

Machines **2014**, *2*, 255–274; doi:10.3390/machines2040255

OPEN ACCESS

machines

ISSN 2075-1702

www.mdpi.com/journal/machines

Article

Data-Driven Methods for the Detection of Causal Structures in Process Technology

Christian Kühnert ^{1,*} and Jürgen Beyerer ^{1,2}

¹ Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Fraunhoferstraße 1, Karlsruhe 76131, Germany; E-Mail: juergen.beyerer@iosb.fraunhofer.de

² Institute for Anthropomatics, Karlsruhe Institute of Technology, Adenauerring 4, Karlsruhe 76131, Germany

* Author to whom correspondence should be addressed;
E-Mail: christian.kuehnert@iosb.fraunhofer.de; Tel.: +49-721-6091-511;
Fax: +49-721-6091-413.

External Editor: David Mba

Received: 5 March 2014; in revised form: 5 August 2014 / Accepted: 13 October 2014 /

Published: 4 November 2014

Abstract: In modern industrial plants, process units are strongly cross-linked with each other, and disturbances occurring in one unit potentially become plant-wide. This can lead to a flood of alarms at the supervisory control and data acquisition system, hiding the original fault causing the disturbance. Hence, one major aim in fault diagnosis is to backtrack the disturbance propagation path of the disturbance and to localize the root cause of the fault. Since detecting correlation in the data is not sufficient to describe the direction of the propagation path, cause-effect dependencies among process variables need to be detected. Process variables that show a strong causal impact on other variables in the process come into consideration as being the root cause. In this paper, different data-driven methods are proposed, compared and combined that can detect causal relationships in data while solely relying on process data. The information of causal dependencies is used for localization of the root cause of a fault. All proposed methods consist of a statistical part, which determines whether the disturbance traveling from one process variable to a second is significant, and a quantitative part, which calculates the causal information the first process variable has about the second. The methods are tested on simulated data from a chemical stirred-tank reactor and on a laboratory plant.

Keywords: root cause localization; causal structure discovery; time series analysis

1. Introduction

Modern industrial plants are complex systems that need to run over several weeks or months. During a production run, operating conditions can change, which can lead to abnormal behavior of the process. Since modern plants' control and measurement devices are strongly cross-linked with each other, a failure in a major piece of equipment can potentially lead to plant-wide disturbances and for example result in a flood of alarms, making it difficult to localize the root cause of the disturbance. As not all relations of the different process parameters are well known, data-driven methods can be of great help to localize or at least to narrow down on the cause of a disturbance.

Backtracking the disturbance propagation path using data-driven methods means to detect temporal cause-effect relationships in a data set. In detail, this means that statistical relationships and time-shifts are used to reconstruct the propagation direction of the disturbance. Several methods have been already developed to test for temporal causal dependencies in data:

One of the first approaches was made by Granger [1], who compares two autoregressive models. The first model contains only past values of itself; the second model is augmented with past values of another variable. If the augmentation improves the regression, it is assumed that this variable has a causal impact on the other. In [2], an algorithm for root cause localization based on the cross-correlation function is presented for causal analysis, especially when having valve stiction. Schreiber [3] presents a concept, named transfer entropy, which detects causal dependencies by measuring the reduction of uncertainty when one variable predicts future values of the other. Further methods for the detection of causal dependencies are proposed in terms of dynamic Bayesian networks [4,5] or nearest neighbor approaches [6]. An overview about different methods, tested on artificial benchmark data sets and for biosignal analysis, is given in [7].

In this paper, different algorithms are proposed, which are based on the cross-correlation function, the transfer entropy, Granger causality and support vector machines. Next, the results of the different methods are combined, and a root cause priority list is generated. This priority list contains a ranking of the different process variables describing their possibility as being the actual cause of the fault.

All proposed algorithms consist of a statistical test, which determines whether the disturbance traveling from one process variable to a second is significant, and a quantitative part, called the causal strength, which defines the influence the input variable has on the output variable. For all methods, the causal strength takes values between zero(= no causal dependency) and one (= causal dependency). Finally, this value is used for generating the root cause priority list. The paper centralizes the main results given in [8], containing the following main contributions: (1) a proposal of a new algorithm based on support vector machines by using a recursive variable selection and model reduction approach; (2) the development of a design approach to combine all methods into one causal matrix and transfer into a root cause priority list; (3) the extension of an existing method based on the cross-correlation function by

using permutation tests for the significance test; and (4) the development of a visualization method for the causal matrices. The paper is structured as follows:

In Section 2, the foundations for the detection of causal dependencies in dynamic systems are introduced. Section 3 explains the algorithms. In Section 4, it is pointed out how the results of the different methods can be combined and how the root cause priority list is calculated. Additionally, a new way to visualize the results from the different methods is described. Finally, Section 5 tests the methods on data from a simulated chemical stirred-tank reactor and on an experimental laboratory plant.

2. Detecting Causal Dependencies in Process Measurements

Definition (causal system): A time-invariant system is causal if, for all input signals with $u_1(t) \equiv u_2(t)$ and $t \leq t_1$, for any t_1 , the output signals for $0 \leq t \leq t_1$ show the characteristic $y_1\{x_0, u_1(t)\} \equiv y_2\{x_0, u_2(t)\}$ (with x_0 : the initial state). Systems that are not causal are called acausal. In causal systems, the input signals for $t > t_1$ do not have an impact on the behavior of the output signal until time t_1 . Additionally, for causal systems, the impulse response for $t < 0$ is zero.

This definition of causality results in several system theoretic consequences. For all static systems $f : \mathbb{R} \rightarrow \mathbb{R}$, causality exists, even if u or y is seen as the input signal, since in both cases, the output signal is not depending on future values, but on the current value of the input signal.

Additionally, all linear time-continuous systems with one input and one output signal are causal, since $Y(s) = G(s)U(s)$ and $U(s) = G^{-1}(s)Y(s)$ explain the data equally well ($U(s), Y(s)$: Laplace-transformed input and output signals). In that case, it is only possible to test if the transfer function $G(s)$ is realizable (order of the numerator polynom \leq order of the denominator polynom). Still, this does not contain information about causality.

To backtrack the disturbance propagation path in a plant, a signal decay time or a signal dead time needs to be present in the measurements. Decay times between $u(t)$ and $y(t)$ characterize a smoothing effect, so that $y(t)$ seems to be delayed with respect to $u(t)$. A dead time T_D between $u(t)$ and $y(t)$ exists, if $y(t)$ does not depend on $u(\tau)$ in $\tau \in (t - T_D, t]$, but only on $u(t)$ in $\tau \leq t - T_D$. In other words, dead times describe the interval a change in the input signal needs to become visible on the output of the system. In industrial processes dead times exist, e.g., in tubes, when measuring a fluid concentration on different positioned sensors.

Another point of view on causality is given by Pearl [9]. The central idea is that a cause C increases the probability of the appearance of an effect E . This means that C can only be the cause of E , if $P(E = 1|C = 1) > P(E = 1|C = 0)$ is fulfilled. However, an increase of the probability can only be done through an active intervention. Therefore, Pearl introduces the do-operator, which forces setting C on a fixed value. In other words, C only has a causal impact on E , if $P(E = 1|\text{do}(C = 1)) > P(E = 1|\text{do}(C = 0))$ is fulfilled. This approach is the only possibility to avoid the detection of false causal dependencies.

