



## Generating Diagnostic Bayesian Networks from Qualitative Causal Models

Kirchhübel, Denis; Jørgensen, Thomas Martini

*Published in:*

Proceedings of the 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation

*Publication date:*

2019

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Kirchhübel, D., & Jørgensen, T. M. (2019). Generating Diagnostic Bayesian Networks from Qualitative Causal Models. In *Proceedings of the 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation* IEEE.

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# **Proceedings**

## **2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)**

Paraninfo Building, University of Zaragoza  
Zaragoza, Spain  
10 - 13 September, 2019

Organized by

University of Zaragoza, Spain

Sponsored by

The Institute of Electrical and Electronics Engineers (IEEE)  
IEEE Industrial Electronics Society (IES)  
Aragón Institute for Engineering Research (I3A), Spain

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Operations Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved. Copyright ©2019 by IEEE.

IEEE Catalog Number: CFP19ETF-ART

ISBN: 978-1-7281-0303-7

ISSN: 1946-0759

# Generating Diagnostic Bayesian Networks from Qualitative Causal Models

Denis Kirchhübel  
Department of Electrical Engineering  
Technical University of Denmark  
Kgs. Lyngby, Denmark  
dekir@elektro.dtu.dk

Thomas Martini Jørgensen  
Department of Applied Mathematics  
and Computer Science  
Technical University of Denmark  
Kgs. Lyngby, Denmark  
tmjq@dtu.dk

**Abstract**—Safety and efficiency of modern industrial plants can be improved by providing operators with effective digital assistants to diagnose abnormal situations occurring in the plant. To make sense of a large number of alarms, root cause analysis can help pinpoint the origin of an abnormal situation. We investigate the translation of qualitative causal models into Bayesian belief networks (BBN) to utilize efficient tools for probability inference. The diagnosis result of a fault scenario of the Tennessee-Eastman-Process highlight the feasibility of the principle approach and the ongoing research aims to fully leverage the potential of BBN.

**Keywords**—*Bayesian methods, Expert systems, Process Control, Fault diagnosis, Fault-trees*

## I. INTRODUCTION

The impact of situational awareness in the decision-making process of operators of industrial processes has long been a concern in safety critical operations. In order to address the challenge of human factors the performance of the operator interface naturally is essential. Situational awareness is necessary to maintain reliability, anticipate events and respond appropriately when or before they occur. Comprehension of the situation is based on a synthesis of perceived information. This comprehension requires an understanding of the significance of the presented elements beyond simply being aware of information. By developing digital assistants to support the operator, the operators can make better-informed decisions.

Reducing the number of alarms presented to an operator, by means of alarm management has been the subject of many improvement efforts in industry. This reflects in the guidelines and standards defined for process industries, most significantly the Engineering Equipment Materials Users' Association (EEMUA) publication 191 [1]. Hollifield et al. [2] have compiled an overview of the current best-practice in industry. It is necessary to combine all available information to provide accurate assistance: recorded data, as well as design and operation knowledge [3].

Digital diagnostic assistants for “human in the loop” real-time operations can be valuable in identifying root causes of failures [4], [5] or even predict the onset of disturbances that could lead to failures [6], [7], thereby reducing downtime and limiting stress for the operators. To establish a model of the system, technical documentation like P&IDs and process flow diagrams, are important [3], but it is also vital to “harvest” expertise and experience from engineers and operators [8] or empirical data [5]. Based on this data the causality between

offsets in the plant can be modeled and the fault propagation can be analyzed to identify root causes [7], [9].

Bayesian Belief Networks (BBN) have been proposed as means of diagnosing faults based causal process representations. BBN yields a way of representing uncertainty about the causal relationships and with efficient Bayesian inference one can update likelihood estimates of the unobserved states and the potential root causes, implying the possibility of ranking them. One approach to establish the process representation on which Bayesian Network analysis can be added is to apply Hazard and Operability Studies (Hazop) [10] or fault-tree analysis [11], [12], which are usually prepared for risk assessment. Peng et al. [13] also describe the application of Bayesian inference to distinguish root causes identified by fault propagation in a multi-logic causal model. However, producing and maintaining accurate representations of a process in safety documents in a consistent digital format is a lengthy process involving process, safety and operations experts.

Multilevel Flow Modeling (MFM) facilitates the generation of causal models. Representing the process as a hierarchy of mass, energy, and control flows by functional concepts provides a structured approach to modelling causality. [14] Using the abstract causal model reduces the knowledge engineering effort and fault propagation is used to propose root cause candidates. To identify the actual root cause BBN can then be used, similar to the approach in [13].

