Bayesian Networks Approach to Causal Modeling of a Domain

A Bayesian network (BN) is a model of a domain containing uncertainty. For industrial processes, the uncertainty can be originating from the incomplete understanding of the complexity of the domain, from occurrence of stochastic events leading to randomness in the process behaviour, from the process condition at the time a given control or maintenance actions is to be performed, or a combination of these.

A Bayesian network is a set of child and parents nodes representing random variables and a set of links connecting these nodes to build a directed acyclic graph (DAG) [3, 14]. DAG expresses the absence of directed path(s) starting and ending at the same node. The nodes correspond one-to-one with the domain variables of the probability distributions such that there is one conditional probability distribution (CPD) function for each child node given its parents. The CPD expresses the strengths of the (causal) dependency relations of the child node from its parent's configurations of states. Bayesian networks are also called belief networks, Bayesian belief networks or causal probabilistic networks.

The generic mechanism of disturbance (or failure) build up consists of root cause activation, which causes abnormal changes in the process conditions. The last represent effects or symptoms of abnormality. They are registered by sensors or soft sensors. If not identified and corrected, these

[1] This work was performed, while G.Weidl was associated with ABB Corporate Research and E.Dahlquist with ABB Process Industries, both in Västerås, Sweden

abnormal conditions can enable events causing an observed failure. A causal representation of the above factors gives the following chain of events and *transitions*, which is of interest for RCA under uncertainty and for the purpose of decision support on corrective actions, as shown in Fig. 1, column 1.

The BN model for Root Cause Analysis reflects the causal chain of dependency relations as shown in Fig. 1. The dependency relations are between the three symbolic layers of random variables in the problem domain, i.e.

(1) $\{H_i\}$, $\{S_j\}$, $\{F\}$, where $i = 1... n$, $j = 1 … m$

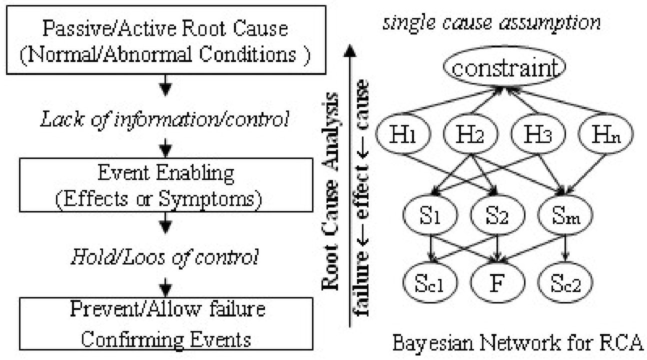

Fig. 1. The conceptual layers of the BN for RCA (column 1) and the corresponding variables in each layer of the BN

In (1), the set of root causes $\{H_i\}$ contains all possible failure sources or conditions, which can enable different events S_j, which precede a failure F or its confirming events Sc_k. The set of variables $\{S_j\}$ contain also early abnormality effects and symptoms, which are observed, measured by sensors or computed by simple statistical or physical models (e.g. mass and energy balances). The three sets of variables $\{H_i\}$, $\{S_j\}$, $\{F\}$, can be viewed as three conceptual BN layers (i.e. root cause ® effect ® failure), see Fig. 1.

An *Object-Oriented Bayesian Network* (OOBN) is a network that, in addition to the usual nodes, contains *instance nodes*. An instance node is a node representing an instance of another sub-BN, which can itself contain instance nodes, whereby an object-oriented network can be viewed as a hierarchical description (or model) of a problem domain.

The use of OOBN for facilitating the construction of large and complex domains, and simple modification of BN fragments was discussed in 8.

We use this idea to model industrial systems and processes, which often are composed of collections of identical or almost identical components. Models of systems often contain repetitive pattern structures (e.g. models of sensors, actuators, process assets). In Bayesian networks, such patterns are network fragments, modelled as sub-BN. In particular, we use OOBN to model all (DCS and computed) signals uncertainties and signals level-trend classifications as small standardized sub-OOBN or fragments within the OOBN of the problem domain.

We also use OOBN for top-down RCA of industrial systems, which allows different levels of modeling abstraction in the plant and process hierarchy. A repeated change of hierarchy is needed partly due to the fact that process engineers, operators and maintenance crew discuss systems in terms of process hierarchies and partly due to mental overload with details of a complex system in simultaneous causal analysis of disturbances. It also proves to be useful for explanation and visualization of analysis conclusions, as well as to gain confidence in the suggested sequence of actions.

In this work, we have used the Hugin software [13] ,]14], which supports the construction of *hierarchical network structures*. A fully object-oriented paradigm for constructing Bayesian networks should also include the notions of subclasses and inheritance, as known from object-oriented programming languages.