An example for the detection of a false causal dependency is a non-measured signal u_z , which has an impact on u , as well as a later one on y , without having a direct dependency pointing from u to y . In [10], it is shown how the approach given by Pearl can be used to detect causal structures in process

data. Since the methods proposed in this paper only rely on observational data, meaning that no active intervention is performed, this approach will not be pursued in this paper.

3. Methods for Reconstructing the Disturbance Propagation Path

3.1. Cross-Correlation Function

The cross-correlation function (CCF) [11] quantifies the linear similarity of two equidistant sampled time series $u[k], y[k]$ that are time-shifted by a constant lag λ . The CCF is defined as:

$$\hat{c}_{uy}[\lambda] = \frac{1}{K} \frac{\sum_{k=1}^{K-\lambda} (u[k] - \hat{\mu}_u)(y[k + \lambda] - \hat{\mu}_y)}{\sqrt{\sum_{k=1}^{K-\lambda} (u[k] - \hat{\mu}_u)^2 \sum_{k=1}^{K-\lambda} (y[k] - \hat{\mu}_y)^2}}, \quad (1)$$

with $\lambda \in \{1 - K, 2 - K, \dots, K - 2, K - 1\}$ and:

$$\hat{\mu}_u := \frac{1}{K} \sum_{k=1}^K u[k], \quad \hat{\mu}_y := \frac{1}{K} \sum_{k=1}^K y[k], \quad (2)$$

with $\hat{c}_{uy}[\lambda] \in [-1, 1]$. $\max |\hat{c}_{uy}[\lambda]| = 1$ means that $u[k]$ and $y[k]$ are perfectly correlated at a shift λ , while values close to zero indicate that there is no correlation.

3.1.1. Significance Test

To check for a significant cause-effect relationship $u \rightarrow y$, two tests are established. The first test is used to check if $\max |\hat{c}_{uy}[\lambda]|$ differs significantly from zero; the second test checks if the causal strength from $u \rightarrow y$ differs significantly from $y \rightarrow u$ to have a clear indication of the direction of the propagation path.

Significant time-shifted correlation: Since the available samples of the two time series are limited, Equation (1) only gives an estimation of the CCF, so that $\max |\hat{c}_{uy}[\lambda]|$ randomly differs from zero for two uncorrelated signals. To test if $\max |\hat{c}_{uy}[\lambda]|$ differs significantly from zero, a hypothesis test is performed. We define:

$$\lambda^{\max} := \arg \max_{\lambda} |\hat{c}_{uy}[\lambda]|$$

with $\lambda \in \{1 - K, \dots, -1, 1, \dots, K - 1\}$. (3)

If $\lambda^{\max} > 0$ holds, there could be a causal relation from $u \rightarrow y$, and the hypothesis test is performed. To check if a significant correlation between the signals is present, a t -test with the null hypothesis defined as having no correlation is used with significance level $\alpha = 0.05$. If this test fails, it is assumed that no cause-effect relationship exists from $u \rightarrow y$.

Significant causal direction: This test covers the possibility that the resulting CCF can have a global maximum for $\lambda > 0$ and a slightly lower local maximum for $\lambda < 0$, meaning that the causal direction is not obvious. Therefore, this test checks if the maximum for $\lambda > 0$ is significantly different from the maximum for $\lambda < 0$. To perform the test, a compound parameter C^{CCF} defined as:

$$C^{\text{CCF}} := \frac{\max_{\lambda>0} |\hat{c}_{uy}[\lambda]| - \max_{\lambda<0} |\hat{c}_{uy}[\lambda]|}{\max_{\lambda>0} |\hat{c}_{uy}[\lambda]| + \max_{\lambda<0} |\hat{c}_{uy}[\lambda]|}, \quad (4)$$

with $-1 \leq C^{\text{CCF}} \leq 1$ is used. A significance value $C^{\text{CCF}} > 0$ indicates a causal dependency from $u \rightarrow y$. As C^{CCF} strongly depends on the characteristics of $u[k]$ and $y[k]$, an adaptive threshold is derived through a 3σ permutation test. Since performing a complete permutation of $u[k]$ destroys all causal information, the resulting value of C^{CCF} should be close to zero. This idea is exploited by calculating random permutations of $u[k]$ and generating several values for C_{π}^{CCF} . The threshold $C_{\text{thresh}}^{\text{CCF}}$ for each pair of variables is calculated as:

$$C_{\text{thresh}}^{\text{CCF}} := \mu_{C_{\pi}^{\text{CCF}}} + 3\sigma_{C_{\pi}^{\text{CCF}}}, \quad (5)$$

with $\mu_{C_{\pi}^{\text{CCF}}}$ being the mean value and $\sigma_{C_{\pi}^{\text{CCF}}}$ the standard deviation. If $C^{\text{CCF}} > C_{\text{thresh}}^{\text{CCF}}$ holds, the test has passed successfully.

If both significance tests are passed, the found causal dependency is assumed to exist.

3.1.2. Causal Strength

The causal strength of $u \rightarrow y$ is defined as:

$$Q^{\text{CCF}} := (\max(0, C^{\text{CCF}}))^{\beta_{\text{CCF}}}, \quad (6)$$

with $0 \leq Q^{\text{CCF}} \leq 1$. β_{CCF} is a design parameter to make the different methods numerically compatible, and its selection will be explained in Section 4. The proposed algorithm for the detection of cause-effect relationships for two time series $u[k]$ and $y[k]$ using the CCF is summarized as follows:

Algorithm 1: Algorithm based on cross-correlation.

1. Compute \hat{c}_{uy} of $u[k]$ and $y[k]$ and determine λ^{max} ;
 2. If $\lambda^{\text{max}} < 0$, then $u \nrightarrow y$, else test for a significant correlation;
 3. Calculate C^{CCF} and corresponding threshold $C_{\text{thresh}}^{\text{CCF}}$;
 4. Check if $C^{\text{CCF}} > C_{\text{thresh}}^{\text{CCF}}$; calculate the causal strength $u \rightarrow y$ by Q^{CCF}
-

3.2. Granger Causality

The Granger causality (GC) has been introduced by Clive Granger [1] and is traditionally used in the field of economics [12,13]. Recently, the application of GC is of growing interest, especially in the field of neuroscience [14] and biology [15].

The central idea is that it is assumed that a signal $u_i[k]$ has a causal influence on $y[k]$ if past values from $u_i[k]$ and $y[k]$ result in a better prediction of $y[k]$ than using only past values from $y[k]$ for prediction. GC takes into account that besides the signal $u_i[k]$, $y[k]$ can depend on further signals $u_l[k] = u_1[k], \dots, u_p[k]; l \neq i, y[k]$. A comparison is done using two vector autoregressive models and performing a one-step-ahead prediction. If the prediction error of the first model is substantially smaller than the one from the second model, a causal dependency $u_i \rightarrow y$ is concluded. For the proposed algorithm, each time series is once selected as output $y := u_m$, while the left $r - 1$ time series are used as input. This can be formulated with n defining the model order as:

$$E_{\bar{U}_i Y} = \sum_{k=n+1}^K \left(y[k] - \hat{a}_0 - \sum_{j=1}^n \hat{a}_j y[k-j] - \sum_{\substack{l=1, \\ l \neq i, m}}^r \sum_{j=1}^n \hat{b}_{lj} u_l[k-j] \right)^2, \quad (7)$$

$$E_{U_i Y} = \sum_{k=n+1}^K \left(y[k] - \hat{a}_0 - \sum_{j=1}^n \hat{a}_j y[k-j] - \sum_{\substack{l=1, \\ l \neq m}}^r \sum_{j=1}^n \hat{b}_{lj} u_l[k-j] \right)^2, \quad (8)$$

By definition, the parameters \hat{a}_j , \hat{b}_{lj} in Equations (7) and (8) result from separate estimations. The impact of the i -th input signal is measured in terms of the sum of the squares of residuals without u_i as $E_{\bar{U}_i Y}$ and with u_i as $E_{U_i Y}$ and is used to test for causal significance and for calculating the causal strength.

