The International Journal of Advanced Manufacturing Technology manuscript No.
(will be inserted by the editor)

Input Variable Selection for Model-Based Production Control and Optimisation

Miha Glavan · Dejan Gradišar · Maja Atanasijević-Kunc · Stanko Strmčnik · Gašper Mušič

Received: date / Accepted: date

Abstract For model-based production control and optimisation it is crucial to properly identify those input variables that have the strongest influence on production performance. This way, production operators can focus only on the relevant variables, and production control problems can be reduced. In order to identify previously unknown relationships among the production variables, hidden knowledge in historical production data needs to be explored. In the article, two decisive steps are considered. First, an input variable selection methodology, typically applied for selecting model regressors, is applied. Next, the appropriateness of the selected inputs and their manipulative strength is validated by an operating-space-based controllability analysis. To use the most appropriate input variable selection approach, different input selection methodologies are compared with synthetic datasets. Moreover, a case study of Tennessee Eastman process is applied to demonstrate a complete input variable selection procedure for model-based production control and optimisation.

Keywords model-based production control · holistic production control · input variable selection · controllability

Miha Glavan · Dejan Gradišar · Stanko Strmčnik
Jožef Stefan Institute
Jamova 39
1000 Ljubljana, Slovenia
E-mail: miha.glavan@ijs.si

Maja Atanasijević-Kunc · Gašper Mušič
Faculty of Electrical Engineering, University of Ljubljana
Tržaška 25
1000 Ljubljana, Slovenia

1 Introduction

Over the last decades, different approaches have been considered in order to establish a modern, flexible and cost effective industrial environment.

The process engineering community have focused increasingly on plant-wide control, where structural and strategic decisions involved in control system design are integrated for a complete production plant [23]. The original control problem is hierarchically decomposed and heuristic logic is developed to keep process variability, and therefore the operational plant objectives, within acceptable limits [7, 23, 46].

Another approach uses closed-loop control techniques at the production control level. A widely adopted solution in industry is the use of so-called Real-Time Optimisation (RTO) [9, 20, 36]. Engell [9] defines RTO as a model-based upper level control system that is operated in a closed loop and provides set-points to the lower-level control systems in order to maintain the process operation as close as possible to the economic optimum despite disturbances and other process changes.

To establish such centralised closed-loop control and optimisation a first-principle model of production is required. As Engell [9] has stated, steady-state models are available for many processes nowadays, as they are used extensively in the process design phase. However, there are still many production processes for which physical models are not available, and due to the overall complexity of modern plants, a considerable effort is required to adequately formulate the dynamics of a process.

On the other hand, with the rapid development of IT technology, a vast amount of information about current production plants is available. Often massive databases of historical plant data are recorded, but a lack of solutions of how to effectively exploit this recorded data is noticeable. Recorded data could represent a solid foundation for developing production models while employing model identification and data-mining methods. With a production model at hand, concepts for production control and optimisation can be applied in a similar manner as for process control (e.g., Intelligence-based Manufacturing approach [28], Holistic Production Control - HPC approach [15, 54, 55]).

The main challenge of data-based production control and optimisation strategies is how to derive an appropriate production model. The model has to include enough details of the production process to reflect the dominant dynamics for production control. The model should be relatively simple in comparison to models used at the process control level, yet because of the overall complexity and the limitations of process-testing, this task is extremely complex.

One of the key steps of production model derivation is to identify appropriate model inputs and outputs. Outputs are typically selected by the production managers in accordance with the overall objective of production process. A more complex task is to select appropriate model inputs. One reason is that, as Smits et al [44] point out, "in practice many times one has to select the relevant inputs from a possibly large set of candidate inputs". And the other reason is that one has to find a compromise between two different goals. The first one is to obtain a simple model with the best prediction ability, and the second one is to obtain an appropriate model for optimisation and control purposes.

The first goal can be addressed by employing data-mining approaches, where the most relevant inputs need to be selected from a large set of candidate inputs. Relationships within the available data need to be discovered in order to identify suitable predictors of the model output [30]. Such input variable selection (IVS) methods select the most informative inputs but this does not guarantee that the selected set of variables also have the strongest manipulative effects.

The second goal could be pursued by the application of the classical control system design approach. The main inputs are typically selected on the basis of different selection criteria [52], i.e.: accessability, state controllability and state observability, IO controllability, efficiency of manipulation and estimation, robust stability and robust performance. However, such an approach is based on the detailed analysis of the controlled process, which is hardly possible in a production environment with a huge number of interdependent variables, and where first principle models are not available. Also, to make it useful for plantwide systems, where only empirical models based on process data are available, some extensions are required. For this scope, it would be useful to extend the operating-space-based controllability analysis presented by Georgakis et al [14].

In this article a new input variable selection approach appropriate for model-based production control and optimisation is suggested. The approach is based solely on the analysis of historical process data and combines both of the aforementioned approaches. The novelty of approach is characterised by IVS methodology used as a prefilter step, and a newly proposed data-based controllability analysis which is applied to gain some extra knowledge about the appropriateness of the selected inputs for control tasks.

The paper is organised as follows. In Section 2, the case-study is introduced and problem addressed in the article is exposed. In Section 3, a review of the IVS methods is presented. Moreover, synthetic datasets are used to compare the performance of some IVS methods and to select the most useful ones. A short overview of the IO controllability measures is presented in Section 4. A special focus is given to the space-based methodology, which is extended in order to be successfully applied to model-based production control and optimisation. Concluding remarks are given in Section 5.

2 Problem illustration by a case study

The problem of selecting the most relevant manipulating variables is common to different model-based production control and optimisation concepts. To illustrate the problem more practically, a problem definition will be given in the frame of the Holistic Production Control (HPC) of the well known Tennessee Eastman (TE) benchmark process [8]. Although this article is focused on a specific problem, the presented approach can still be viewed more generally and can be easily extended to other model-based production control and optimisation concepts.

The idea of HPC concept was first introduced by Zorzut et al [54, 55]. The concept was further developed by Glavan et al [15], where main design steps were discussed and demonstrated on the TE case study. In these articles, the selection of the most important manipulative variables was recognised as one of the crucial HPC design problems. Though, it was not yet adequately addressed there. The success of the HPC realisation significantly depends on the reasonable compromise between the reduction of the optimisation problem and consideration of the proper set of manipulative variables.

2.1 Brief introduction of the Holistic Production Control concept

The HPC concept [15, 54, 55] is based on the assumption that for certain classes of production it is possible to control and optimise the production process with the aid of a simplified empirical model. The HPC concept is schematically shown in Fig. 1. The *production process block* covers the overall production process, including the stabilising process control level. The input variables of production comprise reference values for process control loops ($\mathbf{K_u}$) and/or other manipulative variables not used within stabilisation loops, and measured disturbances ($\mathbf{d}$). In order to reduce the burden of high dimensional production data and to extract only the current business-related status of the process, production Performance Indicators (pPI) are considered. The idea of performance measures is widely applied to assess the objectives within a production process [3, 13, 35]. Production performance indicators ($\mathbf{K}$) are calculated on-line from the process measurements and aggregate different production variables ($\mathbf{y}$).