1.2. An overview of causal and non-causal object oriented modelling languages

In general, the object-oriented (OO) constructs incorporate encapsulation, inheritance and hierarchy. Its common purpose is efficient modelling and simulation, providing a convenient language for reuse and exchange of models.

Examples of non-causal object-oriented languages for modelling of physical and chemical systems and processes include: Modelica, gPROMS, ASCEND, NMF/IDA, Omola, etc. As compared to the known simulation languages, these object-oriented languages offer several advances: 1) non-causal modelling based on differential and algebraic equations, describing the physics of the domain, provided it is well understood; 2) multi-physics domain modelling within the same application model, incorporating a combination of electrical, mechanical, thermodynamic, hydraulic etc. sub-models; 3) a general type system that unifies object-orientation, multiple inheritance, and templates within a single class construct [20].

Thus, the main difference is in the causal probabilistic handling of uncertainties in our OOBN approach (exploiting hybrid information obtained from both measured and calculated variables from physical models), while the above mentioned OO languages are using non-causal modelling, based on differential and algebraic equations of the problem domain. On the other hand, we use a hybrid approach, which is a combination of OOBN with first level statistical diagnostic packages and physical models (e.g. pressure-flow nets) serving as agents in the system design and providing evidence for automated reasoning on abnormality in process operation.

1.3. The OOBN approach and contribution overview

The main advantage due to the Bayesian approach is much faster operator guidance, *without limiting* the failure analysis to only one possible root cause. Instead, a list of root causes ranked after probabilities will give quick and flexible decision support to the operator with explanation facility based on causality.

We have previously developed a number of generic sub-OOBN for signals classification, process performance monitoring and diagnosis [9-12].

With this work, we present an application of a combined methodology for root cause analysis and decision support. The main contributions of this work include:

- The use of OOBN for RCA & DS in process operation to ensure causal modeling of interdependency of events; to ease the construction, reusability and modification of BN; to reduce the overall complexity of the network; to provide explanations with overview at different levels of industrial plant hierarchy
- The development of OOBN model for adaptive signal classification by mixture models and prediction of the development of signals' level-trend.
- The system design for Root Cause Analysis (RCA) & Decision Support (DS), incorporating the agents for handling of uncertainties and reflecting the information flow
- The development of pressure-flow network as physical (non-causal) model of the screening process. It provides soft sensor (non-measurable) information for evidence in the reasoning under uncertainties
- Risk assessment of disturbances, estimation of their most probable root causes for predictive maintenance on demand.

The most efficient sequence of corrective actions is obtained from a probability-cost function. The cost is represented by the expected average cost of corrective (process and asset management) actions. We demonstrate the application on a pulp screening process.

2. Statement of the Problem

Disturbance analysis (RCA and DS) in industrial process control could be a time-consuming task leading to big production losses. The overall goal of RCA and DS is to extract from DCS-data volumes the necessary information for early assessment of abnormalities and provide efficient troubleshooting advice in process operation and for maintenance on demand.

The following issues are treated in this paper. The disturbance analysis system should provide reliable handling of uncertainties in acquisition of knowledge and data, including both discrete and continuous signals. The signal classification should be adaptive to changes in process operation mode and account for both normal and abnormal/faulty operation conditions. Prediction of the level-trend development should ensure early risk assessment and warning on abnormality in order to be able to propose early treatment using efficient sequences of actions. Therefore for predictive maintenance on demand, the cost estimations should also anticipate the potential production losses.

The system performance should adapt to natural process changes and allow user interaction. An operator steering a complex process will prefer a transparent decision support system to a black box system. If the system explains the underlying mechanism for its conclusions and suggestions, the operator can compare these with his experience and take the needed corrective actions with confidence.

2.1. Conditions for a process or device

Conditions are the process states at a particular time. Conditions are used to determine whether a plant-wide disturbance analysis should run or not. Process condition is a condition that depends on the state of a production process. The state is affected by external factors. An abnormal condition is a condition caused by disturbances that prevent the process parameters to stay within control limits that define the range of normal process operation. It causes degradation of targeted process performance, resulting in the inability to deliver a pre-specified state of output. A critical condition is a condition that causes failure to meet targets, unexpected process destructive effects or dangerous consequences. It requires urgent corrective actions.

2.1. Interaction of Condition Monitoring, RCA and DS

RCA is a structured procedure, which guides the failure analyst from the disturbance or failure event to its cause(s). In standard process control the deviation of a single parameter outside its normal range will trigger an alarm. To prevent a large number of false alarms, the thresholds of the variables should not be chosen too sensitive. But this approach will indicate failures only late at an advanced stage.

Process condition monitoring interacting with RCA will use more sensitive thresholds. The large number of triggered “alarms” is first analyzed internally by the RCA system. Only if the change of some variables in context with the behavior of all other process parameters suggests the development of a failure, the operator is informed and advised on actions.

