Causal Structure Learning in Process Engineering Using Bayes Nets and Soft Interventions

Christian Kuehnert, Thomas Bernard and Christian Frey
Fraunhofer IOSB
Email: {christian.kuehnert, thomas.bernard, christian.frey}@iosb.fraunhofer.de

***Abstract*—As modern industrial processes become more and more complex, machine learning is increasingly used to gain additional knowledge about the process from production data. Up to now these methods are usually based on correlations between process parameters and hence can not describe cause-effect relationships. To overcome this problem we propose a hybrid (constraint and scoring-based) structure learning method based on a Bayesian Network to detect the causal structure out of its data. Starting with an empty graph, first the underlying undirected graph is learned and edges coming from root nodes are directed using a constraint-based approach. Next a scoring-based method is used in order to calculate the uncertainty for each possible directed edge and out of this construct the Bayesian Network. Finally soft interventions are applied to the process in order to learn the real causal structure. Our approach is tested in simulations on a chemical stirred tank reactor and on an experimental laboratory plant.**

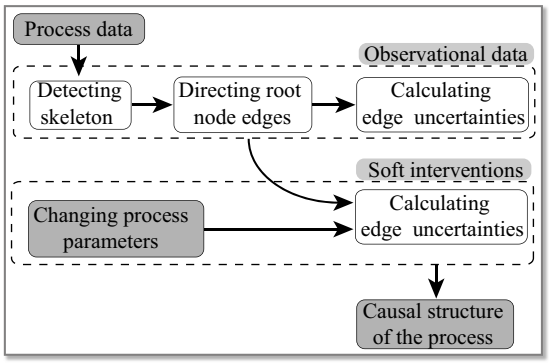

Fig. 1. Developed method to learn the causal structure using observational and interventional data in process engineering.

I. Introduction

Generally modern processes are complex systems in which not all relationships are known (i.e. partial reactions in chemical processes). To detect important and yet unknown relations between different process parameters, recently machine learning algorithms attracted more attention ([20],[6],[9]).
However up to now these methods are usually based on correlations and hence can not describe the effect and the cause between the found different process parameters. That's why without having a thorough expert background it is not possible to conclude how the application of the found knowledge (i.e. in form of changing process control) will have effect on the process.
One approach to overcome this problem and to detect causal structures in process data is to use Bayesian networks (BNs) [14] that are compact graphical representations of joint probability distributions. Recently these models have aroused much attention especially in the area of system biology [3],[10] and cognitive science [5]. Still in process engineering BN are only used for fault diagnosis [7] and monitoring [19] tasks but not for the detection of cause-effect relationships in data. For this reason we will examine the use of BN to detect unknown causal associations in measurement data coming from process factors. In addition, we will use interventional data ([12]) to distinguish between possible causal structures found in the observational data.
Regarding Bayesian Networks there are essentially two kinds of approaches for learning the structure. The first one tries to find a graph which satisfies all the statistical constraints implied by the empirical conditional independences in the data. The idea of this approach is exploited in [15] in form of the PC-algorithm and theoretically leads to a subset of directed acyclic graphs that all share the same set of probabilistic dependency relations. Still one special problem of constraint-based algorithms is that they face problems when the sample size is low [1]. Furthermore, due to repeating the same statistical tests, errors propagate through the network and can lead to bad results [17].
The second approach uses some scoring metric to find a graph which returns the model with the highest scoring. This approach is usually more robust than performing a constraint-based method but does not imply that the most probable model also describes the underlying causal structure.
Furthermore, as explained in [14] both approaches suffer from the fundamental problem that they can only identify the model up to its Markov equivalence. This means that both methods can only find models that imply the same set of (conditional) independences implied in the data.
The only way to distinguish between Markov equivalent models is to perform interventions. An intervention means that some manipulation has been performed from the outside on some part of the process. Interventions can be modeled in different ways (see [2] for an overview). Traditionally they are modeled by clamping a variable to a fixed value (i.e. fixing a temperature or a valve in a process). Unfortunately, usually this can not be applied to industrial processes as this would in most cases harm the installed plant.
For this reason we propose to use *soft interventions* as described in [12] to distinguish between the found Markov