The performance of the model strongly depends on the model order n . A too small of a value for n leads to a large prediction error, and setting n too large results in an overfitted model. Therefore, the Akaike information criterion (AIC) [16] is used for model order estimation, while taking into account the prediction errors, the sample size K , the number of variables r and the model order n . For model order estimation, the loss function is once set to $V = E_{U_i Y}$ for the unrestricted and to $V = E_{\bar{U}_i Y}$ for the restricted model. The AIC is defined as:

$$\text{AIC}(n) = \log(V) + \frac{2np^2}{K}, \quad (9)$$

while the estimated order is selected as $\arg \min_n \text{AIC}(n)$. Furthermore, all models are tested for consistency by performing a Durbin-Watson statistic [17] on the residuals.

3.2.1. Significance Test

This test is performed to check whether $E_{\bar{U}_i Y}$ and $E_{U_i Y}$ differ significantly. According to [18], $E_{\bar{U}_i Y}$ and $E_{U_i Y}$ follow χ^2 distributions, and an F -test can be performed to verify if the time series $u_i[k]$ has a causal influence on $y[k]$. The test is performed on the restricted and unrestricted model under the null hypothesis that $E_{U_i Y} < E_{\bar{U}_i Y}$ with significance level $\alpha = 0.05$.

3.2.2. Causal Strength

The strength of the significant causal relationship $u_i \rightarrow y$ is defined as:

$$Q^{GC} := \left(1 - \frac{E_{U_i Y}}{E_{\bar{U}_i Y}}\right)^{\beta_{GC}}, \quad (10)$$

with $0 \leq Q^{GC} \leq 1$, since $E_{U_i Y} < E_{\bar{U}_i Y}$ counts. The design parameter β_{GC} will be determined in Section 4. The suggested algorithm is summarized as follows:

Algorithm 2: Algorithm based on Granger causality.

1. Compute $E_{\bar{U}_i Y}$ and $E_{U_i Y}$ using the model order estimated through AIC;
 2. Test for model consistency for both models using the Durbin–Watson statistic;
 3. Perform a significance test based on an F -test for $E_{U_i Y} < E_{\bar{U}_i Y}$;
 4. If the causal dependency is significant, calculate the causal strength $u_i \rightarrow y$ by Q^{GC} ;
-

3.3. Transfer Entropy

The transfer entropy (TE) is an information theoretic measure and was first introduced by Schreiber [3]. Applications of the TE can be found, e.g., in neuroscience [19,20] and in financial data analysis [21]. In the field of process engineering, research has been conducted by Bauer [6], who uses TE for the causal analysis of measurements taken from chemical processes.

From its definition, it can be used to detect causal dependencies by testing how much information is transferred from $u[k]$ to $y[k]$ and how much information is transferred from $y[k]$ to $u[k]$. The transition probability for $y[k]$ is defined as $P(y_{n+1}|\mathbf{y})$, which is used as short notation for:

$$P(y_{n+1} = d_{n+1} | y_n = d_n, \dots, y_1 = d_1), \quad (11)$$

with n defining the time horizon and $d_v \in \{1, \dots, D_y\}$ the quantization levels of $y[k]$. The transition probability $P(y_{n+1}|\mathbf{y}, \mathbf{u})$ is defined accordingly. The transfer entropy is then given as:

$$\text{TE}_{uy}^*(\lambda) = \sum_{\substack{d_1=1 \\ \vdots \\ d_{n+1}=1}}^{D_y} \sum_{\substack{e_1=1 \\ \vdots \\ e_n=1}}^{D_u} P(y_{n+1}, \mathbf{y}, \mathbf{u}) \log \frac{P(y_{n+1}|\mathbf{y}, \mathbf{u})}{P(y_{n+1}|\mathbf{y})}. \quad (12)$$

According to [21], the boundaries of the TE are $0 \leq \text{TE}_{uy}^*(\lambda) \leq H_y$, with H_y being the entropy of the output signal. To capture dead times in the data, the parameter λ is introduced to perform a backward-shifting of $u[k]$, and Equation (12) is calculated for different $u[k - \lambda]$. For the calculation of causal dependencies, the maximum value of the transfer entropy is set to $\text{TE}_{uy} := \arg \max_{\lambda} (\text{TE}_{uy}^*(\lambda))$.

To calculate the value of the time horizon n , the residual sum of squares of several vector autoregressive models is calculated:

$$\hat{\sigma}_y^2 = \sum_{k=n+1}^K \left(y[k] - \hat{a}_0 - \sum_{j=1}^n \hat{a}_j y[k-j] \right)^2, \quad (13)$$

with \hat{a}_j resulting from a least squares estimation. The used order n_{TE} is then chosen as the minimum of the Akaike information criterion defined as $AIC(n) = \log \hat{\sigma}_y^2 + \frac{2n}{K}$ [16].

3.3.1. Significance Test

Testing for a significant causal relationship is done by a test introduced in [3]. The key idea is to generate an adaptive threshold TE_{uy}^{thresh} based on the permuted input time series $u_\pi[k]$ and the generation of several values for $TE_{u_\pi y}$. The values of $TE_{u_\pi y}$ are finally used to calculate the threshold TE_{uy}^{thresh} in terms of a 3σ -test:

$$TE_{uy}^{thresh} := \mu_{TE_{u_\pi y}} + 3\sigma_{TE_{u_\pi y}}, \quad (14)$$

with mean value $\mu_{TE_{u_\pi y}}$ and standard deviation $\sigma_{TE_{u_\pi y}}$. If $TE_{uy} > TE_{uy}^{thresh}$ holds, a causal dependency $u \rightarrow y$ is concluded, and the causal strength can be calculated.

3.3.2. Causal Strength

By definition $TE_{uy} \leq H_y$, meaning that for normalizing to values between zero and one, the causal strength of the transfer entropy can be defined as:

$$Q^{TE} := \left(\frac{TE_{uy}}{H_y} \right)^{\beta_{TE}}. \quad (15)$$

The design parameter β_{TE} will be chosen in Section 4. The suggested algorithm for the detection of cause-effect relations is summarized as:

Algorithm 3: Algorithm based on transfer entropy.