3. Methodology Used

The task of failure identification during production breakdown, its isolation and elimination is a troubleshooting task. While the task of detecting early abnormality is a task for adaptive operation with predictive RCA and maintenance on demand. Therefore, these two tasks have different probability-cost functions as discussed in [11]. We combine both tasks under the notion of asset management. It aims at predicting both process disturbances and unplanned production stops, and to minimize production losses. Thus, the priority is set to determine an efficient sequence of actions, which will ensure the minimal production losses and will maximize the company profit. To provide a solution to troubleshooting and maintenance on demand, an extended methodology is suggested, able to:

1. detect a failure at an early abnormality stage with a reduced number of hardware sensors
2. find the most likely causes of abnormality
3. propose an efficient sequence of corrective actions and observations

For any abnormal case, once identified, the system is searching to find the root cause of observed or predicted disturbance. The basic algorithm of RCA, as implemented in this application, is a modification of the decision-theoretic troubleshooting algorithm, where costs are assumed to be order independent to ensure an optimal sequence of actions [6]. We have used the estimated average cost, as described in [18] and [19]. Moreover, we have extended this methodology to pro-active solutions with early treatment of abnormality in order to avoid potential losses of production [19].

3.1. Handling of Uncertainties

The necessary data to determine the condition of a process and its devices is provided by DCS-signals, alarms, event lists, equipment data, maintenance reports, and a number of first level diagnostic packages (Fig. 2).

3.1.1. Agents for Handling of Uncertainties

First level asset diagnostic packages serve as agents in the system architecture. These include diagnostics of small asset units, e.g. sensors, actuators, control loops, soft sensors, see Fig. 2. They provide information on the degree of reliability of sensors readings (by data reconciliation), sensor status (by sensor diagnosis), calculated signal-trends (by trend diagnosis), actuators and other process assets conditions. This reduces the degree of uncertainty in the acquired evidence. Asset is used here as a collective notion to include actuators (valves, pumps), other process assets (e.g. digester screens; pipes, can be represented as fake valves) and in general, even equipment failures as a root cause of signal deviations. More details on the system architecture are given in [10].

In Fig. 2, the Dynamic Data Reconciliation Agent is utilizing simulations from a Pressure-Flow Network, which we describe in the following sub-section.

3.1.2. Fluid dynamics models and uncertainties in knowledge acquisition

The idea is to use the simulations provided by a fluid dynamic model (e.g. pressure-flow net) as soft sensors' evidence from thereof computed (non-measurable) process variables. A pressure-flow model can provide for example estimates on some parent configurations in the BN. There are several sources of uncertainties in this physical model estimation, since modeling inputs for the actual valve openings might be different than the ones indicated by DCS measurements. Moreover, the state of the screening plate (normal, clogged, hole or cracks) will still represent uncertainty of the outcome of such estimations. This is because clusters of small particles or long fibers in the pulp flow can clog part of the plate screening area, which is not directly considered in the flow dynamics model. One can also model this effect in the fluid dynamics simulations by a function expressing the gradual reduction of a plate screening area.

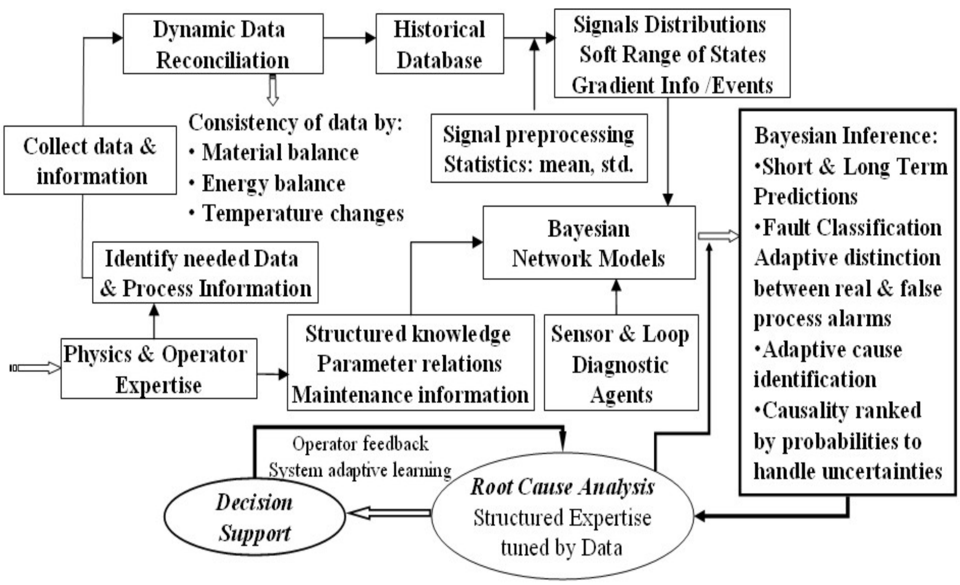

Fig. 2. System Design with Information Flow for Root Cause Analysis and Decision Support under Uncertainties