equivalent models. The advantage of this modeling is that depending on a user selected pushing factor, the force of the impact of the intervention can be described in the BN. Namely small changes as for example changing the set point are described as *soft pushing* and interventions where a parameter is fixed to one value are described as *hard pushing*.
To learn the underlying causal structure of a process we propose the following method outlined in figure 1. Starting with an empty graph, first the skeleton (the underlying undirected graph) is learned and edges coming from root nodes are directed using a constraint-based approach. Using this approach means that the found partially directed graph is in terms of causality correct. Next a scoring-based method as presented in [3] is used. This method calculates the uncertainty for each possible directed edge and out of this constructs the BN. Calculating the uncertainties in the graph tells if it can be trusted in the the found BN or not. Finally we perform soft interventions to learn the real causal structure.
The performance is tested on a chemical stirred tank reactor and on an experimental plant where liquid is pumped around in cycles.
The paper is structured as follows: In section II BNs are reviewed and learning through soft interventions is explained. Section III explains the developed hybrid approach for detecting causal structures in observational data. Finally section IV shows the obtained results using our approach with and without using interventional data on a .

II. Learning Bayesian Networks

Formally a BN consists of $X_i, X_j, X_k, ...$ nodes that represent in our case process parameters and are linked together using directed or undirected edges.
Assuming that all nodes X are available and a priori globally independent, a BN can be computed using the product of the marginal likelihod of each family (a node and its parents) in the net.

$$P(X^{1:N}|G) = \prod_{i=1}^{d} p(x_i^{1:N}|x_{G_i}^{1:N}) \tag{1}$$

In here x_i^n describes the value of the node i and n describes the possible states of the considered family, denoted by G_i.
For a family the marginal likelihood can be calculated using the following equation including the model parameter θ_i^0.

$$p(x_i^{1:N}|x_{G_i}^{1:N}, F_i^{1:N}) = \int [\prod_{n;F_i^n=0} p(x_i^n|x_{G_i}, \theta_i^0)] p(\theta_i^0) d\theta_i^0 \tag{2}$$

For a family the parameter priors have to satisfy local independences as well. For this the BDeu (Bayesian Dirichlet equivalence uniform) prior is selected which is described in [8] by using the following equation:

$$p(x_i^{1:N}|x_{G_i}^{1:N}) = \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})} \prod_{k=1}^{q_i} \frac{\Gamma(\alpha_{ij}+N_{ij})}{\Gamma(\alpha_{ij})} \tag{3}$$

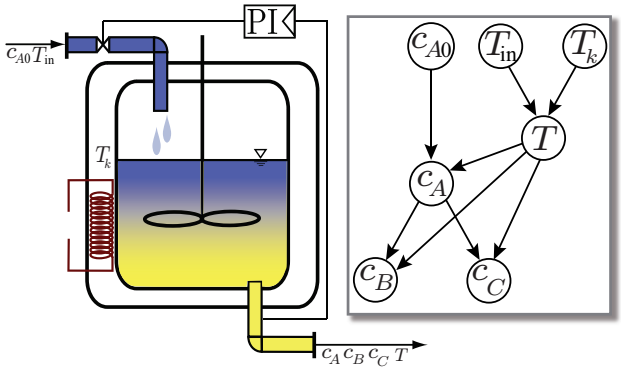

Fig. 2. Illustration of the chemical stirred tank reactor and the underlying causal structure derived from the differential equations.

While

$$\alpha_{ij} = \frac{1}{q_i r_i} \tag{4}$$

Hereby r_i describes the number of states in X_i, q_i the number of possible states of the graph X_{G_i} and Γ describes the Gamma function.
It is assumed that the available data is not generated by experiments but available as usual production data including some statistical significant variation for input and output data and that the number of selected process parameters is sufficient to detect all causal dependencies. In addition it is assumed that the causal structure stays unvaried if interventions are performed on the process [14].