This paper examines more closely the application of BBN for on-line fault diagnosis based on causal models. The approach is demonstrated on a case of the Tennessee Eastman Process simulator (TEP) [15]. The paper is organized as follows. First we present a causal model of the thermal aspects in the TEP based on the concept of Multilevel Flow Models (MFM) [14]. Next, we outline different approaches to generate a BBN from the causal model. Finally, we obtain the probabilities of all possible root causes and compare the results with the true fault for the investigated TE scenario.

## II. CAUSAL MODEL OF THE TENNESSEE EASTMAN PROCESS

The MFM causal model shown in Fig. 1 represents the thermal aspects of the TEP. The MFM methodology decomposes the system into mass and energy flows as a hierarchy of means and supported objectives of the plant. By using abstract function primitives to represent the different flows the model can be related to the design intentions and human understanding of the plant operation [7]. Some of the abstract

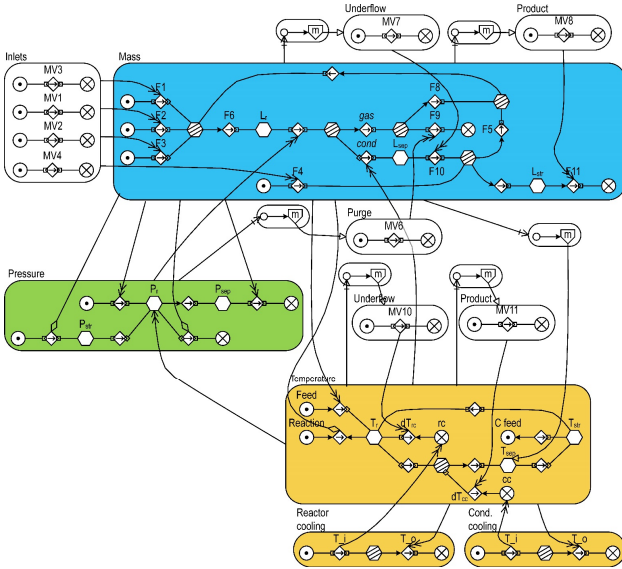


Fig. 2 Multilevel Flow Model of the TEP. Decomposition of the process by flow function primitives [14] for mass flow (blue), pressure (green), heat transfer (yellow), and control loops (white).

mass or energy flow functions in the model directly relate to measurable quantities in the process, such as pressures, temperatures, and flow rates, as well as manipulated variables. The implicit causal model in MFM yields a causal di-graph for the measurable quantities of the TEP shown in Fig. 2.

For the diagnosis, we consider the scenario of a high condenser coolant temperature, indicated by two alarms: F9 high and F5 low. The alarms are generated by a 2% band around the steady state value [16] using the simulation data provided by Rikker [17].

### III. BAYESIAN BELIEF NETS

A Bayesian belief network is a probabilistic graphical model that describes variables and their conditional dependencies based on a directed acyclic graph. Efficient algorithms exist for both inference of probabilities and learning of the causal structure. As such, it represents an established methodology for analyzing complex causal dependencies between faults [18], [19]. There are a number of synonyms in the literature all corresponding to a Bayesian Belief Net (BBN). These include Bayes nets, directed acyclic graphs, and probabilistic networks.

A BBN models the joint probability distribution of the combined states of the system under consideration - in our case given by the nodes in the (combined) fault-tree(s). The BBN is defined by a directed acyclic graph in which each edge corresponds to a conditional dependency and each node corresponds to a unique random variable. In addition to the graph structure, the BBN contains Conditional Probability Tables (CPT) for all nodes having one or more parents and by marginal probability tables for the root nodes (nodes without parents). Generally, one can say that a BBN is a solution to model complex systems because they perform the factorization of the variables joint distribution based on the conditional dependencies. The main objective of BBNs is to compute the distribution probabilities of a set of unknown variables given the observation of one or more other variables. The detailed principles of this modeling tool are explained in [18], [19].

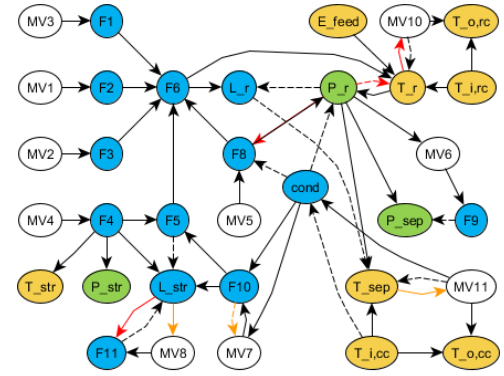


Fig. 1 Signed directed graph generated from MFM model. Positive (solid) and negative (dashed) edges are directed from cause to effect. Minor reciprocal influence (red) and influence of controlled variables (yellow) removed to generate BBN. Nodes are coloured according to MFM perspectives.