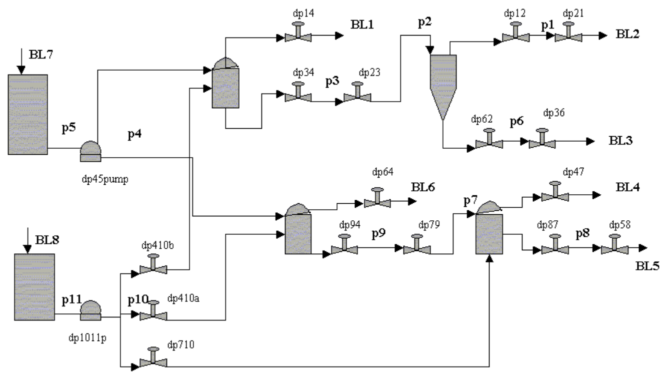

Fig. 3. Pressure-flow network of pulp screening process with screens in different process sections and series connection of screens

A pressure-flow model (see Fig. 3) can be used to specify mathematical expression of the relation between a node *(flow_accept or flow_reject)* and its parents. Two types of equations can be used to build the pressure-flow equation system of the screening process:

- Mass conservation equations for all flow splitting-points

(2) $$S\ In\_flow = S\ Out\_flow$$

which for the screen becomes

$$flow\_inject + flush\_flow = flow\_accept + flow\_reject$$

- Pressure drop equations at all pressure changing components of the network. Three types of pressure changing components *D(Pressure_at_component)* are considered: pumps, valves and screening plate

$$Pressure\_after=Pressure\_before+D(Pressure\_at\_component)$$

The pressure change due to a pump is given as:

(3) $$Dp\_pump = dp_0 - [\ (dp_0 - dp_n)*q^2 / q_n^2\ ]$$

where dp_0, dp_n, q_n are the dimensioning parameters of the pump, i.e. dp_0 - the pressure at flow $q=0$ through the pump; dp_n the pressure at normal flow $q= q_n = [N\ kg/h]$.

The pressure change due to a valve is given as:

(4) $$Dp\_valve = r*q^2 / (a^2 v^2)$$

where q is the flow, r is the density of pulp, a is the admittance factor and v is the valve opening in % units. Due to non-measurable process disturbances, the parameters r, a and v incorporate uncertainties in the expressions for *Δp_accept_valve, Dp_reject_valve and Dp_screen_plate.*

Usually, the pulp flow concentration (density) measurements are unreliable. The building and dynamics of fibers and clusters in the pulp flow are not modeled by physical models, or even if modeled - not calculated on-line since they are computation time expensive. The active screening area in relation to the clogged plate surface is difficult to estimate. The flow of accepted pulp consistency can be reduced due to many other factors, besides screen plate clogging. Uncertainty in measurements is one of the motivations for adaptation.

Uncertainty in computed pressure-flow balance is another argument for adaptation. The pressure-flow equation system build from (2), (3), (4) can be expressed as: $f(x) = 0$, where the vector $x = \{q, p\}$ is its solution with components $q = \{ q_1, q_2, ..., q_n\}$ for the flow and $p = \{ p_1, p_2, ..., p_m \}$ for the pressure.

We use the Newton's iteration method to find the solution $x = \{ p, q\}$ of the pressure-flow equation system, as follows:

(5) $$x^k = x^{k-1} - J^{-1}(x^{k-1})\, f(x^{k-1}), \quad k = 1, 2, 3, \dots$$

where the Jacobian of the system is given by

$$J(x) = \{¶f_i(x)/¶\, x_i\}.$$

At $k = 1$, $x^{k-1} = x^0$ is the *initial guess* of the iteration procedure. Since the Newton's iteration method represents tangential search, as closer is the initial guess to the real solution, as bigger is the chance to find the correct pressure-flow net configuration and thus keep the system control in balance.

The above provides good motivation for using adaptation. The Newton iteration provide good mathematical model as estimates on $x = \{q, p\}$. The pressure-flow balance should hold at each time step and is computed in quasi-stationary regime (by use of the general purpose software *Maple*), since there is a possibility to change at each time step the initial flow entering the system by taking into account process input changes. The adaptation will then compensate the classification of those states of the system, which the models do not fully capture, since the domain changes over time.