A. Learning through soft interventions

Hard pushing is usually not suitable for industrial processes as this would mean that some parameters would have had to be clamped to exactly one value and this would in most cases harm the process. For that reason soft interventions introduced by [12] are used as hereby variables do not have to be fixed to one value.
Soft interventions mean that an intervention just increases the likelihod that a node enters a target state x_i^* depending on a pushing force ω_i between $[0...\infty]$. If ω_i is chosen to be 0 there has been some small mechanism change, without any target state, for $\omega_i \to \infty$ an intervention has been performed where the process variable is effectively cut off from its parents.
In the mathematical framework of a BN an intervention is modeled in form of so called intervention nodes F_i^n (for $F =$ force). Is $F_i^n = 1$, this means there has been an intervention on node i and the model parameters θ_i^1 are used. Is $F_i^n = 0$ the model parameters without intervention θ_i^0 are selected.
For this equation 2 has to be extended to consider the interventions in the marginal likelihood. Now the likelihood is separated into two categories, namely in a part where the nodes are observed and a part where there has been an intervention. In other words θ_i^0 and θ_i^1 need to have different

hyper parameters.

$$p(x_i^{1:N}|x_{G_i}^{1:N}, F_i^{1:N}) = \int[\prod_{n;F_i^n=0} p(x_i^n|x_{G_i}, \theta_i^0)]p(\theta_i^0)d\theta_i^0 \times \int[\prod_{n;F_i^n=1} p(x_i^n|x_{G_i}, \theta_i^1)]p(\theta_i^1)d\theta_i^1 \quad (5)$$

This leads to the following description using the BDeu prior:

$$p(x_i^{1:N}|x_{G_i}^{1:N}) = \prod_{j;F_i^n=0} \frac{\Gamma(\alpha_{ij}^0)}{\Gamma(\alpha_{ij}^0+N_{ij}^0)} \prod_{k;F_i^n=0} \frac{\Gamma(\alpha_{ij}^0+N_{ij}^0)}{\Gamma(\alpha_{ij}^0)} \times \prod_{j;F_i^n=1} \frac{\Gamma(\alpha_{ij}^1)}{\Gamma(\alpha_{ij}^1+N_{ij}^1)} \prod_{k;F_i^n=1} \frac{\Gamma(\alpha_{ij}^1+N_{ij}^1)}{\Gamma(\alpha_{ij}^1)} \quad (6)$$

Equation 6 shows that the pushing of the intervention solely depends on the selection of the parameter value α_{ij}^1. This can be modeled by assuming that there is a deterministic relationship between α_{ij}^0 and α_{ij}^1 described over a pushing factor ω_i as follows:

$$\alpha_{ij}^1 = \alpha_{ij}^0 + \omega_i e_t \quad (7)$$

In this equation $t = x_i^\star$ is the target value and the vector $e_t = (0, ..., 0, 1, 0, ..., 0)$ whereas the value of one is describing the direction in which the BDeu prior is pushed in direction of the target value.

III. Hybrid structure learning approach

To learn the causal structure our developed method starts using the MMPC algorithm [17] to find the underlying skeleton (graph without directed edges) in the data. Out of the found skeleton the root nodes are detected by looking for collisions as this step has the advantage that assuming faithfulness [14] in the data the found structure is in terms of causality correct. Out of the found partially directed acyclic graph the most possible BN is found by calculating the uncertainty of each edge in the BN. This has the advantage compared to other scoring-based methods that it can be seen directly if it can be trusted on the found BN or not (if uncertainty in the BN is high or low).

A. Step 1: Detecting skeleton

This step reduces the number of possibles graphs that can possibly be constructed out of the available measurement data using the scoring-based approach described in section III-C.
For calculation of the skeleton the MMPC (max-min parents and children) algorithm is used which is separated in two phases summarized here for completeness. For a more detailed explanation, we refer to [17].
In the growing (first) phase, nodes that are conditionally dependent on a node X_i are selected as a possible set of parents and children according to a heuristic function (max-min heuristic). Iteratively a node X_i enters the candidate set if it maximizes the minimum association (defined as the minimal conditional dependency of a node X_i and X_j) to X_j given the current candidate set. This means that each time the node is selected which is most unlikely to be conditionally independent from the subset of the current candidate set. The phase is finished after all dependent variables have entered the candidate set.
In the shrinking (second) phase, the false-positive variables that entered the candidate set during the growing phase are removed. False positives are variables that are independent of X_i given some subset of all variables.