1. Compute model order n_{TE} of the transfer entropy using VAR models and AIC;
 2. Compute $TE_{uy}^*(\lambda)$ of the time series $u[k]$ and $y[k]$,
set $TE_{uy} = \max(TE_{uy}^*(\lambda))$;
 3. Calculate TE_{uy}^{thresh} and check if $TE_{uy} > TE_{uy}^{thresh}$;
 4. Set Q^{TE} as the resulting value of the causal strength $u \rightarrow y$
-

3.4. Support Vector Machines for Regression

Support vector machines (SVM) are learning methods that are used for supervised learning. Originally, they were developed by Vapnik [22] for classification and later on were extended towards

regression and time series prediction [23]. In industrial processes, they are used, e.g., for fault detection and diagnosis [24] or optimization [25]. For the detection of causal structures in data, a model reduction approach is proposed.

Given a training set with $\{x_i, z_i\}_{i=1}^K$ with $x_i \in \mathbb{R}^n$ and $z_i \in \mathbb{R}$, an SVR means to find the regression function:

$$f(x) = \langle w, x \rangle + b, \quad (16)$$

containing as parameters the normal vector w , the bias b and $\langle \cdot, \cdot \rangle$, denoting the scalar product. The function $f(x)$ should have at most a deviation ϵ from the values z_i for the whole training data set, while seeking a normal vector w , which is as small as possible. As the selection of a too small insensitivity zone ϵ would lead to equations with infeasible constraints, slack variables $\xi, \hat{\xi} \in \mathbb{R}_{\geq 0}$ and an additional weighting parameter $C \in \mathbb{R}_{>0}$ are introduced leading to the objective function:

$$\begin{aligned} \min_{w, \xi, \hat{\xi}} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^K (\xi_i + \hat{\xi}_i) \\ \text{with} \quad & f(x_i) - z_i \leq \epsilon + \xi_i \\ \text{and} \quad & z_i - f(x_i) \leq \epsilon + \hat{\xi}_i. \end{aligned} \quad (17)$$

The optimization problem is solved by means of its dual containing the Lagrange multipliers $\alpha, \hat{\alpha}$. The solutions of the regression function is finally given by:

$$f(x) = \sum_{i=1}^K (\alpha_i - \hat{\alpha}_i) \langle x_i, x \rangle + b. \quad (18)$$

Kernel functions: One of the main reasons why SVMs are employed is their ability to deal with nonlinear dependencies by introducing so-called kernel functions; see, e.g., [26,27]. The input data \mathbb{R}^n is mapped into some feature space \mathcal{F} with a possibly higher dimension by using a nonlinear transformation function Φ and searching for the flattest function in the defined feature space. Since SVMs solely depend on the calculation of dot products, the computational complexity for this transformation remains feasible. A kernel function, defined as $k(x, x') := \langle \Phi(x), \Phi(x') \rangle$, can be plugged into Equation (18), resulting in the final regression function:

$$f(x) = \sum_{i=1}^N (\alpha_i - \hat{\alpha}_i) k(x_i, x) + b. \quad (19)$$

In this paper, only Gaussian kernels, defined as $k(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$, are used for the detection of causal structures. To optimize the parameters ϵ, C and σ , a downhill simplex algorithm [28] is used.

3.4.1. Detecting Causal Dependencies

To detect cause-effect dependencies, an input data set $\Psi_{uy} = \{u[k-1], \dots, u[k-n], y[k-1], \dots, y[k-n]\}$ for predicting $y[k]$ is generated. The SVM is once trained and optimized regarding ϵ, C and σ on the complete data set. The time horizon n is estimated as given in Equation (13) for the transfer entropy.

To detect cause-effect dependencies, an input data set $\Psi_{uy} = \{u[k-1], \dots, u[k-n], y[k-1], \dots, y[k-n]\}$ for predicting $y[k]$ is generated. The SVM is once trained and optimized regarding ϵ , C and σ on the complete data set. The time horizon n is estimated as given in Equation (13).

In the first step, the variables in Ψ_{uy} are ranked in terms of their prediction accuracy of $y[k]$ by performing a recursive variable elimination algorithm based on the Lagrange multipliers as proposed by [29].

In the second step, a relevant subset of input variables is selected. If the resulting subset contains one or several past values of $u[k]$, it is assumed that u causes y . For the selection of the size of the subset, an F -test [11] ($\alpha = 0.05$) is performed on the resulting residual sum of squares of the two SVMs, while the first SVM contains ψ variables and the second SVM $\psi + 1$ variables. If the null hypothesis cannot be rejected, the residual sum of squares does not change significantly, and the found subset of variables is set to size ψ .

3.4.2. Causal Strength

Similar to Granger causality, the causal strength is calculated based on the comparison of the squared sum of residuals. The causal strength Q^{SVM} is calculated through a comparison of the two different squared sums of residuals, named E_{uy} and E_y , while $E_{uy} \leq E_y$. In detail, E_{uy} is calculated using the above explained SVM with the subset of input variables resulting from the initial set Ψ_{uy} . For the prediction of $y[k]$, the residual sum of squares E_y is calculated by performing the same algorithm, only starting with the reduced set $\Psi_y = \{y[k-1], \dots, y[k-n]\}$, which does not contain the time series $u[k]$, and by using the same parameters ϵ_{opt} , C_{opt} and σ_{opt} .

Again a tuning parameter β_{SVM} is defined. Selecting the parameter value is postponed to Section 4. The resulting value Q^{SVM} is therefore defined as:

$$Q^{\text{SVM}} := \left(1 - \frac{E_{uy}}{E_y}\right)^{\beta_{\text{SVM}}}, \quad (20)$$

with $0 \leq Q^{\text{SVM}} \leq 1$, where zero equals no causal dependency and one means maximum causal strength. The complete algorithm using support vector machines for the detection of causal dependencies is summarized as Algorithm 4.

Algorithm 4: Algorithm based on support vector machines.

1. Estimate the time horizon n_{SVM} using a VAR model and AIC to generate Ψ_{uy} ;
 2. Train SVM and fit user-selected parameters ϵ , C , σ using downhill simplex algorithm and check the consistency of the SVM using the Durbin–Watson statistic;
 3. Perform variable selection and calculate subset;
 4. If u is in the subset, set Q^{SVM} as the resulting value of the causal strength $u \rightarrow y$;
-

As for each pair of variables a distinct SVM needs to be trained and tuned, this algorithm is the most complex one of the four presented. Regarding the transfer entropy, transition probabilities need to be calculated, meaning that this method can also become computationally intense. The cross-correlation function and Granger causality, being linear measures, are rather cheap to compute. For a detailed comparison of the different algorithms, containing a large set of benchmark data, refer to [8].

4. Reconstruction of the Disturbance Propagation Path and Localization of the Root Cause

Each pair of process variables results in a value that represents the causal influence one variable has on the other. For displaying the complete information, all relationships are written into a causal matrix $Q \in \mathbb{R}^{r \times r}$ defined as:

$$Q := \begin{bmatrix} - & q_{X_2 \rightarrow X_1} & \cdots & q_{X_r \rightarrow X_1} \\ q_{X_1 \rightarrow X_2} & - & \cdots & q_{X_r \rightarrow X_2} \\ \vdots & \vdots & \ddots & \vdots \\ q_{X_1 \rightarrow X_r} & q_{X_2 \rightarrow X_r} & \cdots & - \end{bmatrix}, \quad (21)$$

consisting of the causal strengths $q \in \{0, \dots, 1\}$ of the process variables X_i with $i = 1, \dots, r$. In the matrix, the row index represents the variable that is the causing candidate and the column index representing the effect candidate. Values close to zero describe weak causal strengths, and values close to one describe strong ones.