A fluid dynamics model gives a correlation between flows and pressure drops. In many modern valves it is custom to include measurements of pressure before and after the valve opening, as well as the position of the ball or slider in the valve. This gives a great chance to make flow measurements as well as pressure measurements all over modern plants in the future, when advanced busses like profibus and field bus foundation and other s are being used more extensively. This will give better information on the process performance as well as the performance of the valves, all flow meters and other types of sensors. Today this information is just measured locally in the valves, but in the future they will be registered also in the process computers and process data bases.

Pro- and contras in knowledge acquisition

Physical simulations models (like fluid dynamics, or intergated fluid dynamic models with built in control strategies) are more rigorous, although time consuming and sometimes unsuitable for on-line use. In this respect, the knowledge of process physics can be used with advantage for causal probabilistic (BN) modeling and exploited (time efficiently) on-line. In cases, where the underlying physics of the problem domain is not well understood, the only alternative left is to use simplifying assumptions in BN as described in [3] and [10].

3.1.3. Continuous Distributions on Soft Range of States

The system is receiving as evidence both discrete and continuous signals. The event lists contain only Boolean variables. The variation range of continuous signals (DCS-measured or thereof computed signals (e.g. from physical models)) are also discretized into a number of soft numeric intervals, represented as states of a BN node. We use discretization and not continuous nodes explicitly, since we want to capture both the continuous variation of the signal during normal process operation, as well as its non-continuous disturbances (or discrete faulty deviations outside normal variations). This is realized by use of mixture models [15], [16].

Let S be a continuous variable. Assume that S can be partitioned into sets $s_{1\ \dots}\ s_n$ such that the probability density function $P(S)$ can be approximated by a finite sum over its n soft interval states s_i

(6) $$P(S) = S_{i=1..n} P(s_i)\, P(S|s_i)$$

i.e. $P(S)$ is partitioned into n sub-CPD $P(S|s_i)$, each with probability $P(s_i)$ as a root cause of S. Gaussian mixtures are used most commonly as sub-CPD, since they allow to approximate any other probability distribution.

The sub-distribution of each soft interval state is chosen as a localized function with one peak and it is decreasing monotonically with the distance from the peak. Gaussian mixtures are used most commonly, since they allow to approximate any other probability distribution. We also use Gaussian distribution on selected soft interval states to represent the most characteristic values of a continuous signal during normal and faulty operation.

3.2. Generic OOBN Models

3.2.1. Adaptive Signal Classification

The process is usually operated at several normal operation modes dependent on production rate, process load, etc. During standard operation modes, the variations of process variables are inside their boarders allowed from process control. Faulty change of operation mode, faulty process operation, as well as asset faults can be the root causes of abnormal process deviations. This can cause degradation of process output (e.g. quality, quantity) or failure in process assets, when exploited under improper conditions.

Dependent on operation mode and set-points c_p of parameters, the signal's level and trend have different normal and abnormal states, see the first slice (at time t_0) of Fig. 4. Normal operation mode is characterized by a number of set-points and their typical signal variations under normal and abnormal process operation. From data analysis, we have found that certain failure events are enabled during process transition between consequent operation modes (e.g. mode change for increase of production rate). For such cases, it is necessary to use signal classification, which is adaptive to changes in normal process operation.

The node sensor reading s_r represents the continuously measured value of a process variable. The node sensor status s_s represents the condition of the sensor instrument used to perform the measurement of a process variable. The node real value R_t represents the actual development of the process variable at time t. The node sensor diagnosis s_d receives input on the sensor status from the sensor diagnostics agent. The sensor diagnostics conclusions are affected by the sensor status (true/false), while the real value can be restored by use of information from the dynamic data reconciliation agent (Fig. 2).

Fig. 4. Prediction of classified DCS-signals, based on present and past values of process variable

Suppose a DCS-signal S is discretized over soft interval states as given in (6). In cases of faulty operation or root causes originating in abnormal condition of process assets, the real value of a process variable is dependent on the set-points of different operation modes and on the status of the asset A_s. Then, the conditional probability distribution (CPD) of the real value becomes a mixture of two Gaussian distributions around the set point during normal and abnormal process operation

$$P(R_t|c_p, A_s) = \begin{cases} Normal\ (c_p,\ x_{\pm}{}^{\chi} s), & A_s \\ Normal\ (m_{abn},\ x_{abn}{}^{\chi} s_{abn}), & o.w. \end{cases}$$

For variables, which are directly manipulated in process control loops, the set point serves as mean m in the Gaussian distribution with the respective scaled variance $x_{\pm}{}^{\chi} s$ around the set-point of the operation mode. The variance is scaled by a factor $x_{\pm}$ in order to avoid too many "*internal alarms*" for RCA. For variables without set-point, but which covariate with controlled variables, we calculate the mean of the relation (real value/set point) or alternatively any physical or statistical function expressing their correlation. The dependence on the status of the asset is expressed with a Gaussian distribution, where m_{abn} is the real alarm threshold and $x_{abn}{}^{\chi} s_{abn}$ is the scaled variance which lower variation limit provides more sensitive alarm threshold for early risk assessment of abnormality. Alternatively, for signals deviation, which is characterized by low frequency of failure events, we use a mixture of Gaussian distributions on normal operation behavior and Poisson distribution during faulty operation.