Building a BBN involves both a structural part (the graph) and a quantitative part (the probability tables). Both of these parts can be learned from data. However, developing a BBN for a complex system entirely by learning from historical diagnostic cases, although very attractive, is rarely an option due to lack of data. On the other hand updating an existing model from data is often feasible. It is also important to note that the structural part (the backbone of the causal dependencies) is more difficult to learn than the parameter values of the probability tables. Accordingly, it is advantageous to draw structural information from an available causal representation of the system. There have been a number of publications showing how to map a fault-tree to a BBN [11], [12]. It is in principle straightforward and the OR gates can be directly represented by deterministic CPTs. In addition we can then easily represent uncertainty in the anticipated propagation of causes to consequences by modifying the CPTs using so-called Noisy OR gates [20].

### IV. GENERATING BAYESIAN NETS

In the following, we outline three different paths of generating a BBN from the presented causal model. Firstly, a recipe for removing cycles in the causal di-graph is presented, generating a BBN with trinary states for all variables in the graph. Subsequently, two different ways of interpreting the fault-trees generated by back tracing into a BBN.

#### A. Trinary representation

If it is possible to translate a causal graph directly into a BBN, each variable can be considered as having one of three states: no deviation (normal), too low value (low), or too high value (high). However, control loops in the system will reflect as cycles in a causal graph, whereas a BBN needs to be acyclic by definition. We propose a recipe to resolve these cycles and create a BBN from a causal di-graph, or signed di-graph. The approach is based on the following assumptions:

- During normal operation, reciprocal influence between two variables has a dominant direction, due to the process and control design.
- Actuated variables are likely root causes, e.g. a fault in the sensor, controller implementation, or the actuator can cause an offset.

TABLE 1 CPT FOR INTERMEDIATED NODE  $v_i$  BASED ON PARENT  $v_j$ 

$v_i \backslash v_j$	$s_{ji} = I$		
	high	normal	low
high	1	0	0
normal	0	1	0
low	0	0	1

---

$v_i \backslash v_j$	$s_{ji} = -I$		
	high	normal	low
high	0	0	1
normal	0	1	0
low	1	0	0

Consequently, the following steps are performed to generate a BBN:

1. Identify and keep only dominant influence between reciprocal variables. See Fig. 2.
2. Remove edges from controlled variable to corresponding actuated variable, if they are part of a cycle. See Fig. 2.
3. Generate a CPT for each intermediate node using Table 1. Each entering edge  $e_{ji}$  of node  $v_i$  has sign  $s_{ji}$ . If  $v_i$  has multiple entering edges the cross product of the states is formed as the sum of the high or low observations, the normal state is only probable if all parent nodes indicate normal.

#### B. Mapping causal paths

The assumptions presented in A will often be too simplistic to cover all scenarios in a plant, e.g. if a cycle led to remove the influence of  $P_r$  on  $MV6$  by rule A.2 there would be no propagation causing  $F9$  high. Alternatively, we explore a situation specific approach to generating the BBN. Given specific fault observations, the causal model can be used to trace back the causal path and identify possible root causes. In doing so, only a subset of the graph will be traversed and occurring cycles are detected. The back-tracing spans a fault-tree but common consequences linking branches of the fault-tree are being ignored. To penalize longer propagations uncertainty is introduced for each traversed edge in the causal model, assuming a higher likelihood of local causes rather than long propagation paths in the system.

Independent back tracing of multiple faults will lead to individual fault-trees. To reach a meaningful diagnosis nodes representing the same variable cannot be considered independently per fault-tree or causal path, as they refer to the same physical entity.

TABLE 2 CPT FOR BINARY ROOT CAUSE BASED ON TRINARY PARENT NODE

$v_i \backslash v_j$	trinary		
	high	normal	low
high			
fault occurred	1	1	0
normal	0	1	1

---

$v_i \backslash v_j$	trinary		
	high	normal	low
low			
fault occurred	0	1	1
normal	1	1	0

Two ways of combining the generated fault-trees are considered here: (B1) combining all recurring nodes or (B2) maintaining split trees unless the causal paths overlap. The former approach requires re-examining nodes previously detected to form cycles as the combination of independent fault-trees can recreate those cycles. The latter approach avoids these cycles by only combining identical causal paths. In this way, we ensure to have single nodes for root causes but we may have multiple copies of intermediate nodes.