4.1. Balancing and Combining Causal Matrices

Since each method uses a different mathematical approach, the causal matrices of the four methods are not comparable directly. To make them comparable to each other, the prior introduced exponential fitting parameters $\beta^{\text{CCF}}, \beta^{\text{TE}}, \beta^{\text{GC}}, \beta^{\text{SVM}} \in [0, \infty)$ are used.

The proposed design approach is based on the assumption that, on average, all methods will work equally well on the data set. In that case, equally well means that for the found significant causal dependencies, all causal matrices will result in the same mean value. Hence, the value of each β parameter is fitted in a way so that the matrices $Q^{\text{CCF}}, Q^{\text{TE}}, Q^{\text{GC}}$ and Q^{SVM} give the same mean for the significant cause-effect relationships. Regarding the investigated use cases in Section 5, the mean value for the causal matrix from each method is set to 0.5.

Finally, to calculate the combined causal matrix, the mean is taken over all balanced causal matrices for all causal dependencies. In that case, non-significant causal dependencies are set to zero.

4.2. Root Cause Priority List

This list contains a ranking of the analyzed process variables with regard of their possibility of being the actual root cause. As a consequence, a value defined as RC is associated with each variable. This is done by summing up the causal influence one variable has onto the other variables defined as:

$$\text{RC}_n := \sum_{i=1, i \neq n}^r q_{X_n \rightarrow X_i}. \quad (22)$$

The variable having the maximum value of RC is ranked first, meaning that this variable is most likely to be the root cause of the disturbance. Table 1 outlines the representation of the root cause priority list.

Table 1. Root cause priority list from the causal matrix.

Rank	Process variable	RC
1	X_n	$\sum_{i=1, i \neq n}^r q_{X_n \rightarrow X_i}$
\vdots	\vdots	\vdots
r	X_k	$\sum_{i=1, i \neq k}^r q_{X_k \rightarrow X_i}$

4.3. Visualizing Causal Matrices

Several techniques have been already developed that deal with the visualization of the causal matrices. In [18], circular directional charts are suggested, and bubble charts are proposed in [6]. In [5], it is suggested to use heat maps to illustrate causal dependencies. Still, all of the methods have as a drawback that only one causal matrix can be visualized at a time. Hence, to compare the different causal matrices better, several ways for visualization are utilized in this paper.

Partially-directed graph: In these graphs, process variables are represented by nodes and causal dependencies by directed edges. It is possible that several edges point onto one node or that several edges leave one node. The main purpose of this representation is to give a fast overview of the disturbance propagation path, while the root cause is the first variable of the chain. Furthermore, the size of the arrowhead is used to indicate the strength of the causal dependency. The graph represents the combined causal matrix.

Doughnut chart: These graphs are circular charts that are divided into several sectors, while having a blank center. To represent the causal matrices, the quantity of each sector results from the calculated entries in Q plus one blank sector. The value in the middle of the doughnut represents the combined causal strength.

Bar chart: Bar charts represent values in the form of rectangular bars while having their length proportional to the causal strength. This visualization avoids the drawback of the doughnut chart, as the different sections are hard to compare, since they are bent.

5. Use Cases

Two use cases, namely a simulated continuously stirred-tank-reactor and a laboratory plant, are used to test the methods. Further experiments on the laboratory plant and more simulation results on the tank-reactor, as well as on other benchmark data sets can be found in [8].

5.1. Continuously-Stirred-Tank Reactor

To study the performance of the different methods, the model of a continuously-stirred-tank reactor (CSTR), explained in [30], is used. The underlying chemical reaction scheme consists of two irreversible follow-up reactions, where an educt A reacts to an intermediate product B , and this reacts to the resulting product C . The reactants are dissolved in a fluid and can be measured in terms of the three concentrations c_A, c_B, c_C at the outlet of the CSTR. The CSTR is continuously filled, while the fluid has the reactant concentration c_{in} and the temperature ϑ_{fl} . V describes the volume of the CSTR and F the selected volume flow rate. The parameters k_1 and k_2 are empirical parameters and describe the relationship between the temperature and speed of the chemical reaction. E_1 and E_2 are the activation energies of the reactants, and R is the universal gas constant. The parameter values are given in Table 2.

Table 2. Continuously-stirred-tank reactor (CSTR) parameters taken from [30].

Parameter	Value	Unit
F	100	L/min
V	100	L
k_1	7.2×10^{10}	1/min
k_2	5.2×10^{10}	1/min
E_1/R	8750	K
E_2/R	9750	K

Finally, the underlying differential equations of the CSTR are:

$$\begin{aligned}
 \dot{c}_A(t) &= \frac{F}{V}(c_{in}(t) - c_A(t)) - k_1 c_A(t) e^{-E_1/(R\vartheta_{fl})}, \\
 \dot{c}_B(t) &= k_1 c_A(t) e^{-E_1/(R\vartheta_{fl})} - k_2 c_B(t) e^{-E_2/(R\vartheta_{fl})} \\
 &\quad - \frac{F}{V} c_B(t), \\
 \dot{c}_C(t) &= k_2 c_B(t) e^{-E_2/(R\vartheta_{fl})} - \frac{F}{V} c_C(t).
 \end{aligned} \tag{23}$$

The set-points of the two input variables are chosen as $\vartheta_{fl,OP} = 350$ K and $c_{in,OP} = 1$ mol/L, while ϑ_{fl} is superposed with white noise having $\mathcal{N}(0, 3 \text{ K}^2)$ and c_{in} with white noise having $\mathcal{N}(0, 0.1 (\text{mol/L})^2)$. To calculate the causal matrices, in total, $K = 1,000$ samples are used. An extract of the data set used for the analysis is given in Figure 1. From the differential equations, it is expected that the methods deliver as a result the disturbance propagation path $\vartheta_{fl} \rightarrow c_{in} \rightarrow c_A \rightarrow c_B \rightarrow c_C$. ϑ_{fl} should be ranked in first position, since it has an impact on all three chemical reactions.

The results are illustrated in Figure 2, with the red squares marking the expected causal dependencies. The design parameters result in $\beta_{CCF} = 0.51$, $\beta_{TE} = 0.11$, $\beta_{GC} = 0.59$ and $\beta_{SVM} = 0.28$.

Figure 1. Simulated data from the CSTR used for the analysis.