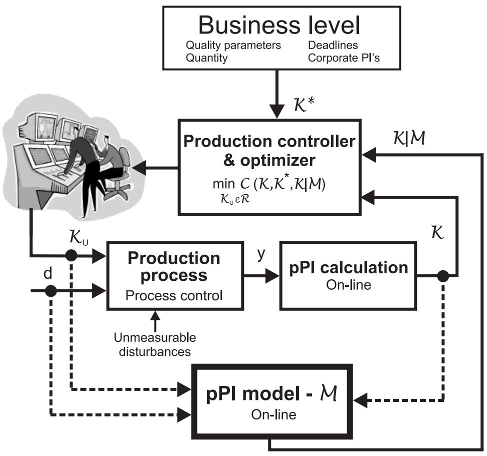


Fig. 1: Concept of holistic production control.

Model-based control and optimisation of the pPIs is considered within the HPC concept to achieve business-level goals. An appropriate model describing the behavior of the process projected on pPIs is required (the *pPI model*) and it is expected to be derived from the historical process data. Based on current pPIs values $\mathbf{K}$, the predicted model outputs $\mathbf{K}|\mathbf{M}$ and the reference values $\mathbf{K}^*$ (i.e. planned business goals), the *Production controller & Optimiser* determines the appropriate manipulative values $\mathbf{K_u}$ and thus supports the production manager. Such control of pPIs would represent a decision support system able to help the manager to determine the appropriate corrections of process inputs in order to realise demands from the business control level and to optimise the production process.

2.2 Tennessee Eastman benchmark process

The TE benchmark process was introduced by Downs and Vogel [8] as a model of a real chemical production process. The model represents a test problem for researchers to experiment with different control-related solutions. As shown in Fig. 2, the process consists of five major units: a chemical reactor, a product condenser, a vapor-liquid separator, a product stripper and a recycle compressor. Four reactants (A, C, D, E) and an inert component (B) enter the process, where four exothermic, irreversible reactions result in two products (G, H) and one byproduct (F). The process products leave the process through stream 11, where they are separated in a downstream refining section. The production process has 41 measured variables ($\mathbf{y}$) and 12 different manipulative variables ($\mathbf{u}$).

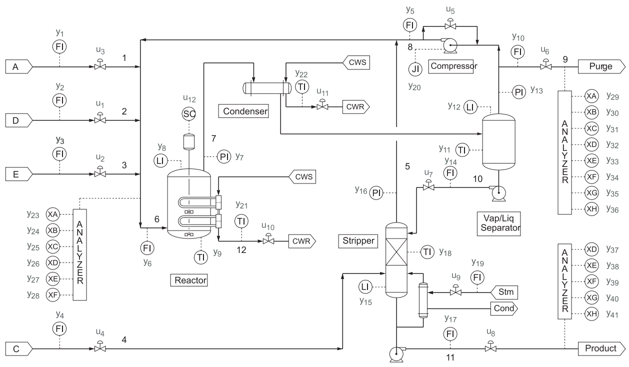


Fig. 2: Production scheme of the Tennessee Eastman process.

A specific combination of the production rate and/or the product mix are usually demanded by the market or some capacity limitations. Therefore, six typical operational modes (see Tab. 1) are defined.

Table 1: List of the Tennessee Eastman production modes defined by Downs and Vogel [8].

mode	G/H mass ratio	Production rate (stream 11)
1	50/50	$14076 kg/h$ ($22.9 m^3/h$)
2	10/90	$14077 kg/h$ ($22.8 m^3/h$)
3	90/10	$11111 kg/h$ ($18.1 m^3/h$)
4	50/50	maximum
5	10/90	maximum
6	90/10	maximum

2.3 Definition of pPIs

The first step to derive a production model for HPC is to define the pPIs. Selected pPIs should be defined in accordance with the production's economic objectives, where production efficiency is aggregated from a vast amount of available process measurements.

As the first pPI, an estimation of the production *Cost* is selected. The formulation of the cost function was introduced by the authors of the TE model [8]. The costs are calculated from the process measurements in units of $\$/h$ and are formulated as shown in (1). The first row of the equation evaluates the costs of the compressor work and steam expenses, while the other rows evaluate the loss of the components leaving the process with the product and purge.

$$\begin{aligned} Cost =&0.0318 \cdot y_{19} + 0.0536 \cdot y_{20} + \\ &0.0921 \cdot y_{17} \left[22.06 \cdot y_{37} + 14.56 \cdot y_{38} + 17.89 \cdot y_{39}\right] + \\ &0.04479 \cdot y_{10} \big[2.209 \cdot y_{29} + 6.177 \cdot y_{31} + 22.06 \cdot y_{32} + \\ &14.56 \cdot y_{33} + 17.89 \cdot y_{34} + 30.44 \cdot y_{35} + 22.94 \cdot y_{36}\big] \end{aligned} \tag{1}$$

Next, we want to express the productivity of the process. The definition of this PI is quite straightforward, as the quantity of product leaving the process is directly measured ($Production = y_{17}$). Since the product quality can be viewed as a desired mass ratio between the two final products (G and H), an indicator for the process quality can also be directly derived from the process measurements ($Quality = y_{40}$).

2.4 Manipulative variables

The TE process is a highly unstable system and without low-level process control it exceeds the process safety limits and automatically shuts down within an hour. We used the system that was stabilised with the low-level control presented in the work of Larsson et al [24], where nine outputs are controlled with cascade loops.

A few changes to the original low-level control were needed, in order to realise HPC concept [15]. As *Production* and *Quality* were already controlled, the *production rate* and *Product ratio* loops were removed and two new manipulative variables were defined (i.e., F_p and r_2). In this way the control of the *Quality* and *Production* was intentionally moved out from the process level. However, they will be controlled at the production level, where the HPC is realised. The TE process with a modified low-level control has 9 manipulative variables, which represent setpoint adjustments to the low level control loops as shown in Tab. 2.

Table 2: Process manipulative variables.

notation	Controlled variable setpoints
F_p	Production rate index
R_2	Striper level
R_3	Separator level
R_4	Reactor level
R_5	Reactor pressure
R_7	% C in purge
R_8	Recycle rate
R_9	Reactor temperature
r_2	D/E feed rates

2.5 Problem statement

The presented case study is a relatively simple example of the model-based production control and optimisation problem. On the one hand there are three main business objectives expressed as pPIs (*Cost*, *Production* and *Quality*). On the other hand, there are many different manipulative variables (see Tab. 2) which should be modified to influence the observed pPIs.

To simplify the pPI model structure and to reduce the optimisation problem, only the manipulative inputs with the greatest impact on the selected pPIs have to be determined. Due to lack of information about the internal relations of the production variables, there are no clear answers, which manipulative inputs should be selected. Moreover, in a real production scenario it is reasonable to expect an even broader set of candidate variables.

The following sections suggest how to extract the valuable information from the historical production data in order to solve the presented variable selection problem.

3 Input variable selection methods

Progress made in sensor technology and data management allows us to gather data sets of ever increasing sizes. Therefore, the most natural way would be to employ data-mining methods (e.g. [6, 39]). Data-mining methodology to identify relevant manipulative variables is known as Input Variable Selection (IVS) and is widely adopted in the field of empirical modelling.