The nodes of the mapped BBN each represent a specific deviation with two states – “fault occurred” or “ok”. To incorporate the uncertainty introduced by the propagation Noisy-OR gates define the conditional probability of “fault occurred” if a given parent node also has a “fault occurred” state but with the respective uncertainty depending on the length of the causal path between parent and intermediate node.

#### C. Connecting root causes

While the complete BBN representation of the causal model with trinary nodes correctly interprets states that are mutually exclusive, a BBN generated by combining fault-trees can contain nodes referring to the same variable in mutually exclusive states without representing their relation. As outlined by Lampis [21] a common n-ary parent node incorporating all fault states and “normal” can link mutually exclusive root nodes.

Extending method B with the notion introduced in A, a trinary node with the states “high”, “normal”, and “low” is added as parent to the root nodes in the BBN created from independent fault-trees. Since the intermediate nodes are not coherently linked in the same manner, the uncertainty according to the propagation length is maintained for the intermediate nodes. Table 2 represents the CPT for a binary root cause depending on the trinary parent node.

TABLE 3 ROOT CAUSE PROBABILITIES (HIGH / LOW IN %) FOR TEP SCENARIO HIGH CONDENSER COOLANT TEMPERATURE  $T_{i,cc}$ 

Root cause	(A) Trinary	(B1) Combined	(B2) Split	(C1) Combined & Trinary
$T_{i,cc}$	56.71 / 14.08	53.37 / 50.04	54.80 / 50.07	35.55 / 31.12
MV11	14.08 / 56.71	---	---	---
MV4	17.96 / 46.11	50.04 / 56.18	50.16 / 64.09	29.24 / 37.43
MV5	37.88 / 26.82	50.05 / 50.13	51.24 / 50.33	33.30 / 33.37
$E_{feed}$	36.63 / 29.81	50.07 / 50.04	50.78 / 50.22	33.34 / 33.32
$T_{i,rc}$	29.81 / 36.63	50.04 / 50.06	50.22 / 50.78	33.33 / 33.34
MV1/ MV2/ MV3	34.62 / 31.87	50.07 / 50.04	50.70 / 50.11	33.35 / 33.32

## V. DIAGNOSIS RESULTS

Table 3 shows the probabilities inferred by the respective BBNs with the CPTs described before and no prior knowledge about the marginal distributions (“stupid prior”). Hugin Researcher [22] was used to design and diagnose the BBNs. In case of the trinary representations the marginal probability for “high” and “low” is 33.33 %. In the binary cases either of the independent root nodes for “high” and “low” can show “fault occurred” with 50%. Accordingly, the deviation from 33.33 % or 50 %, respectively, after Bayesian inference, indicates the likelihood of the fault being the root cause.

From Table 3 we observe that all BBNs put the actual root cause as first or second highest probability. However, the significant shortcoming of the fault-tree based BBNs (B and C) is reflected in the consistently higher probability of MV4\_low as root cause, since it could immediately cause F5 low. However, the missing link between the mutually exclusive states prevents the interpretation of this contribution. On the other hand, the presented trinary BBN does not contain the same notion of uncertainty for longer propagations represented by the Noisy-OR gates, since the noisy combination of more than binary states is not trivial, and at the same time some causalities had to be removed from the model to create a valid trinary BBN.

The diagnosis achieved by the two approaches of binary BBN yield the same ranking of causes. The introduction of trinary root causes for the binary model does not affect the ranking of the root causes, but allows a clearer distinction between the fault states of a single variable. It is noted that the required post-processing of the split paths can be significantly smaller, if the overlap of consistent causal paths is already considered during the back tracing in the causal model.

## VI. CONCLUSION

In summary, the presented investigation reveals a great potential in using BBN to interpret the root cause analysis based on causal models and shortlist the most relevant root causes to support operators. However, generating a general diagnostic BBN from a causal model is limited by the acyclic nature of BBNs. On the other hand, back tracing multiple observations and combining their fault-trees into a BBN disregards causal connections that could improve the diagnostic power of the BBN.

The combination of the comprehensive trinary and simple binary representation of quantitative fault states improves the diagnosis of the binary network but cannot fully capture causality of a given situation. The ongoing work focuses on identifying an efficient method to map a given causal model into a BBN maintaining as much as possible of the causal structure in a given situation while leveraging the possibilities of a trinary representation of process variables.