It is obvious that if the measurement instrument is not properly functioning, then the *real-value* and the "*sensor reading*" need not be the same. Therefore, the sensor reading from any DCS-measurement is conditionally dependent on random changes in two variables: real value under measurement and sensor status of the instrument. Its probability distribution is expressed as a mixture of normal and uniform distributions for the real value when the sensor status is true or false respectively

$$P(s_r|R_t, s_s) = \begin{cases} Normal\,(R_t, x^{\chi} s), & s_s \\ Uniform(y_{min}, y_{max}), & o.w. \end{cases}$$

where the uniform distribution is defined on the entire interval of signal variation.

The signal trend at any time step is directly influenced by random changes in the real value at both previous and present time steps. For robustness, the trend is calculated as the derivative on the averaged time history of the signal sampled with a time step $Dt = t_i - t_{i-1}$:

$$(7) \quad trend_{signal} = DS/Dt = \{m(R_t(t_0)) - m(R_t(t_{-1}))\}/(t_0 - t_{-1})$$

where the mean $m(R_t(t_i)) = S_{j=(i-N)\ldots i} R_t(t_j)/N$ is averaged between the time points t_{i-N} and t_i ($i=0$ current time) over the real value $R_t(t_i)$ of a signal at time point t_i.

This provides a filter of the noise in the signal behavior. To model the degree of uncertainty in the signal trend, we use a diagnosis trend agent. In this case, we use the uncertainty (historically calculated) for each of the N historical points and base on this the diagnosis of the derivative variable. The considered pulp screening process has slow dynamics. Therefore, the trend is expressed by the derivative calculated over a floating interval window containing *20* points distributed uniformly on *10* minutes period.

The signal class is conditionally dependent on random changes in two variables: signal level and signal trend. Its probability distribution is then defined to adapt the classification to changing operation mode.

The generic BN model for adaptive signal classification into levels and trends (see the first time-slice of Fig. 4) is one of the generic building blocks in any RCA model of monitored industrial process with its equipment and asset components in every process section. Fig. 4 shows a past-present-future combination of such generic blocks for the purpose of risk assessment and early warnings on abnormality (e.g. screen plate clogging).

3.2.2. One Step Look Ahead Prediction of Signals

The temporal Bayesian network models are used to predict the development of the signals and evaluate their risk of abnormal deviation due to disturbances. This signals' prediction is used as evidence in the RCA model. Therefore, the RCA can provide early warnings on root cause activation. In that case, the control system can examine with short disturbance (e.g. opening or closing of valve) whether the suggested root cause is the real one and if confirmed (by the operator) the necessary corrective action is undertaken at an early stage of failure development.

In addition, temporal BNs can be used to express causality dependencies reflecting the dynamic character of the process. For example, alarm filtering deals mainly with time-delay effects. In Fig. 4 we use only three time steps to model an infinite step process. In the reality, such temporal network is a static network, since it represents a finite and fixed number of time slices and it can reason only with a finite series of observations coming from a dynamic process or system. For real time applications, it is desired to include in the model as many time slices as possible. The last can cause an inefficient and time consuming inference, since evidence propagation would involve all time slices although probability update is desired only for a limited number of time slices.

A computation scheme, which can handle infinite series of observations in dynamic BN has been described in [17]. It changes dynamically the width of the time slice window, as well as the number of backward smoothing and forecasting time slices. Thus, it can provide flexible and selective inference. It also supports inclusion and modification of time-delayed observations. The forecasting of signal development and time-delays will be incorporated in the proposed RCA system in the near future. Another issue is concerned with suitable approximations for handling of large number of temporal relations between the different time slices.

3.2.3. Early Warning Based on Risk Assessment: A Case Story on Pulp Screening

The pulp is obtained as a result of cooking of wood-chips in a digester. The screening of pulp is a filtering process. In order to predict the condition of the screening process and to demonstrate the concept, we have selected a characteristic group of process random variables: S1 is the differential pressure signal; S2, S3 are the flows on accept and reject side, S4 is the consumed power by the equipment during screening process operation. As noted for the signal classification model, the different production rates during normal screening operation are represented each by one specific combination of process variables. All deviations from operation mode-set combinations are symptoms of expected abnormality in the process or its equipment.