3.1 Literature overview

IVS represents an important step in the model identification procedure. A useful subset of original inputs should be selected with consideration given to the identified relationship within the available data. As Guyon and Elisseeff [16] have pointed out, the main benefits of the prior reduction of inputs are improvement of prediction performance, better understanding of the system and a more cost effective model (faster training and predicting).

If the predictors of the model output are underspecified, selected variables do not fully describe the observed behavior. On the other hand if the input set consists of irrelevant or redundant input variables, with little or no predictive power, the size of the model is increased, data processing time is increased, and insignificant information is added. It is obvious that with an appropriate input variable selection the model is simplified, model accuracy is enhanced and the curse of dimensionality is eliminated.

As presented in the reviews of the IVS approaches [1, 16, 21, 30], selection algorithms can be broadly grouped into three classes: *filter*, *wrapper* or *embedded* algorithms.

The *filter* approach separates the IVS step and the learning of the final prediction model [1]. Filters are independent of the applied modelling methodology, as irrelevant inputs are filtered out before the final model training begins. Such input analysis is mainly based on the statistical tests, properties of the functions and on the use of a simple model.

On the other hand, the *wrapper* methodology employs the prediction of a given learning machine to assess the relative usefulness of variable subsets [16]. Here, input selection is treated as a means to optimise the model structure, where either all or a subset of the possible input sets are compared. The input set that yields the optimal generalisation performance of the calibrated model is selected.

Embedded methods directly incorporate variable selection as a part of the model training process. During training, the irrelevant and redundant input variables are progressively removed (e.g., regularisation, pruning). Compared to the wrapper methods, the computational complexity is reduced, as the retraining of a predictor for every investigated subset can be avoided and only a single model is trained.

Wrapper and filter methods share four main steps [19, 30]: generation of candidate subsets, subset evaluation and selection, stopping criteria and final model validation. An optimal solution can be reached by evaluating all the possible subsets. The number of possible subsets selected from a set of d potential variables is $(2^d - 1)$. As the exhaustive search is computationally infeasible in high-dimensional problems, computationally more efficient input selection strategies are required. Heuristic search or stepwise algorithms (forward selection, backward elimination or their combination) are widely applied to decrease the computational burden. In the next step, every candidate subset is evaluated against some criterion to measure their informativeness and when the stopping criteria is met, the tested input subset is selected.

It should be noted that for systems with dynamical characteristics, the IVS problem is additionally augmented by the selection of the proper time-delayed regressors of the inputs. Usually the maximum anticipated order of the system is defined in advance, and appropriate regressors are included in the IVS procedure. The complexity of the IVS procedure is therefore enlarged, as the number of potential inputs can quickly increase.

3.2 Input variable selection for model-based production control and optimisation

In this article, we will focus on some of the IVS methods from the literature, which can be employed to rank the candidate inputs. The drawback of such an approach is that the k-most relevant variables do not strictly yield an optimal model [30]. However, for closed-loop control and optimisation, priority can be given to the evaluation of the manipulative effects of the tested inputs, rather than to find the final model with optimal generalisation capabilities.

The following paragraphs provide a brief overview of the methods outlined in the literature, which are tested and compared in this article.

One of the most widely used IVS methods is input variable ranking based on the linear *Pearson's correlation coefficient* (Corr.) [30]. To exclude redundant inputs a *partial correlation measure* (Part. corr.) was proposed, as it is able to indicate the relationship between two variables while excluding the effect of other variables [30]. A forward selection approach is adopted for partial correlation to exclude inputs correlated to already selected input set.

Distance correlation (dCorr), also called Brownian distance covariance, was introduced by Székely et al [49] as a true statistical dependence measure between two random variables in arbitrary dimension. Unlike the Pearson's correlation criterion, where only linear relationships between two variables are considered, a distance correlation indicates true independence.

Mutual information (MI) dependence measure takes into account the probabilistic distribution of variables. MI is based on information theory and the notion of Shannon entropy [40] and is sensitive also to nonlinear statistical dependencies. In this article a solution from TIM toolbox [38] was used, where k-nearest neighbours statistics is used [22]. To measure the information between two observations that is not contained in a third one, a *Partial mutual information* (PMI) was proposed. Again TIM toolbox [38] was used, where k-th nearest neighbour statistics has been generalised to directly estimate a PMI as shown in [12].

Gamma test (GT) employs k-th nearest neighbours measures to provide an estimate of the model's output variance that cannot be accounted for by a smooth data model. The method was introduced by Stefánsson et al [45] and later formally mathematically justified by Evans and Jones [10].

The idea of using the statistical *Analysis of Variance* (ANOVA) to select appropriate regressors in dynamical systems was intensively studied by Lind and Ljung, e.g. [26]. The N-way ANOVA method from Statistical toolbox of Matlab [29] was used in this article.

The *Variable Importance in Projection* (PLS VIP) measure is often applied to assess how important the candidate variables are for the projections, in order to find latent variables in *Partial Least Squares* (PLS) regression [5]. PLS regression was evaluated with the NIPALS algorithm and the optimal number of latent variables was determined by k-fold cross validation. VIP scores were then applied to rank all the inputs.

Non-negative Garrote (NNGarr.) introduced by Breiman [2], where a regularised solution of least square estimate is sought. The method tends to eliminate some of the variables, while parameters of others are reduced. In this article, k-fold cross validation has been applied to determine the optimal design parameter and the regression parameters of the final model have been used to rank the predicted importance of the inputs.

Tibshirani [50] has introduced a method called *Least Absolute Shrinkage and Selection Operator* (LASSO). Again, a regularised solution to the least square problem is sought. Free design parameter has been determined with a k-fold cross validation and again the regression parameters are used to rank the inputs.

Another considered method is *input sensitivity analysis of trained neural model* (SA). Here, a weight decay regularisation is adopted [18]. This way, redundant weights are decayed during model training and irrelevant inputs are progressively suppressed. When the model is fully trained, a simple input sensitivity analysis of the neural model [33] is applied to determine the influence of each individual input on the model's prediction.

Li and Peng [25] have presented a pre-filter method to distinguish the most influential inputs for nonlinear regression modelling. To decrease the computational complexity of training, the nonlinear-in-the-parameter model was substituted with the polynomial non-linear regression model which is *linear-in-the-parameters* (LIP). To identify only the significant nonlinear input terms an iterative algorithm was introduced [25] for explicitly computing the contribution of each candidate to the final cost function.

3.3 Comparison of the IVS methods on the data sets with known attributes

In this section, the input ranking methods presented in 3.2 are validated on the synthetic datasets taken from the IVS literature (see Tab. 3). The aim of the comparison is to find the most appropriate methods for selecting inputs when designing model-based production control and optimisation. Synthetic problems are used, as the underlaying structure of the model is known and therefore the IVS results can be evaluated.