## ACKNOWLEDGMENT

The authors thank Hugin Expert A/S for providing the trial license of Hugin Researcher used to generate the results in this work. This work is funded by the Danish Hydrocarbon Research and Technology Centre (DHRTC) as part of the Operations and Maintenance Technology programme.

## REFERENCES

- [1] Engineering Equipment Materials Users' Association, “Alarm systems – A guide to design, management and procurement,” London, EEMUA Publication 191, 2013.
- [2] B. R. Hollifield and E. Habibi, *The alarm management handbook: a comprehensive guide*. PAS, 2006.
- [3] J. Wang, F. Yang, T. Chen, and S. L. Shah, “An Overview of Industrial Alarm Systems: Main Causes for Alarm Overloading, Research Status, and Open Problems,” *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 1045–1061, Apr. 2016.
- [4] S. Natarajan and R. Srinivasan, “Implementation of multi agents based system for process supervision in large-scale chemical plants,” *Comput. Chem. Eng.*, vol. 60, pp. 182–196, 2014.
- [5] D. T. Nguyen, Q. B. Duong, E. Zamai, and M. K. Shahzad, “Fault diagnosis for the complex manufacturing system,” *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, vol. 230, no. 2, pp. 178–194, Apr. 2016.
- [6] J. Zhu, C. Wang, C. Li, X. Gao, and J. Zhao, “Dynamic alarm prediction for critical alarms using a probabilistic model,” *Chinese J. Chem. Eng.*, vol. 24, no. 7, pp. 881–885, 2016.
- [7] X. Zhang, “Assessing Operational Situations,” Technical University of Denmark, 2015.
- [8] B. Cai, H. Liu, and M. Xie, “A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks,” *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, 2016.
- [9] E. Arroyo Esquivel, “Capturing and Exploiting Plant Topology and Process Information as a Basis to Support Engineering and Operational Activities in Process Plants,” Helmut-Schmidt-Universität, Hamburg, 2017.
- [10] J. Hu, L. Zhang, Z. Cai, and Y. Wang, “An intelligent fault diagnosis system for process plant using a functional HAZOP and DBN integrated methodology,” *Eng. Appl. Artif. Intell.*, vol. 45, pp. 119–135, 2015.
- [11] K. W. Przytula and R. Milford, “An Efficient Framework for the Conversion of Fault Trees to Diagnostic Bayesian Network Models,” in *2006 IEEE Aerospace Conference*, pp. 1–14.
- [12] A. Bobbio, L. Portinale, M. Minichino, and E. Ciancamerla, “Comparing Fault Trees and Bayesian Networks for Dependability Analysis,” Springer, Berlin, Heidelberg, 1999, pp. 310–322.
- [13] D. Peng, Z. Geng, and Q. Zhu, “A Multilogic Probabilistic Signed Directed Graph Fault Diagnosis Approach Based on Bayesian Inference,” *Ind. Eng. Chem. Res.*, vol. 53, no. 23, pp. 9792–9804, Jun. 2014.
- [14] M. Lind, “An overview of multilevel flow modeling,” *Nucl. Saf. Simul.*, vol. 4, no. 3, pp. 186–191, Sep. 2013.
- [15] J. J. Downs and E. F. Vogel, “A plant-wide industrial process control problem,” *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.
- [16] X. Ma and D. Li, “A Hybrid Fault Diagnosis Method Based on Fuzzy Signed Directed Graph and Neighborhood Rough Set,” in *2017 IEEE 6th Data Driven Control and Learning Systems Conference*, 2017, pp. 253–259.
- [17] N. L. Ricker, “Tennessee Eastman Challenge Archive,” *University of Washington*, 2019. [Online]. Available: [depts.washington.edu/control/LARRY/TE/download.html](https://depts.washington.edu/control/LARRY/TE/download.html). [Accessed: 03-Apr-2019].
- [18] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.
- [19] F. V. Jensen, *An introduction to Bayesian networks*. UCL Press, 1996.
- [20] A. Antonucci, “The Imprecise Noisy-OR Gate,” *Inf. Fusion*, pp. 709–715, 2011.
- [21] M. Lampis, “Application of Bayesian Belief Networks to System Fault Diagnostics,” Loughborough University, 2010.
- [22] U. B. Kjærulff and A. L. Madsen, *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*, vol. 22. New York, NY: Springer New York, 2013.