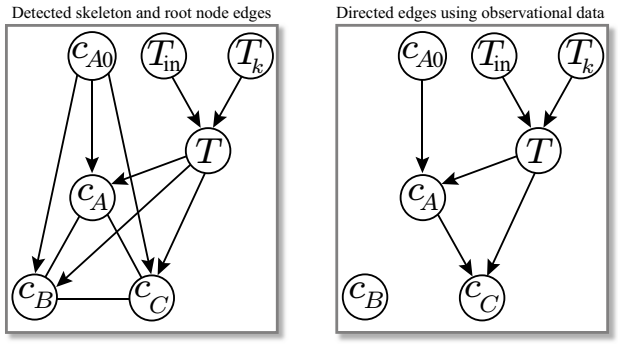

Fig. 3. Calculation of the causal structure after 1000 productions using only observational data. The left plot shows the found skeleton and the collisions coming from root nodes, the right plot shows the selected edges after calculating the edge uncertainties.

B. Step 2: Directing root node edges

After finding the skeleton the root nodes of the graph are detected. This step reduces the possible BNs that can be constructed by calculating the edge uncertainties further.
To find these nodes the detection of collisions in the skeleton is performed. The idea behind collisions is shown in Table I with the representation of the conditional independences. Finding that the nodes X_1 and X_3 are independent conditional on X_2 is consistent with a chain and a fork. This means that these structures can not be detected uniquely but collisions can be used to direct edges in the skeleton and imply a unique representation concerning the conditional independence.
In industrial processes root nodes represent input parameters. That means finding these nodes reduces significantly the number of possible graphs that can be detected when calculating the edge uncertainties.

TABLE I
Possible representations of a DAG and its (conditional) independences.

DAG	Representation	(Conditional) Independences
Chain	$X_i \to X_k \to X_j$	$X_i \perp X_j \mid X_k$
Fork	$X_i \leftarrow X_k \to X_j$	$X_i \perp X_j \mid X_k$
Collision	$X_i \to X_k \leftarrow X_j$	$X_i \perp X_j$ but $X_i \not\perp X_j \mid X_k$

As [17] shows that constraint-based methods lose their statistical power as the test for conditional independence is repeatedly applied, only collisions that include at least one root node are used for describing the graph. The other edges are directed using the scoring based approach described in the following section.
To calculate the conditional independences implied in the data the G^2 statistic as described in [15] is used.

C. Step 3: Calculating edge uncertainties

In this step the resulting graph from step one is used to calculate the uncertainties of each edge. The clear advantage of this method is that it shows on which edges in the BN exist uncertainties and if it can be trusted on the found graph or not.
The weighting of the edges is performed by calculating the marginal likelihood function by using the equation as described in [3] :

$$P(f_{x_k \to x_j} | G_j, X^{1:N}) = \frac{P(f_{x_k \to x_j}, X^{1:N} | G_j)}{P(X^{1:N} | G_j)} \quad (8)$$

In this equation the nominator and the denominator can be calculated separately. The nominator is calculated through the sum of the individual marginal likelihood of each Graph described as follows:

$$P(f_{x_k \to x_j}, X^{1:N} | G_j) = \sum_{G_k \in G_j} P(G_k | G_j) P(X^{1:N} | G_k) \quad (9)$$

The denominator results in the sum of the marginal likelihood from each possible family for each node and a following multiplication over the number of nodes:

$$P(X^{1:N} | G_j) = \prod_{k}^{j} \sum_{G_k \in G_j} P(G_k | G_j) p(x_k^{1:N} | x_{G_k}^{1:N}) \quad (10)$$

The value for the direction of each edge $X_i \to X_j$ can differ from zero to one. In this case a value of zero means that the edge $X_i \to X_j$ does not exist, a value of one means that the direction of the edge exists for sure and 0.5 describes the maximum value of the uncertainty of the edge namely it would symbolize an undirected edge.