3.3.1 Data sets with known attributes

Properties of the data sets are given in Tab. 3 and the underlaying model structures are shown in Eq. (2–11).

$$y = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + e \tag{2}$$

$$y = \frac{\sin(\sqrt{x_1^2 + x_2^2})}{\sqrt{x_1^2 + x_2^2}} + e \tag{3}$$

$$y(k) = 0.3y(k-1) - 0.6y(k-4) - 0.5y(k-9) + e(k) \tag{4}$$

$$y(k) = 2e^{-0.1y(k-1)^2}y(k-1) - e^{-0.1y(k-1)^2}y(k-2) + e(k) \tag{5}$$

$$y(k) = \begin{cases} -2y(k-1) + e(k) & \text{if } y(k-1) < 0 \\ 0.4y(k-1) + e(k) & \text{if } y(k-1) \geq 0 \end{cases} \tag{6}$$

$$y(k) = \begin{cases} -0.5y(k-6) + 0.5y(k-10) + 0.1e(k) & \text{if } y(k-6) \leq 0 \\ 0.8y(k-10) + 0.1e(k) & \text{if } y(k-6) > 0 \end{cases} \tag{7}$$

$$\begin{aligned} y(k) = {} & 0.3y(k-1) + 0.6y(k-2) + 0.6\sin(\pi x(k-1)) + \\ & 0.3\sin(3\pi x(k-1)) + 0.1\sin(5\pi x(k-1)) + e(k) \end{aligned} \tag{8}$$

$$y(k) = \sin(2.3x(k)x(k-2)) + e^{x(k)} - x(k-2)^{5/2} + y(k-1) \tag{9}$$

$$\begin{aligned} y_0(k) &= \tfrac{-2x(k-2)x(k-3)}{y(k-1)} + e^{(-x(k-1)y(k-1))} - x(k-3)^2 \\ y(k) &= y_0(k) + e(k) \end{aligned} \tag{10}$$

$$y(k) = \frac{y(k-1)y(k-2)(y(k-1)+2.5)}{1 + y(k-1)^2 + y(k-2)^2} + x(k-1) + e(k) \tag{11}$$

The regression vector, where all considered variables are included is defined as:

$$\begin{aligned} \mathbf{X} = [& y(k-1) \dots y(k-n_a), x_1(k) \dots x_1(k-n_b+1), \\ & \dots, x_{n_x}(k) \dots x_{n_x}(k-n_b+1)] \end{aligned} \tag{12}$$

where n_a and n_b denote the number of delayed inputs and outputs included in the regression vector and n_x refers to the number of all potential inputs.

The first set of examples (Eq. 2–3) are static functions, the next four examples (Eq. 4–7) are time-series models, while the last set of problems (Eq. 8–11) illustrate dynamical systems with exogenous inputs.

Table 3: Properties of the experiments

id	eq.	samples	n_x	n_a	n_b	source
1	(2)	500	15	0	1	[11]
2	(3)	500	15	0	1	[31]
3	(4)	400	0	15	0	[11, 41]
4	(5)	3000	0	8	0	[4, 27]
5	(6)	3000	0	8	0	[4, 27]
6	(7)	400	0	15	0	[41]
7	(8)	1000	1	5	5	[32, 47]
8	(9)	300	1	4	4	[42]
9	(10)	300	1	5	5	[42]
10	(11)	1000	1	5	5	[32, 47]

All inputs $[x_1 \dots x_{n_x}]$ and additive noise e are initialised randomly with uniform (experiments 1 and 7) or gaussian distribution – $N(0,1)$ (other experiments). Apart from experiment 9, where the noise distribution is $N(0, 0.1|y_0|)$ and output y is limited to $[-1, 3]$. Prior to any input selection, all data were standardised, where each variable was divided by its standard deviation after mean centering.

3.3.2 Evaluation of the IVS methods

For each problem 50 different datasets were generated, and for each dataset the IVS procedure was employed. To enable direct comparison of the methods, IVS scores for each method have been scaled to achieve a unit sum of the scores for all candidate inputs.

Some of the tested methods implement the iterative forward selection strategy, where in each iteration the variable with the highest score is selected. This way the decisive scores do not strictly descend as the selection progresses. Thus, to properly rank the importance of the inputs and to preserve the order of the selection, in this case the following correction of the scores is employed:

$$K_n = \begin{cases} \frac{1}{\frac{1}{K_{n-1}} + \frac{1}{S_n}} & \text{if } S_n \neq 0 \vee K_{n-1} \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Index n indicates the consecutive selection round and is limited to the number of the input variables. With (13) the maximal input score in n-th selection round (S_n) is corrected, that the equivalent modified score (K_n) is less than the modified score of the variable selected in the previous step (K_{n-1}).

To evaluate the performance of different IVS methods, a confusion matrix based measure has been applied, as presented in [5]. The adopted confusion matrix is shown in Tab. 4, where a is the number of correctly classified irrelevant inputs, b is the number of incorrectly classified irrelevant inputs; c and d similarly refer to relevant inputs. Next, we define sensitivity as the proportion of selected relevant predictors among relevant predictors ($\mathit{Sensitivity} = \frac{d}{c+d}$) and specificity as the proportion of unselected irrelevant predictors among irrelevant predictors ($\mathit{Specificity} = \frac{a}{a+b}$). The geometric mean of sensitivity and specificity defines the adopted performance measure [5]:

$$G = \sqrt{\mathit{Sensitivity} \cdot \mathit{Specificity}} \tag{14}$$

The value of G ranges from 0 to 1, where a G close to 1 indicates that most predictors are classified correctly.

The disadvantage of such a performance measure is the assumption that the number of true inputs is known in advance. This is almost never true in real scenarios, where it is often hard to resolve when to stop with the selection of important variables. Typical stopping criteria varies by the applied IVS algorithm. It is common to simply select k-most important variables; to define the threshold for the obtained score; to use cross-validation methods; or to simply compare the differences among the scores.

As this article only focuses on the ranking capabilities of the methods and no specific stopping criteria are considered, we introduce another performance measure to point out how the methods are able to disregard the irrelevant inputs. In this criterion the sum of the IVS scores for the irrelevant inputs is compared to the total sum of the scores for all inputs:

$$H = (1 - \sum_{i \in \{\text{irr. inputs}\}} S_i / \sum_{i \in \{\text{all inputs}\}} S_i) \tag{15}$$

where S_i refers to the score for the i-th input. Values of H close to 1 indicate distinctive results, where the method was able to suppress the influence of irrelevant inputs.

3.3.3 Comments on the results

An evaluation of the variable selection results for the tested synthetic problems is presented in Tab. 5 and 6 for performance measure of G and H, respectively.

In the first static problem (2) the detection of inputs x_3, x_4 and also x_5, which contribute the most to the overall function dynamics, can be distinguished using any method. All except the LASSO method had problems to recognise inputs x_1 and x_2, which are clearly close to the noise threshold. On the other hand, the second static problem is a highly non-linear function, with a small signal-to-noise ratio. Only the *LIP* method and *dCorr* were able to clearly detect both correct inputs

Table 4: Confusion matrix of the classes, where a, b, c and d indicate the number of elements in each class.

		Predicted classes	
		Irrelevant predictor	Relevant predictor
True classes	Irrelevant predictor	a	b
	Relevant predictor	c	d

undoubtedly, while other methods struggled to point out any of the inputs.