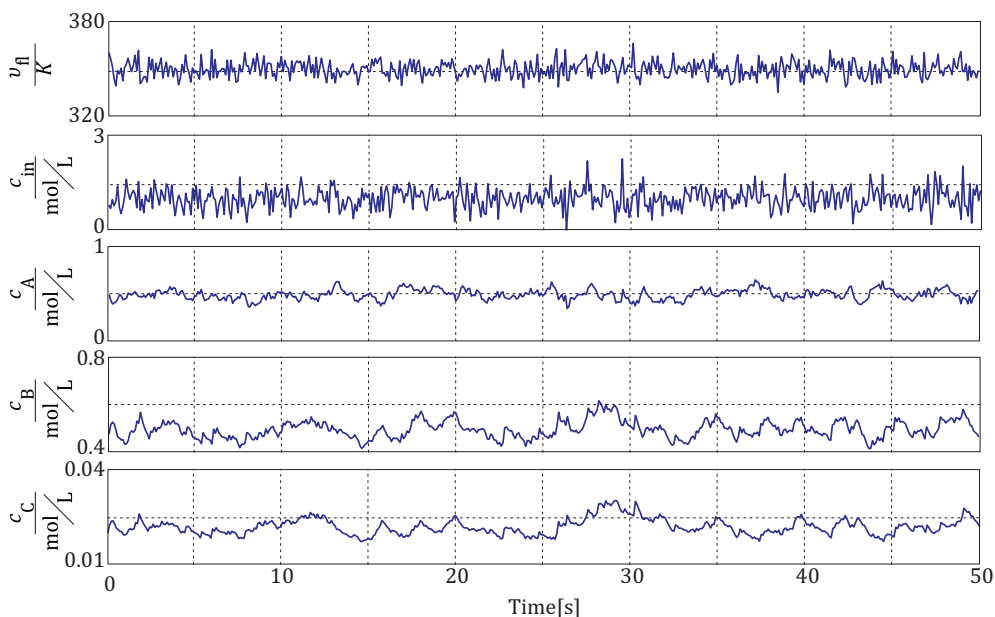
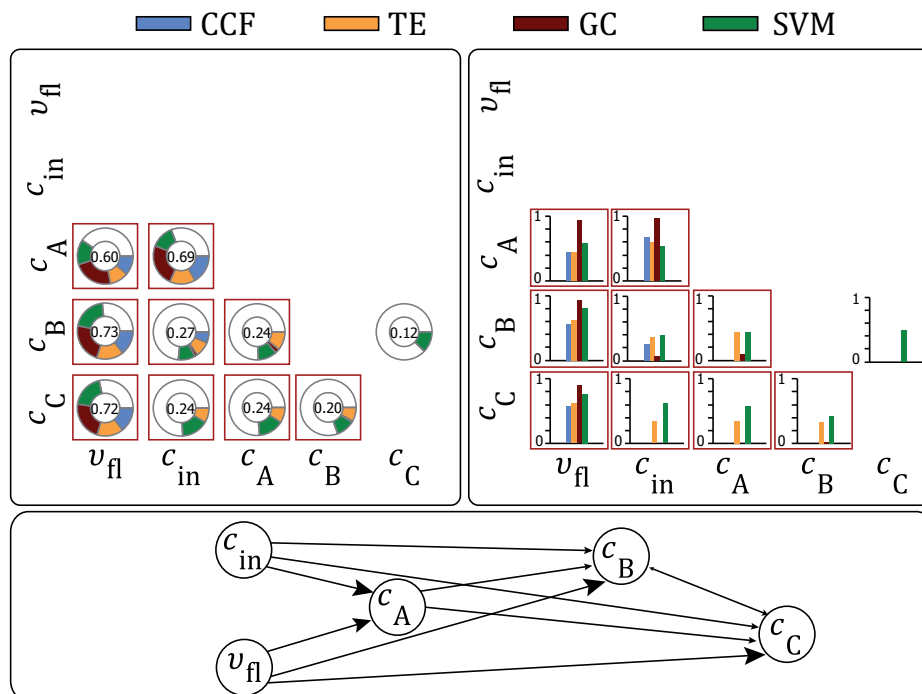


Figure 2. Causal matrices for the stirred-tank-reactor. The red squares describe the expected causal dependencies to be detected by the methods.



Analyzing the bar chart in Figure 2 illustrates that all methods detect a large causal strength of v_{fl} pointing towards the other process variables. This becomes obvious when taking into account the underlying differential equations of the CSTR (see Equation (23)), as the temperature has a direct impact on all three concentrations. Furthermore, the result indicates that the nonlinearity implied in the exponential function has been correctly fitted by the Granger causality and cross-correlation function.

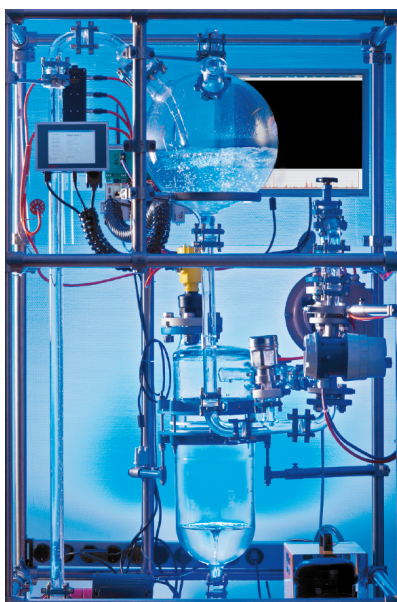
Another strong causal strength has been found from $c_{in} \rightarrow c_A$. This can also be explained through the differential equations, as c_{in} has a direct impact on c_A . The relationship $c_{in} \rightarrow c_B$ is the only indirect causal dependency detected by all methods, and $c_A \rightarrow c_B$ and $c_B \rightarrow c_C$ are detected by GC, TE and the SVM. The cause-effect dependency $c_{in} \rightarrow c_C$ and $c_A \rightarrow c_C$ has been found by TE and the SVM. Except the SVM, which detects the wrong causal dependency $c_C \rightarrow c_B$, no wrong causal dependencies are found by the other methods. Furthermore, the transfer entropy is the only method that detects all expected causal dependencies. The propagation path from the combination of the methods is given in Table 3.

Table 3. Root cause priority list calculated from the causal matrix in Figure 2.

Rank	Process variable	RC
1	ϑ_{fl}	2.05
2	c_{in}	1.20
3	c_A	0.48
4	c_B	0.20
5	c_C	0.12

The resulting root cause priority list is given in Table 3. ϑ_{fl} is correctly detected as being the root cause, and c_{in} , being the source of disturbances, is ranked second. Furthermore, the results show that causal dependencies at the end of the reaction chain result in weaker causal strengths or that these dependencies do not pass significance tests. The reason is that the disturbances are low-pass filtered each time a reaction takes place, so that less fluctuations for inferring causal dependencies are present in the data. Additionally, when merging the methods into one resulting causal matrix, the correct causal dependencies obtain a much stronger weighting compared to the causal matrices resulting from only one method at a time.

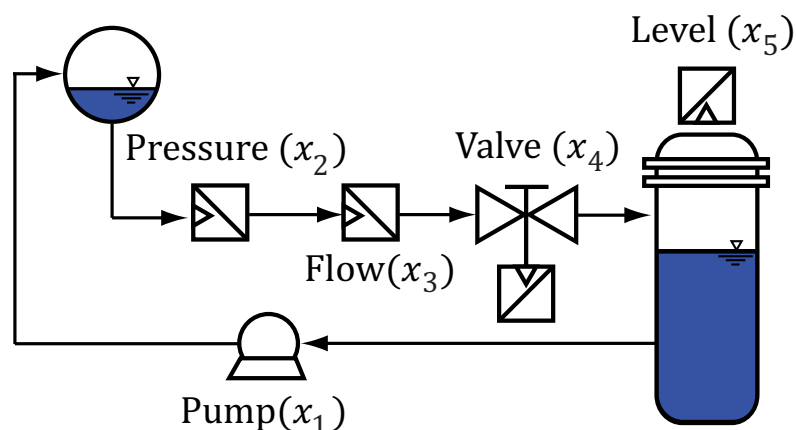
Figure 3. Experimental setup of the laboratory plant.