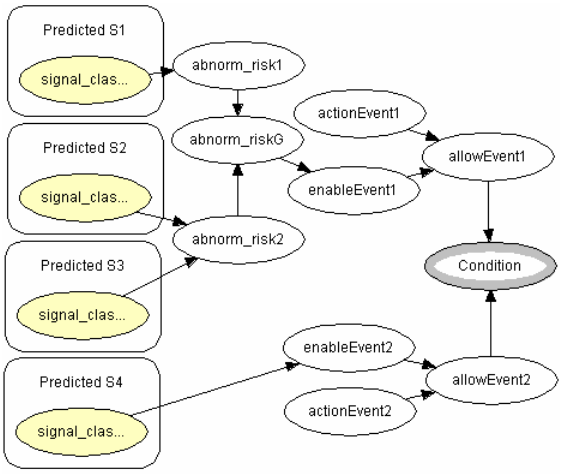

Fig. 5. Assessment of Abnormality Risk and Equipment/Sub-process Condition

Instead of reactive troubleshooting, a long term strategy requires a proactive system with early warnings and corrective (control or maintenance) actions, which prevent abnormality to develop into a failure. For this purpose, we combine in the BN model (Fig. 5) the predicted *signal class* outputs as intermediate variables for risk assessment.

This is based on certain combination of random variables (e.g. signals S1-S4 in our screening application). The pressure-flow combinations of S1, S2, S3 are responsible for a number of mutually exclusive states of the event node *enable Event*, while S4 can be the cause of Event2, which dependents on stochastic circumstances might occur simultaneously, but not always independently of Event1. When abnormality event is enabled, a corrective action from the operator/maintenance or DCS can prevent (or allow) undesired event (failure), leading to abnormal or critical condition of the equipment or a sub-process (Fig. 5).

By analogy, we build the OOBN models at higher plant hierarchy levels (e.g. process diagnosis, control and performance management levels).

3.4. Explanation and Adaptation

Based on the causal character of the OOBN models, the operator can feed his own educated observations into the inference system, which then evaluates alternative actions with respect to their technical and economical impact.

A user explanation interface should include a ranked list of most probable root causes (see Fig. 6), a list of evidence (symptoms) for inference, as well as conclusions on possible effects. Moreover, one can examine the dependency on evidence through the sensor status and update the RCA conclusions. The independence relations induced by a set of nodes in a directed acyclic graph (DAG) are determined using the d-separation criterion [1]. In case there is more than one path between the root cause and the failure, the entropy is calculated for each of the connecting paths and compared before the propagation of evidence and after it. Then, the path with the highest entropy is presented to the operator in order to explain the conclusions. For large BNs, additional properties, as coloring of the most probable scenario of causes and effects allow visualization of the explanations.

Most Probable Root Causes:	**Operator Feedback:**
Rotor drive malfunction: 0.55	True / False
Slipping rotor shaft: 0.35	True / False
Too high concentration: 0.1	True / False
Explanation Symptoms (Measurable):	**Sensor Status (from Diagnosis Agent)**
Pressure Difference: High	True / False
Power consumption: Normal	True / False
Concentration in accept: Normal	True / False
Concentration in reject: High	True / False
Accept flow: Low	True / False
Reject flow: Normal	True / False
Failure Effects:	**RCA Update on Effects:**
Screen plate clogged	True / False

Fig. 6. GUI-functionality for presentation of RCA-results and collection of user feedback

In any real process application, RCA needs adaptation to incorporate the ongoing changes in process behavior. A suitable adaptation algorithm is the sequential learning with fading [4]. The fading is a convenient feature after maintenance activity on the plant. The sequential learning is performed on the actual root cause nodes and corresponding evidence for that particularly observed case. This is based on feedback from DCS and on operator/maintenance reports, see Fig. 6.

The object oriented BN framework fits very naturally into any industrial IT environment, which utilizes object oriented integration of applications as containers of different applications communicating via the aspect integration platform in order to allow overall process optimization.

4. Conclusions

The outlined application has been a subject for feasibility study of our methodology in pulp process operation. It could be generally applicable in process industries. A methodology for systematic (and automated) system testing still needs to be developed, in order to ensure efficient system development, especially when applications grow in size.

The experience shows, that simple updates of typical repetitive structures (e.g. sensors) in a BN may turn into annoying and time-consuming task. Instead, we have used OOBN with advantage for RCA & DS in process operation, i.e. OOBN ensure causal modeling of interdependency of events, simplifies modification and reusability of BN.

The use of OOBN has simplified the development of the model for adaptive signal classification and prediction of the development of signals' level-trend. Moreover, the overall RCA-model complexity has been reduced at different levels of industrial plant hierarchy. This provides an overview for explanations of RCA-conclusions. In addition, this OOBN are used in a next level OOBN for risk assessment of disturbances and estimation of their most probable root causes for predictive maintenance on demand.