Most of the tested methods performed well in the time-series problems. In general, the methods, which were able to distinguish collinearity among the candidate variables (forward selection methods, regularisation methods), have clearly pointed out the true time delayed regressors. Similar conclusions can also be observed in the dynamical experiments with exogenous inputs, where the detection of collinearity among the timed delayed inputs and outputs is crucial.

The superiority of the regularisation methods is determined when indicator H is considered (see Tab. 6). These methods tend to completely remove the influence of the remaining inputs, therefore the considered performance indicator clearly favors their results. The same goes for the *LIP* method, which successfully neglects the irrelevant inputs. On the other hand, from the results of some methods it was hard to distinguish the exact threshold between important variables and those with no impact (e.g., *GT*, *ANOVA*).

On the basis of the presented IVS experiments, no particular method is superior for all of the problems. Some methods gave more accurate and unambiguous results for some experiments, while other methods gave better results for other experiments. Therefore, variable selection should be based on the consideration of different methods and the important inputs should be selected, when one notices consistent results among the methods. In the presented case study we will therefore consider the results of the methods which gave the most clear and exact results: *Part. corr.*, *pMI*, *LIP*, *SA*, *NNGarr.* and *LASSO*.

3.4 Case study: selection of input variables

As already pointed out, the idea behind the input reduction is to minimise the model-based control and optimisation problem on one hand, and to preserve enough manipulative power with respect to the controlled pPIs, on the other hand.

IVS methods, which were identified as the most suitable in the previous section, were employed on the characteristic production dataset of the TE case study. At the end, the inputs that have been shown to be the most influential by the concurrence of different methodologies are selected.

To remove the scores corresponding to irrelevant inputs, a simple stopping criteria was applied. The additional random input (*probe*) was added to the IVS procedure. Scores smaller than the *probe*'s score were excluded form the selection and the average of the remaining scores and their distributions over different IVS methods were examined in order to select the most influential inputs for each of the observed pPIs, as represented in Fig. 3.

Note, that delayed regressors of the inputs are included in the IVS procedure to consider the process dynamics ($n_b = 3$). But, as only the most informative input is sought, the average score for each of the inputs is displayed.

While considering Figures 3b and 3c, the inputs F_p and r_2 should be selected as the inputs with the strongest influence on *Production* and *Quality*. On the other hand, selection of the inputs which have a strong influence on *Cost* has shown to be a more demanding task, as none of the inputs clearly outperformed the others (see Fig. 3a). Besides the already selected inputs F_p and r_2, which also have a strong influence on the *Cost* pPI, the inputs R_4, R_7, R_9 are additionally selected as they achieved the highest scores from the IVS methods.

4 IO controllability analysis for model-based production control and optimisation

The main purpose of the IVS methods is to identify predictors of the model output, which yields improvement of the model prediction. Irrelevant, noisy and redundant inputs are discarded from the model and only the most informative (explanatory) variables are taken into account.

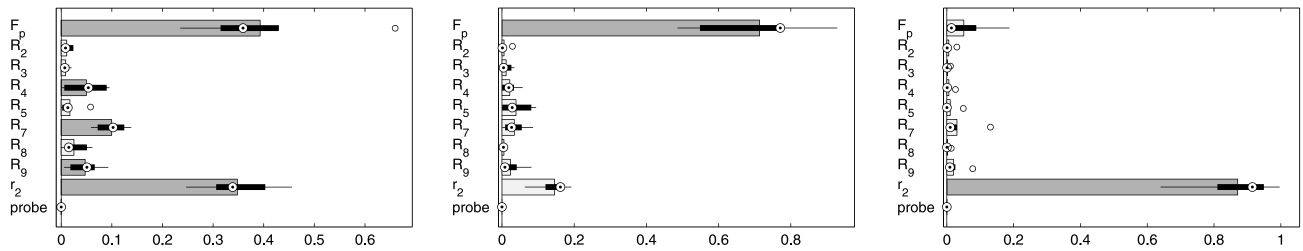
If we use such an approach for the identification of the most influential control inputs, an assumption is made that the most informative manipulative variables have the strongest influence on the controlled variables. But from the view of control-design this is often not enough, as there is no guarantee that the selected set of inputs is sufficient to cover the desired output range. Strict physical limitations of the variable could severely

Table 5: Evaluation of the tested IVS methods – correctness of the results (performance measure G).

	IVS methods											
Ex. id	Corr.	Part. corr.	dCorr	MI	PMI	GT	ANOVA	PLS VIP	NNGarr.	LASSO	SA	LIP
1	0.7454	0.7546	0.7551	0.6593	0.6743	0.7450	0.7455	0.7486	0.7918	1.0000	0.7643	0.7073
2	0.1766	0.1902	0.6861	0.3525	0.2638	0.4725	0.1902	0.1766	0.4748	0.2917	0.2030	0.7299
3	0.6697	1.0000	0.6799	0.6697	0.7395	0.9869	1.0000	0.7504	1.0000	0.9782	0.9913	1.0000
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7022	1.0000	1.0000	0.8936	1.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0.8589	0.9551	0.9166	0.8910	0.8589	0.9166	0.9551	0.8333	0.9166	0.7948	0.7820	0.9423
7	0.7559	1.0000	0.7559	0.7506	0.9463	1.0000	1.0000	0.9756	1.0000	1.0000	0.9951	1.0000
8	0.4472	0.7842	0.4472	0.4472	0.4472	0.7772	0.8166	0.4472	0.4472	0.5944	0.5311	1.0000
9	0.5228	0.5445	0.6963	0.7435	0.7304	0.8313	0.5399	0.5226	0.5624	0.5729	0.7559	0.6876
10	0.5576	0.6478	0.5844	0.6946	0.6985	0.8165	0.6590	0.5576	0.6648	0.9605	0.9556	0.7067
$\sum/10$	0.6734	0.7877	0.7522	0.7208	0.7359	0.8546	0.7906	0.6714	0.7858	0.8193	0.7872	0.8774

Table 6: Evaluation of the tested IVS methods – clearness of the results (performance measure H).