5.2. Experimental Laboratory Plant

To evaluate the methods on real-world data, a fault is generated in an experimental laboratory plant that pumps water in cycles. A photo of the plant is given in Figure 4, and the connection of the different process devices is sketched in Figure 4. The process starts by setting a pump (x_1) positioned on the lower side of the plant into feed-forward control to transfer water into the ball-shaped upper tank. From the upper tank, the water passes several measurement devices before flowing into a lower cylindrical tank. Finally, the water flows from the lower tank back to the pump and closes the water cycle. Between the two tanks, pressure (x_2) and flow (x_3) are measured. With a valve (x_4), placed between the flow meter and the lower tank, the water flow can be controlled. Additionally, the filling level (x_5) is measured in the lower tank.

Figure 4. Schematic drawing of the laboratory plant.



To generate a fault, a connection cable between the valve and compressor is removed and reattached randomly. The pump is set to 50% of its maximum feeding rate. The resulting data are illustrated in Figure 5. The instance the valve closes, the water is blocked from flowing from the upper to the lower tank. As it reopens, the process goes back into stationary phase. This means for the process variables that the flow reduces, while the level meter measures a continuous reduction of the water in the lower tank. The hydraulic pressure increases, until the pump stops delivering water from the lower to the upper tank. It is expected that the valve (x_4) is detected as being the root cause of the disturbance. The results are given in Figure 6, where the red squares mark the expected causal dependencies, and the design parameters result in $\beta_{CCF} = 0.33$, $\beta_{TE} = 0.26$, $\beta_{GC} = 0.31$ and $\beta_{SVM} = 0.82$.

No method can detect all expected causal dependencies, but all expected causal dependencies are found by at least two methods. Table 4 gives the root cause priority list from the combined causal matrix. The valve is set on position one and is therefore correctly detected as being the root cause of the fault.

Like in the data from the CSTR, the outcome shows that when merging the methods into one resulting causal matrix, the correct causal dependencies obtain in a stronger weighting, meaning that the wrongly detected causal dependencies from some methods become less relevant.

Figure 5. Data of the laboratory plant when having a faulty valve. Sampling rate $T_s = 2$ s.

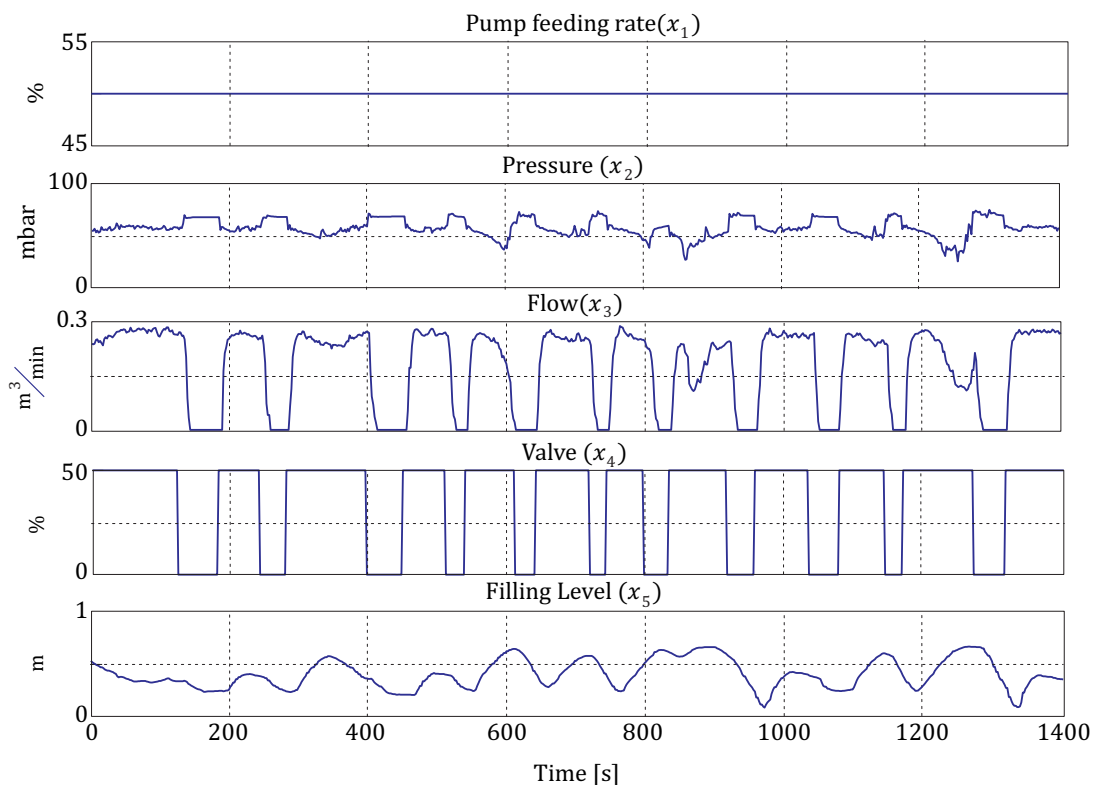


Figure 6. Causal matrices for the laboratory plant. The red squares describe the expected causal dependencies to be detected by the methods.

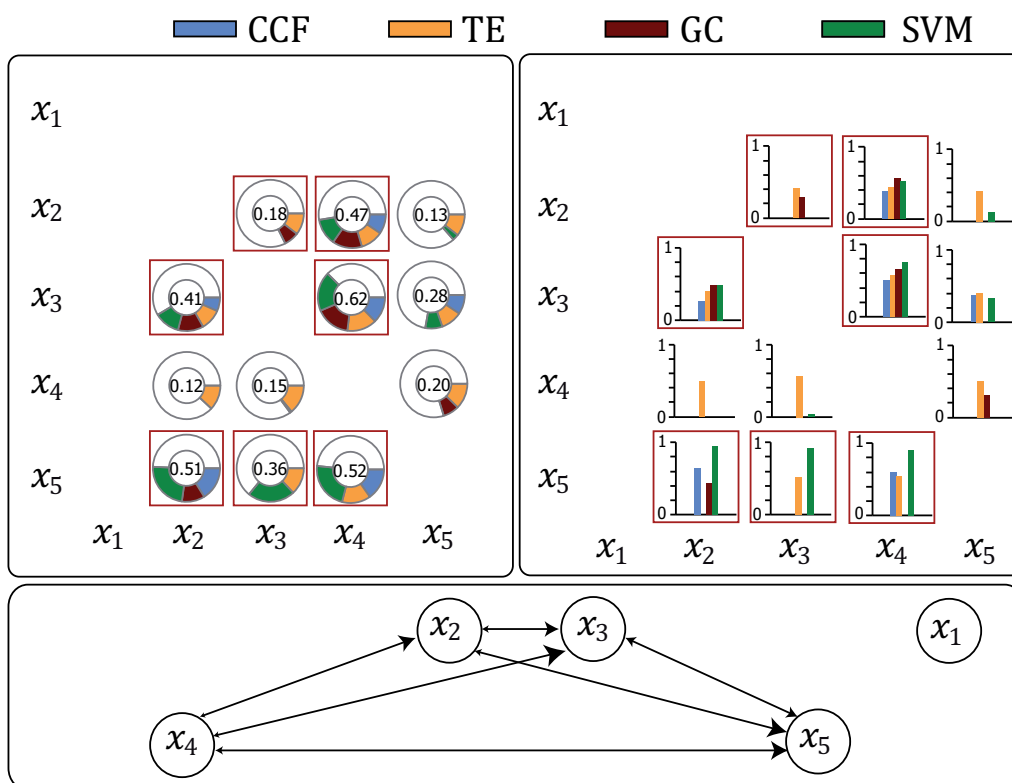


Table 4. Root cause priority list calculated from the causal matrix in Figure 6.