IV. Results

The proposed approach is used to learn the causal structure of a chemical stirred tank reactor and an experimental setup. For validation and visualization of the experimental results the L_1 edge error is calculated as described in [13]. This error is defined as follows:

$$\begin{aligned} L_1 = \; & \sum_{i=1}^{n} \sum_{j=i+1}^{n} I_{G^\star}(X_i \to X_j)(1 - P(X_i \leftarrow X_j) \\ & + I_{G^\star}(X_i \to X_j)(1 - P(X_i \leftarrow X_j) \\ & + I_{G^\star}(X_i \perp X_j)(1 - P(X_i \perp X_j) \end{aligned} \quad (11)$$

The parameter $I_{G^\star}(X_i \to X_j)$ is used as an indicator. If it has the value 1, this means that there is a directed edge $X_i \to X_j$ in the real graph. $G^\star$. As the L_1 edge error is calculated using the edge uncertainties, the error can have a non integral value.

A. Chemical stirred tank reactor

To validate the method data coming from a chemical continuous stirred tank reactor with jacket cooling is used for analysis. The reactor is fed with a solution of a reactant with the concentration c_{A0}. The reactant undergoes a chemical reaction following the so called van de Vusse scheme [18]. Thereby the desired main reaction with the concentrations

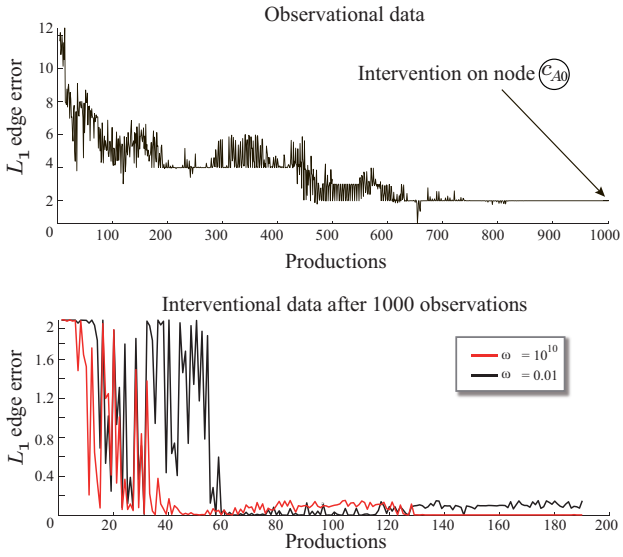

Fig. 4. $L1$ edge error of the chemical stirred tank reactor using only observational data and performing interventions. The lower plot shows two different types of intervention namely performing soft and hard pushing

$c_{A0} \to c_A$ is accompanied by two follow up reactions $c_A \to c_B$ and $c_A \to c_C$ that are both resulting in unwanted byproducts.
The control of the inflow $\dot{V}_{in}$ is used as a control to keep the concentration c_A on a constant value. The process is illustrated in figure 2. The relevant equations c_A, c_B and c_C of the process are described as follows:

$$\dot{c}_A = -k_1(T)c_A - k_3(T)c_A + [c_{A0} - c_A]\dot{V}_{in} \quad (12)$$
$$\dot{c}_B = k_1(T)c_A - k_2(T)c_B - c_B\dot{V}_{in} \quad (13)$$
$$\dot{c}_C = k_3(T)c_A - k_2(T)c_C - c_C\dot{V}_{in} \quad (14)$$

The kinetics $k_1(T)$, $k_2(T)$ and $k_3(T)$ are modeled using Arrhenius functions of the temperature and are calculated as follows:

$$k_i(T) = k_{i0} e^{(\frac{-E_i}{T_i + 273.15})} \quad (15)$$

The enthalpy balance regarding fluid temperature T and cooling jacket T_k leads to the following two differential equations.

$$\dot{T} = \alpha[T_k - T] + [T_{in} - T]\dot{V}_{in} \quad (16)$$
$$\dot{T}_k = \beta[T - T_k]\gamma\dot{Q} \quad (17)$$

Within here α and β are compound factors and describe the coefficient of the reactor- and fluid temperature. The parameter γ describes the heat conductive coefficient between heating power and the reactor temperature. As α is much smaller than β the edge in the graph is directed from $T_k \to T$.
The used values are shown in table II. The BN corresponding to the used differential equations is illustrated in figure 2 as well.