The proposed system design for Root Cause Analysis (RCA) & Decision Support (DS) is incorporating agents for handling of uncertainties. We have developed one particular agent, based on fluid dynamics modeling. This is solving an equation system for the pressure-flow network. It provides soft sensor (non-measurable) information for evidence in the reasoning under uncertainties.

This allows us to estimate the risk of abnormality at an early stage and to propose early treatment by an efficient sequence of actions. The last is utilizing cost estimations anticipating also potential production losses. This allows efficient troubleshooting and predictive maintenance on demand [19].

This application demonstrates that fast and flexible disturbance analysis (RCA and DS) is feasible in industrial process control. It need not be a time-consuming task, if a computerized troubleshooting system is deployed. Thus, it can reduce substantially the production losses due to unplanned process breakdowns.

Acknowledgments

We would like to thank Per-Olov Gelin, Juan Jaliff and Bengt Oldberg for the project management at ABB, all ABB colleagues for sharing with us their knowledge and experience on process engineering and operation, as well as for the stimulating discussions on application aspects and integration into the Industrial IT environment.

References

1. Pearl J. (1988): Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann Publishers
2. Cowell R.G., Dawid A.P., Lauritzen S.L., Spiegelhalter D.G. (1999): Probabilistic Networks and Expert Systems, Springer Verlag
3. Jensen F.V. (2001): Bayesian Networks and Decision Graphs, Springer Verlag, New York
4. Spiegelhalter D., Lauritzen S.(1990): "Sequential updating of conditional probabilities on directed graphical structures", Networks 20: 579-605
5. Kalagnanam J., Henrion M. (1990): "A Comparison of decision analysis and expert rules for sequential analysis". In P.Besnard S.Hanks, editors, Uncertainty in Artificial Intelligence 4: 271-281. North-Holland, New York
6. Heckerman D., Breese J., Rommelse K (1995): "Decision-theoretic Troubleshooting" Communications of the ACM, 38(3): 49-56
7. Nauck D., Klawonn F., Kruse R. (1997): Foundations of Neuro-Fuzzy Systems, John Wiley & Sons Ltd.
8. Koller D. and Pfeffer A (1997), Object-oriented Bayesian networks, In Proc. UAI-97, pp. 302-313, Morgan Kaufmann
9. Weidl G., Dahlquist E. (2002) Root Cause Analysis for Pulp and Paper Applications, In proceedings of 10th Control Systems conference, pp.343-347, Stockholm; Sweden, June 3-5, 2002
10. Weidl G. (2002) Root Cause Analysis and Decision Support on Process Operation. Doctoral Dissertation ISBN 91-88834-36-0. Malardalen University, Sweden
11. Weidl, G., Madsen, A.L. and Dahlquist, E. (2002). "Condition Monitoring, Root Cause Analysis and Decision Support on Urgency of Actions". In Book Series FAIA, vol.87, “Soft Computing Systems - Design, Management and Applications”, pp. 221-230. IOS Press, Amsterdam, the Netherlands
12. Weidl G., Vollmar G., Dahlquist E. (2003) Adaptive Root Cause Analysis under uncertainties in industrial process operation, in conference proceedings of FOCAPO (Foundations Of Computer Aided Process Operations), Florida, January 12-15, 2003, USA
13. Andersen, S.K., Olesen, K.G., Jensen, F.V., and Jensen, F. (1990) HUGIN - a shell for building belief universes for expert systems. In Readings in Uncertainty; Shafer, G., and Pearl, J. (eds.); Morgan Kaufmann; 332-337.
14. Jensen, Kjærulff, Lang, Madsen 2002, HUGIN - The Tool for Bayesian Networks and Influence Diagrams,"First European Workshop on Probabilistic Graphical Models",p.212-221
15. McLachlan G.J. and Basford K.E. (1988): Mixture Models: Inference and Applications to Clustering. Marcel Dekker, New York
16. Holst A. (1997): The use of Bayesian Neural Network Model for Classification Tasks, PhD Thesis, dept of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm, Sweden.
17. Kjaerulff U. (1992), A computational scheme for reasoning in dynamic probabilistic networks, In Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence, 121-129, Morgan Kaufmann, San Francisco.
18. C.-H.Wang, S.H.Sheu (2003). Determining the optimal production-maintenance policy with inspection errors: using Markov chain, Computers and Operations Research, vol. 30, pp. 1-17
19. Weidl, G., Madsen, A.L. and Dahlquist, E. (2003). In proceedings of the Bayesian Workshop at the 19th Conference on Uncertainties in Artificial Intelligence, August 7-10, 2003, Acapulco, Mexico
20. Fritzson, P. and Engelson, V., (1998) Modelica - A Unified Object-Oriented Language for System Modeling and Simulation. In Proceedings of European Conference on Object-Oriented Programming ECOOP98, Brussels, Belgium, July 20-24, 1998