Rank	Process Variable	RC
1	Valve opening (x_4)	1.61
2	Pressure (x_2)	1.04
3	Flow (x_3)	0.69
4	Filling level (x_5)	0.61
5	Pump feeding rate (x_1)	0

6. Conclusion and Future Work

Several data-driven methods have been proposed to detect causal dependencies in measurements by exploiting information contained in time-shifts and statistical relationships.

As use cases, the methods were applied to backtrack the disturbance propagation path and localize the root cause with data coming from a simulated stirred-tank reactor and from a laboratory plant. In both cases, the causing variable of the fault was found correctly. Additionally, the results of the use cases showed, that it seems to be useful that more than one method is applied to perform an analysis. Since a found causal dependency is only a hypothesis, it gives more evidence if different methods indicate the same causal dependencies.

There is much room for future research. As all proposed methods localize faults solely from process data, one part of future work will focus on the integration of prior knowledge available from a plant (e.g., through the scheme of the plant or known process characteristics). Additionally, attenuation of fluctuations along their propagation through the system can be used as a further source of information for reconstructing the propagation path. Another topic is how the methods can be extended to work on MIMO systems, as it is also possible that two faults occur at the same time in a plant. Finally, the approach for combining the methods shows good results, but is rather *ad hoc*. Approaches using fuzzy decision making or Bayesian statistics need to be investigated.

Author Contributions

The main contributions that have been presented in this paper are: (1) a new algorithm based on support vector machines for the detection of causal dependencies in data by using a recursive variable selection and model reduction approach; (2) a design approach to combine the different methods into one causal matrix. As each method follows a different mathematical approach, this was done by introducing exponential fitting parameters into each method. The combined causal matrix is in a subsequential step used to generate a root cause priority list to decide which process variable is most likely to be the root cause of the found causal dependencies; (3) an existing method based on the cross-correlation function has been extended by using permutation tests as significance test; and (4) a new visualization method for the representation of causal matrices was developed. This visualization allows a better comparison of the results coming from the different methods, since all causal matrices can be represented in a single graphic.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Granger, C.W.J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* **1969**, *37*, 424–38.
2. Horch, A. *Condition Monitoring of Control Loops*; Trita-S3-REG, Royal Institute of Technology: Stockholm, Sweden, 2000.
3. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.
4. Murphy, K.P. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD Thesis, University of California, Berkeley, CA, USA, 2002.
5. Eaton, D.; Murphy, K. Belief net structure learning from uncertain interventions. *J. Mach. Learn. Res.* **2007**, *1*, 1–48.
6. Bauer, M. *Data-driven Methods for Process Analysis*. PhD Thesis, University College London, London, UK, 2005.
7. Kühnert, C.; Gröll, L.; Heizmann, M.; Mikut, R. Methoden zur datengetriebenen Formulierung und Visualisierung von Kausalitätshypothesen. *Automatisierungstechnik* **2012**, *60*, 630–640.
8. Kuehnert, C. *Data-driven Methods for Fault Localization in Process Technology*; Karlsruher Schriften zur Anthropomatik, KIT Scientific Publishing: Karlsruhe, Germany, 2013.
9. Verma, T.; Pearl, J. A theory of inferred causation. In Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning, Cambridge, MA, USA, 22–25 April 1991.
10. Kühnert, C.; Bernard, T.; Frey, C.W. Causal Structure Learning in Process Engineering Using Bayes Nets and Soft Interventions. In Proceedings of the IEEE International Conference on Industrial Informatics, Caparica, Lisbon, Portugal, 26–29 July 2011; pp. 69–74.
11. Johansson, R. *System Modeling and Identification*; Information and System Sciences Series 1; Prentice Hall: Englewood cliffs, NJ, USA, 1993.
12. Gelper, S.; Croux, C. On the Construction of the European Economic Sentiment Indicator. *Oxf. Bull. Econ. Stat.* **2010**, *72*, 47–62.
13. Saunders, A. The Short-Run Causal Relationship between U.K. Interest Rates, Share Prices and Dividend Yields. *Scott. J. Polit. Econ.* **1979**, *26*, 61–71.
14. Xiaoyan, M.; Kewei, C.; Rui, L.; Xiaotong, W.; Li, Y.; Xia, W. Application of Granger causality analysis to effective connectivity of the default-mode network. In Proceedings of the 2010 IEEE/ICME International Conference on Complex Medical Engineering (CME), Gold Coast, QLD, Australia, 13–15 July 2010; pp. 156–160.
15. Yang, W.; Luo, Q. Modeling Protein-Signaling Networks with Granger Causality Test. In *Frontiers in Computational and Systems Biology*; Computational Biology, Series 15; Springer: London, UK, 2010; pp. 249–257.
16. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.

17. Durbin, J.; Watson, G.S. Testing for serial correlation in least squares regression. *Biometrika* **1950**, *37*, 409–428.
18. Seth, A.K. A MATLAB toolbox for Granger causal connectivity analysis. *J. Neurosci. Methods* **2010**, *186*, 262–273.
19. Chavez, M.; Martinerie, J.; Le Van Quyen, M. Statistical assessment of nonlinear causality: Application to epileptic EEG signals. *J. Neurosci. Methods* **2003**, *124*, 113–128.
20. Staniek, M.; Lehnertz, K. Symbolic transfer entropy: Inferring directionality. *Biomed. Tech.* **2009**, *54*, 323–328.
21. Marschinski, R.; Kantz, H. Analysing the information flow between financial time series. *Eur. Phys. J. B Condens. Matter* **2002**, *30*, 275–281.
22. Vapnik, V.N. *Estimation of Dependencies Based on Empirical Data*; Springer: New York, NY, USA, 1982.
23. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222.
24. Wu, F.; Yin, S.; Karimi, H. Fault Detection and Diagnosis in Process Data Using Support Vector Machines. *J. Appl. Math.* **2014**, *2014*, doi:org/10.1155/2014/732104.
25. Shi, F.; Chen, J.; Xu, Y.; Karimi, H. Optimization of Biodiesel Injection Parameters Based on Support Vector Machine. *Math. Probl. Eng.* **2013**, *2013*, doi:org/10.1155/2013/893084.
26. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, 1st ed.; Cambridge University Press: Cambridge, MA, USA, 2000.
27. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2001.
28. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313.
29. Rakotomamonjy, A. Analysis of SVM regression bounds for variable ranking. *Neurocomputing* **2007**, *70*, 1489–1501.
30. Tenny, M.; Rawling, J. Closed-loop Behavior of Nonlinear Model Predictive Control. *Tech. Rep.* **2002**, *50*, 2142–2154.