1) Generating a feature database: To capture the process dynamics that are required to detect the causal structure in the data the start-up phase of the reactor is used.
For this several start-up runs are performed while the initial

values of the concentrations c_A, c_B, c_C, the fluid temperature T_{Fl} and the reactor temperature T_k are varied from run to run. Furthermore, the mean value of the concentration c_{in} and the temperature T_{in} are varied and the inflow $\dot{V}_{in}$ is kept constant. As the start-up of the process is an instationary phase the generated data cannot be used directly and a feature extraction [11] has to be performed to describe each run. For this the mean value is calculated for each process value between the beginning and the end of the start-phase. By definition the start-phase is finished as the concentration c_A reaches its stationary value.

TABLE II
MODEL PARAMETERS OF CHEMICAL STIRRED TANK REACTOR.

$k_{10}, k_{20} = 1.28 \cdot 10^{12} h^{-1}$	$k_{30} = 9.04 \cdot 10^{9} h^{-1}$
$E_1, E_2 = 9758.3$	$E_3 = 8560$
$\alpha = 10.82 h^{-1}$	$\beta = 86.68 h^{-1}$
$\gamma = 0.1 \frac{K}{kJ}$	
$c_{A,SP} = 1.5 \frac{mol}{m^3}$	$T_{SP} = 115K$
$T_{k,SP} = 108K$	$\dot{V}_{SP} = 0.1 hr^{-1}$
$c_{A0} = 6 \frac{mol}{m^3}$	$T_{in} = 404K$

2) Results: In figure 3 the results after 1000 productions are shown. On the left side the skeleton including the found collisions is shown, on the right side the resulting BN after calculating the edge uncertainties. The previously wrong detect edges from $c_{A0} \rightarrow c_B$ and $c_{A0} \rightarrow c_C$ are removed after calculating the edge uncertainties. Furthermore, the two existing edges, namely the edges $c_A \rightarrow c_B$ and $T \rightarrow c_B$ were removed as well and declared as being unimportant.
In figure 4 the upper plot illustrates the $L1$ edge error against the used number of productions. Having only a small amount of productions shows a high uncertainty included in the edges of the BN as the error fluctuates a lot. Finally after 750 runs the $L1$ edge error converges towards an error of two.
To detect the real real causal structure and finally directing the last two edges of the graph interventional data has to be used. In detail an intervention will be performed on the parameter c_{A0}. For this the process parameter c_{A0} is once clamped to the value 6.5 which corresponds to $\omega \rightarrow \infty$ (hard pushing). For this ω is selected as 10^{10}. In the other case the parameter c_{A0} is fixed to the value 6.5 but still including a variance of $\sigma^2 = 0.5$ (soft pushing). For this case the value of ω is selected to be 0.01. For a thorough discussion on how to exactly set value of ω we refer to [12] as this is out of scope in this publication.
The results of the two interventions are shown in the lower plot of figure 4. This shows that both types of intervention find the right underlying causal structure of the stirred tank reactor. Performing a hard pushing results after 40 productions in the right causal structure, while performing only soft pushing takes 60 productions until the right causal structure has been found. This shows that small parameter changes are already sufficient to detect the real causal structure only that it takes longer until they are found compared to clamping a parameter to a fixed value.

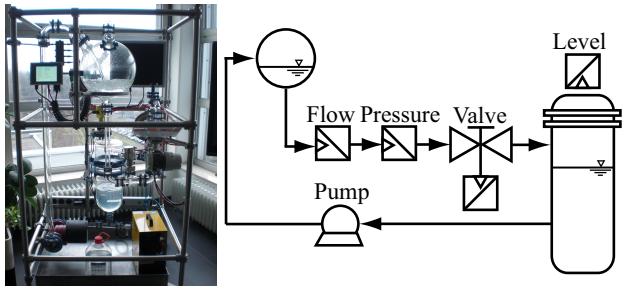

Fig. 5. Experimental setup where liquid is pumped around in cycles. On the right hand side the actor/sensor structure of the process is shown which is reconstructed from data.