	IVS methods											
Ex. id	Corr.	Part. corr.	dCorr	MI	PMI	GT	ANOVA	PLS VIP	NNGarr.	LASSO	SA	LIP
1	0.7408	0.8777	0.6249	0.8892	0.8839	0.5358	0.4604	0.7005	0.9233	1.0000	0.9964	0.8739
2	0.1326	0.1258	0.1786	0.2204	0.1882	0.2478	0.1348	0.1330	0.0892	0.1013	0.1473	0.3179
3	0.3588	0.8652	0.3355	0.4691	0.6900	0.4543	0.3411	0.3445	0.9227	0.9548	0.8947	0.9528
4	0.4181	0.8851	0.4213	0.5202	0.8658	0.3342	0.3195	0.3924	0.8983	0.9664	0.8893	0.9999
5	0.4716	0.5724	0.5154	0.8219	0.8109	0.3674	0.1956	0.4315	0.5322	1.0000	0.8430	0.7482
6	0.3936	0.8219	0.3009	0.5224	0.7906	0.3290	0.2269	0.3482	0.8434	1.0000	0.7745	0.9630
7	0.4165	0.9459	0.4186	0.5392	0.8809	0.3251	0.4746	0.4469	0.9761	0.9993	0.9794	0.9996
8	0.2617	0.8414	0.2730	0.2937	0.8007	0.4943	0.4610	0.2776	0.6293	0.8371	0.9411	0.9995
9	0.5127	0.6466	0.5388	0.5419	0.7650	0.5185	0.4224	0.5043	0.6885	0.9269	0.7238	0.7008
10	0.6348	0.7613	0.5766	0.7896	0.8310	0.5256	0.3518	0.5933	0.7791	0.9399	0.8291	0.9355
$\sum/10$	0.4341	0.7343	0.4184	0.5608	0.7507	0.4132	0.3388	0.4172	0.7282	0.8726	0.8019	0.8491



(a) *Cost* (b) *Production* (c) *Quality*

Fig. 3: Variable selection with consideration to different IVS methodologies.

limit its manipulative effects on production and also disturbances can additionally limit the accessible output range.

The controllability of the process should be considered to verify the adequate manipulation power of the selected set of inputs and to examine the expected control performance of the system.

4.1 IO controllability measures

Input-output controllability is defined as the process ability to achieve an acceptable control performance [43]. That is, being able to keep the outputs within specified bounds or displacements from their reference values in spite of the unknown but bounded variations (e.g., disturbances and plant changes), using the available inputs and measurements of outputs and disturbances.

IO controllability measures can generally be classified into the linear-model-based and nonlinear-model-based approaches [53]. Linear measures (like relative gain array (RGA), the conditional number and singular value decomposition based measures) were commonly applied in the literature (see reviews [43, 52]). But in the last decade, nonlinear methods are also being intensively studied. *Analytical* nonlinear methods focus on the search for the unstable zero dynamics, which limit the perfect control [53]. In contrast to the linear case, where right half plane (RHP) transmission zeros cause an unstable inverse, unstable zero dynamics represent the limitations for nonlinear models. Another class of nonlinear controllability methods are *optimisation* based methods, which can analyse controllability more broadly. With these methodologies, process design and controllability analysis can be integrated into one optimisation problem. The problem is formulated as a mathematical superstructure capable of respecting a dynamic operability, model uncertainty and synthesise optimal controllers at the same time [53].

But in regards to the model-based production control and optimisation procedure, only the IO controllability of the selected subset of manipulative variables should be confirmed. In this scope, an interesting operating-space-based method was introduced by Vinson and Georgakis [51], where steady-state controllability is measured directly as the geometrical comparison between the achievable and desired output spaces. Their work was also further extended on dynamical analysis and for non-square systems [14]. The main idea of their method is presented in the following section and an extension of this concept for model-based control and optimisation is presented in Section 4.3.

4.2 Operating-space-based controllability measure

The operating-space-based controllability measure, presented by Vinson and Georgakis [51], enables one to inspect the ability of the process to reach the full range of the desired output values, within the limited range of process inputs and under the presence of the expected process disturbances.

The available input space (AIS) is defined as the set of values that input variables can take. The space is limited by the operating range of input variables, due to process design and equipment limitations. The idea of the method is to find a geometric representation of the achievable operating region, named as achievable output space (AOS). The relative coverage of the AOS and expected operating region (desired output space – DOS), is defined as the output controllability index (OCI). Similarly, the effect of the expected disturbance space on the AOS can be considered.

To extend the controllability methodologies to plantwide systems, achievable production output space (APOS) was defined by Subramanian and Georgakis [48]. With this, the dimension of the AOS is reduced, since APOS is connected only with the operating region of the exogenous (external) output variables related to production and discards endogenous (internal) outputs.

Operating-space-based controllability analysis can help to identify the design limitations of the plant. All active constraints of process states and inputs can be observed from the boundary points of the APOS space. Different control structures can be tested and the analysis can help to distinguish between several plant designs. Moreover, potential disturbances can be included in the analysis to evaluate how much of the APOS will still be achievable under active disturbances.

4.3 Operating-space-based controllability analysis for model-based production control and optimisation

Controllability measures typically employ detailed first principles models to analyse and compare different process structures. But, while designing model-based production control and optimisation, our knowledge of the process dynamics is usually severely limited. An empirical model needs to be identified from historical process data, as often only such process knowledge is available in order to study the controllability properties of the system.

Let us look at the controllability problem from the stand point of the Model Predictive Control (MPC), which is a natural candidate to be applied at the production control level. When we are using empirical models for MPC, we should be aware of the model extrapolation limits and that the quality of the results directly depends on the quality of the applied empirical model. Similar problems and limitations arise when controllability analysis is applied to the empirical models. Therefore, we should note that one should base decisive conclusions on the results gained inside the model's validity. Limited exploration of the output space should be performed, where the inputs are used within the boundaries of the model-training signal.

Operating-space-based controllability analysis can also be viewed as the prior exploration of the complete process knowledge, incorporated in the model, to be used for MPC. An additional benefit of such analysis is to gain some prior information on what could be expected from the predictive control with the tested model and the selected combination of the inputs and outputs.

Due to the nature of controlled or optimised variables ($\mathbf{y}$), controllability analysis should be studied separately for setpoint-controlled ($\mathbf{y_S}$) and minimisation-oriented (optimised) variables ($\mathbf{y_M}$):

$$\begin{aligned}\mathbf{y} &= [\mathbf{y_S}, \mathbf{y_M}]^T \\ \mathbf{y_S} &= [y_{S1}, y_{S2}, \ldots, y_{Sn}]^T \\ \mathbf{y_M} &= [y_{M1}, y_{M2}, \ldots, y_{Mm}]^T\end{aligned} \tag{16}$$

where n and m refers to the number of the setpoint-controlled and minimisation-oriented variables, respectively.

4.3.1 Controllability analysis of setpoint-controlled variables

To evaluate the Achievable Output Space (AOS) of the setpoint-controlled variables ($\mathbf{y_S}$) we need to examine the achievable borders of each $\mathbf{y_S}$, under the limitations of the available input space and disturbances. Finally, AOS can then be constructed as the space divided by these achievable borders, as it is shown in Fig. 4 for two setpoint-controlled variables.

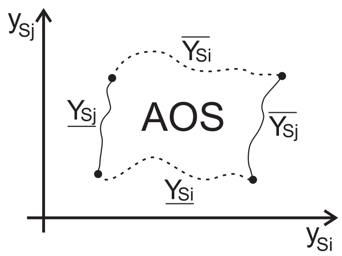


Fig. 4: Achievable output space of two setpoint-controlled variables (y_{Si} and y_{Sj}), defined as the intersection of the their achievable lower and upper boundaries.