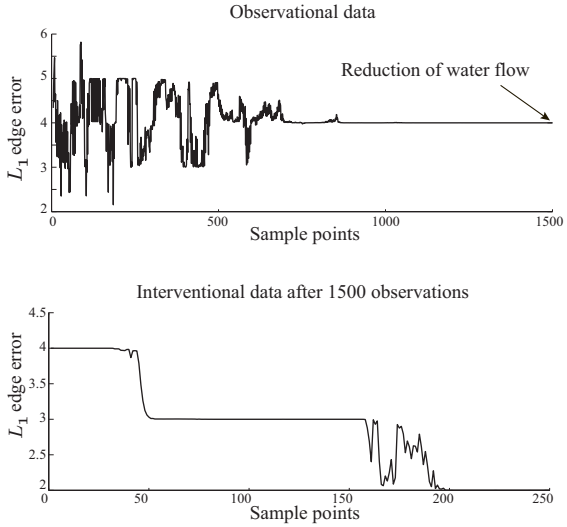

Fig. 6. L_1 edge error for observational data and after intervening on the process. Using only observational data, the BN generated from the data has an L_1 edge error of four. Using interventional data the error can be reduced to two.

B. Experimental laboratory plant

Finally the method is tested on an experimental laboratory plant. This station consists essentially of two containers where liquid is pumped around in cycles with different valve positions [4]. The pumping power is kept constant during the whole time which means that the process is run in feed-forward control. The process as well as the structure of the sensors and actors is shown in figure 5. In addition there is a stop cock between the pump and the flow meter where a (soft) intervention can be performed to push the process into a different set-point.
It has to be mentioned that this process does not include collisions meaning that during the first phase the skeleton will be obtained but no edges will be directed. As well as in the chemical reactor it is assumed that no information except the measured data is available.

1) Results: The results using only observational data are shown in the upper plot in figure 6. This shows that after 700

data points the BN has converged with a L_1 edge error of four after fluctuation between an edge error of three and of five. Figure 7 shows the resulting BN including the skeleton and the directed edges after calculating the edge uncertainties. The skeleton shows that not all conditional independences could be detected. By calculating the edge uncertainties in spite of the chemical reactor no edges were removed but all edges point in the direction of the liquid flow and thus describing the right causal associations.

To reduce the $L1$ edge error and to detect the underlying causal structure as a next step soft interventions are performed. In detail the stop cock is used to reduce the liquid flow in the process by half. For this the pushing factor ω is selected to have the value 0.5.

The results of the intervention are shown in figure 6 in the lower plot. After fifty data points the $L1$ edge error is reduced to three and after a total of 200 data points another edge can be directed and the error is further reduced to a value of two. By performing only an intervention in form of reducing the water flow it is not possible to reduce the edge error further.

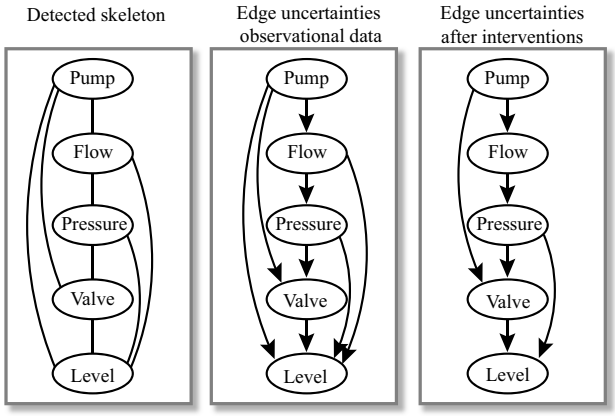

Fig. 7. Different steps of the calculation of the causal structure. After performing all steps the causal structure can be retrieved from the data with two additional directed edges.