The lower ($\underline{\mathbf{Y}_{Si}}$) and upper ($\overline{\mathbf{Y}_{Si}}$) boundaries of the achievable output space for i-th setpoint-controlled variable can be obtained from the optimisation problem solved for each variable (for $\forall i \in \{1, \ldots, n\}$):

$$\begin{aligned}&\underline{\mathbf{Y}_{Si}} = \min_{\mathbf{u}} y_{Si}^{ss} \text{ and } \overline{\mathbf{Y}_{Si}} = \max_{\mathbf{u}} y_{Si}^{ss} \\ &\text{subject to } c(\mathbf{u}, \mathbf{d}) \leq 0 \\ &\text{and for } \forall y_{Sj}^{ss} \in \tilde{\mathbf{O}}_j \ (j \neq i) \text{ and } \forall \mathbf{d} \in \mathbf{D}\end{aligned} \tag{17}$$

Here y_{Si}^{ss} indicates a steady-state response of the empirical production model $y = g(\mathbf{u}, \mathbf{d})$ for the i-th controlled variable, where $\mathbf{u}$, $\mathbf{y}$, $\mathbf{d}$ represent the input, output and disturbance vectors, respectively. The optimisation problem (17) is solved over the expected output space $\tilde{\mathbf{O}}_j$ for the remaining setpoint-controlled variables and over the expected disturbance space $\mathbf{D}$. The optimisation problem is also subjected to the process constraints for manipulative variables and disturbances $c(\mathbf{u}, \mathbf{d}) \leq 0$. Note, that these constraints should reflect the physical process limitations. But, since the presented analysis is limited by the validity of the analysed model, the limitations of the model training data should be applied. From the control point of view, it is expected that training data should cover the area of interest, where the model is expected to operate and, if a model with broader validity can be acquired, the controllability analysis should be repeated.

The AOS, defined as the space limited by the lower and upper boundaries of the different set-point controlled variables, can be compared to our expectations represented by the desired output space (DOS). Each output range is compared to the range that is expected to be covered within the production control and optimisation. From this comparison it can be concluded whether or not the physical boundaries of a particular input should be extended or, if this is impossible, a different set of inputs should be considered.

4.3.2 Controllability analysis of minimisation-oriented variables

In the next step, the controllability analysis for the minimisation-oriented variables is investigated. The definition of the desired production space as the space between two boundaries has no purpose for these variables. Only the lowest achievable value should be identified. The optimisation problem (18) has to be solved over the achievable output space of setpoint-controlled variables $\mathbf{O}$ and the disturbance space $\mathbf{D}$.

$$\begin{aligned}&\forall i \in \{1,\ldots,m\}: \quad \underline{\mathbf{Y}_{Mi}} = \min_{\mathbf{u}} y_{Mi}^{ss} \\ &\text{subject to } c(\mathbf{u},\mathbf{d}) \leq 0 \\ &\text{and for } \forall y_S^{ss} \in \mathbf{O} \text{ in } \forall \mathbf{d} \in \mathbf{D}\end{aligned} \tag{18}$$

The achievable minimisation boundary for each minimisation-oriented variable is evaluated, while manipulating the process with the tested set of inputs. The achievable lower boundary $\underline{\mathbf{Y}_{Mi}}$ should answer whether the input constraints are sufficient or whether they significantly limit the minimisation of the tested variable.

An additional test should be performed to point out what contribution each manipulative variable has on the minimisation of the controlled variable. For each tested input, the lower boundary of the minimisation-oriented variable is compared with the achievable lower boundary, when the tested input is excluded from the process manipulation. The tested input is excluded from the control and set constant to its optimal value, which is defined as the mean of its optimal values from the minimisation problem (18). This way the manipulative contribution of the input u_k to the output variable minimisation can be defined as:

$$\delta\underline{\mathbf{Y}_{Mi}^k} = \underline{\mathbf{Y}_{Mi}}_{(u_k=\text{fixed})} - \underline{\mathbf{Y}_{Mi}} \tag{19}$$

With $\delta\underline{\mathbf{Y}_{Mi}^k}$ we can evaluate if the contribution of k-th input has enough of a manipulation effect on the minimisation of the tested controlled variable. If this is not possible, another input or combination of inputs should be considered to minimise this controlled variable.

4.4 Case study: controllability analysis of the selected inputs

From the IVS analysis of the presented case study (see Section 3.4) it follows, that inputs F_p, R_4, R_7 R_9 and r_2 should be selected as the most influential manipulative variables to control considered pPIs ($Cost$, $Production$, $Quality$). In order to additionally verify the manipulation effects of these input variables, the operating-space-based controllability analysis can be performed.

Neural network regression is applied to identify a pPI model from the historical process data. To assist the identification process, the Neural Network System Identification NNSYSID toolbox for Matlab [34] was used in our case. A dynamical NARX model with a feed-forward neural structure was identified for each output. Each network was trained on the representative training data set with the Levenberg-Marquadt algorithm. Different neural models with alternate topologies were trained and all of the identified neural networks were pruned with the OBS algorithm [17] in order to maximise the generalisation capabilities of the models. In the next step, the identified networks were tested using input data sets not used in the training process. The network with the best performance on the 5-step ahead prediction and simulation response was selected for each pPI. The applied modelling procedure was represented more broadly by Glavan et al [15].

Since the presented controllability analysis is limited only to a steady state, long term responses of the neural model are observed. As the constraints of the process we have applied the boundaries of the training signals in order to limit the extrapolation of the model knowledge in the controllability analysis.

4.4.1 Setpoint-controlled pPIs

The results of the optimisation problem defined with the eq. (17) defines the boundaries of the AOS space for setpoint-controlled pPIs – *Production* and *Quality*. The resulting AOS for setpoint-controlled pPIs is shown in Fig. 5.

From the figure it can be concluded that the predicted achievable output space for setpoint-controlled variables meets our expectations as they include all of the expected production modes. Therefore, the selected inputs and their limitations are recognised as sufficient to cover setpoint-controlled output space of our interest.

Note, that TE process modes 4–6 do not exactly specify the *Production* rates and only specify maximum achievable rates. Therefore, the optimal values published by Ricker [37] are adopted. Ricker's values result from the TE process optimisation problem and represent the optimal steady-state conditions for all of the production rates.

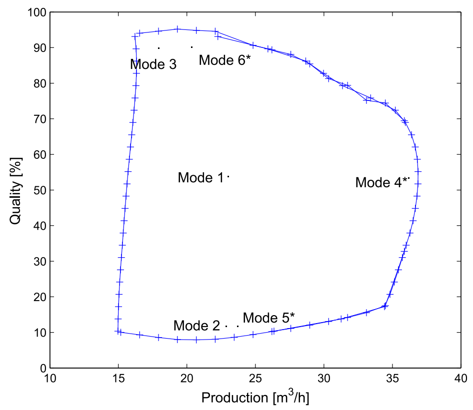


Fig. 5: Achievable production space for setpoint-controlled pPIs (modes indicated with * correspond to the mode values published in [37]).

4.4.2 Minimisation-oriented pPIs

In the next step, controllability analysis is performed for the minimisation-oriented pPI – *Cost*. The optimisation problem (18) rendered the lower boundary of the achievable space shown in Fig. 6. Note that the quality definition follows the literature, where it is defined as the percentage ratio of the two products leaving the process. The indicator's name can be misleading, as this is not the quality in the traditional sense. This can be confusing when interpreting the obtained results, as a higher quality percentage yields lower cost values.