V. Conclusion and future work

It has been investigated if it is possible to detect not only correlations but cause-effect relationships in measurement data coming from industrial processes.

The results show that using the developed hybrid learning method it is possible to discover the underlying causal structure of the process by using only measurement data. For this first a constraint-based method is used to detect the skeleton of the graph and directing the edges coming from root nodes. As a second part the edge uncertainties calculated using a scoring-based method. By using only observational data it is already possible to detect the cause-effect relationship of the main parameters without having any prior knowledge of the process.

In addition soft interventions can be used to detected and direct more edges from the underlying causal structure. It has been shown that this approach is useful as it is usually not possible to clamp nodes to a fixed value without damaging a process but it is possible to perform small parameter changes. The results show that performing soft pushing takes some data points longer than performing hard pushing on the process but comes to the same results.

The performance of this approach has be shown on a theoretical derived chemical stirred tank reactor and finally by using an experimental station the practicability of the approach has been proven. Regarding the stirred tank reactor the complete acyclic graph could be found, regarding the experimental plant the structure could be found with an L_1 edge error of two.

There are many interesting directions for further research. A first part will be to extend this work to develop an active learning strategy based on the calculated edge uncertainties. In addition it should be investigated how soft interventions can be used to uncover the existence of hidden variables that were not included in constructing the BN. Finally the above described technique can be adapted to learn the structure of a dynamic Bayesian network as the factor time is not yet included.

References

[1] D. Dash, M. Druzdzel *Robust independence testing for constraint-based learning of causal structure*. IEEE UAI, 2003.

[2] D. Eaton, K. Murphy *Belief net structure learning from uncertain interventions*. Journal of Machine Learning Research 1, 2000.

[3] N. Friedmann, *Inferring Cellular Networks using Graphical Models*. Science (5659), 2004.

[4] C. Frey, *Diagnosis and Monitoring of Complex Industrial Processes based on Self-Organizing Maps and Watershed Transformations*. IEEE CISMA, 2008.

[5] A. Gopnik, L. Schulz *Causal learning: Psychology, philosophy, and computation*. Oxford University Press, 2007.

[6] C. Hsu *Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry*. Expert Systems with Applications, 2009.

[7] H. Liu et al. *Bayesian Network Method for Fault Diagnosis in a Continuous Tubular Reactor*. Chemical Product and Process Modeling, 2010.

[8] D. Heckerman, D. Geiger and M. Chickering *Learning Bayesian networks: the combination of knowledge and statistical data*. Machine Learning, 1995.

[9] C. Kuehnert, T. Bernard, *Extraction of optimal control patterns in industrial batch processes based on Support Vector Machines*. IEEE Multi-conference on Systems and Control, 2008.

[10] V. Lagani, I. Tsamardinos *Structure-based selection for survival data*. Bioinformatics (26), 2010.

[11] M. Last, A. Kandel, H Bunke *Data Mining in Time Series Databases*. World Scientific, 2004.

[12] F. Markowetz, S. Grossmann and R. Spang *Probabilistic Soft Interventions in Conditional Gaussian Networks*. Artificial Intelligence and Statistics, 2005.

[13] K. Murphy, *Active Learning of Causal Bayes Net Structure*. Technical Report, 2001.

[14] J. Pearl, *Causality*. Cambridge University press, 2000.

[15] P. Spirtes, C. Glymour, R. Scheines *Causation, prediction, and search*. The MIT Press, second edition , 2001.

[16] S. Tong and D. Koller *Active Learning for Structure in Bayesian Networks*. International Joint Conference on Artificial Intelligence , 2001.

[17] I. Tsamardinos, L. Brown, C. Aliferis *The max-min hill-climbing Bayesian network structure learning algorithm*. Machine learning , 2006.

[18] J.G. van de Vusse *Plug-flow type reactor versus tank reactor*. Cehm. Eng. Sci., 1964.

[19] S. Verron, T. Tiplica, A. Kobi *Monitoring of complex processes with Bayesian networks*. Bayesian Networks Sciyo, 2010.

[20] J. Xu, Z. Hou *Notes on Data-driven System Approaches*. Acta Automatica Sinica, 2009.