Additional information can be gained from the analysis of the optimal manipulative variables ($\mathbf{u}_{opt}$), calculated while exploring the lower *Cost* boundary. Percentages of the region where a manipulative variable has saturated are represented in Tab. 7. It can be observed that saturation of the variable R_4 severely limits the optimal performance and it is expected that if this lower boundary could be extended over a wider area, better *Cost* results could be achieved. But if we would like to extend the analysis, the constraints of the manipulative variables should be extended and a new model valid on

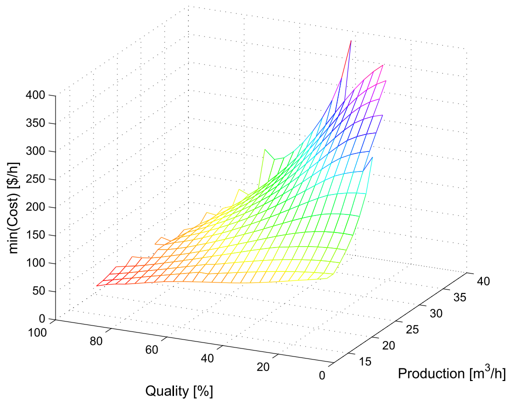


Fig. 6: Lower boundary for minimisation-oriented pPI - *Cost*.

a wider area should be applied. For the presented case study it can be presumed that the applied constraints of the training signal are widespread enough to cover the majority of the working space of the production process. Consequently, it can be concluded that input R_4 is not suitable for *Cost* minimisation.

In the final step, the contribution of each manipulative input on *Cost* minimisation is verified. The inputs are sequentially fixed to the mean of their optimal values, calculated from the *Cost* minimisation step, and again the minimal *Cost* space is examined with the manipulation of the remaining inputs. The difference of the optimal achievable spaces, where all inputs are used and when one input is set constant, are depicted in Figure 7. Also here, the input R_4 is recognised as having almost no *Cost* minimisation effect. On the other hand, the manipulative effects of the inputs R_7 and R_9 are expected to be more significant, especially in the area of low *Quality* and high *Production* values.

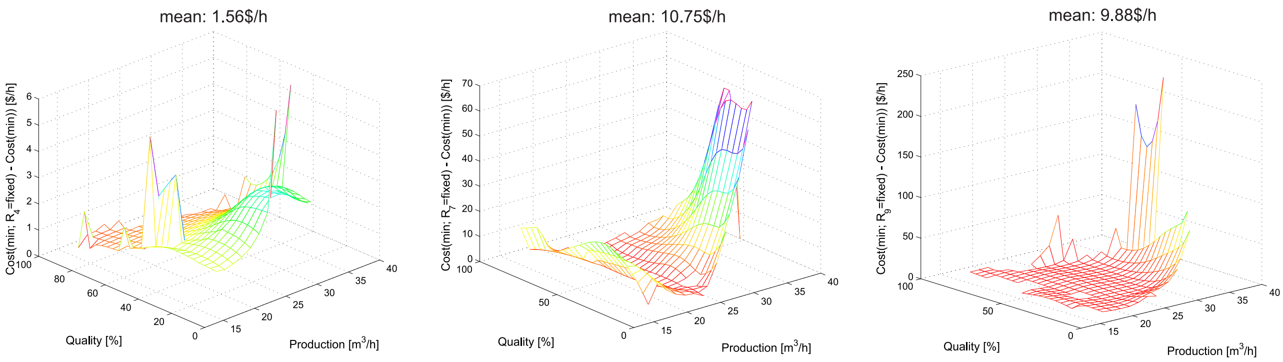
Note, that results for inputs F_p and r_2 are not shown, as excluding these inputs from the control caused the *Production*-*Quality* output space to be unreachable. This is due their strong manipulation effect on the set-point controlled pPIs, which could already be acknowledged from the IVS routine (see Fig. 3).

4.4.3 Comments on the final input subset selection

Based on the controllability analysis, the tested inputs were seen to have sufficient manipulative effects to cover the desired outputs space of setpoint-controlled pPIs. Using the IVS methodology, several inputs were recognised as inputs with a strong influence on the *Cost* pPI. But for minimisation-oriented pPIs, it is crucial that these inputs have an important effect on the direction of pPIs minimisation. The controllability analysis of the

Table 7: Saturation of the manipulative variables for minimised *Costs*.

	F_p	R_4	R_7	R_9	r_2
lower bound	9.17 %	94.80 %	1.22 %	3.96 %	0.31 %
upper bound	5.50 %	2.14 %	5.81 %	24.46 %	1.83 %



(a) R_4 is fixed: $\underline{\delta \mathbf{Y}^{R_4}_{mCost}}$ (b) R_7 is fixed: $\underline{\delta \mathbf{Y}^{R_7}_{mCost}}$ (c) R_9 is fixed: $\underline{\delta \mathbf{Y}^{R_9}_{mCost}}$

Fig. 7: *Cost* minimisation with fixed inputs.

minimisation-oriented pPI showed that the manipulative variable R_4, previously identified as influential, has no direct influence on *Cost* minimisation over different production modes. As the manipulative effects of R_4 on *Cost* minimisation is completely limited by underlying constraints, it should be fixed at its optimal value – the lower boundary. But if it is expected that its value changes during the process operation, R_4 should be included in the production model as a measurable disturbance, since this input still incorporates valuable information for *Cost* prediction.

4.4.4 Evaluation of the controllability results

The predicted output space of setpoint-controlled pPIs can be further evaluated using the results published in the literature. Ricker [37] determined optimal steady-state conditions for all six operating modes. Moreover, the maximal *Production* output space was also explored with the controllability analysis published by Subramanian and Georgakis [48]. If these results are compared to the output space obtained by the proposed approach (see Fig. 5), it can be seen that the output space is not correctly identified in the area around modes 4–6. The points marked with * in fig. 5 represent the maximal production rates for these modes (as defined in Tab. 1). Therefore, the exact AOS borderline should intersect the points for modes 4-6 in Fig. 5. One of the reasons for this inconsistency is the fact that a simplified pPI model was used, whereby internal production states are neglected. And as was shown by Subramanian and Georgakis [48], and also Ricker [37], the true physical limitations of the production rate for these modes represent the internal capacity limits of the feed streams E and D. However, our primary focus in the presented analysis is the selection of the manipulative variables and the evaluation of their limits. Therefore, saturation of the internal states and the required changes to the process design are not our concern at this point, rather only the appropriateness of the inputs is being evaluated.

To evaluate the resulting achievable output space of *Cost*, a comparison with the optimal steady states reported by Ricker [37] is made in Tab. 8. The predictions of the pPI model, manipulated with only five input variables, are close to the optimal values published by Ricker [37]. From this comparison it can be acknowledged that the predictions obtained on the basis of the simple pPI model have a similar tendency to the results from the more detailed analysis, which is crucial when the lower *Cost* boundary is being analysed.

Ricker [37] also showed that lower R_4 directly yields smaller overall *Costs* over different production modes. This finding directly supports our conclusions that the manipulative variable R_4 is inappropriate for *Cost* control.

5 Conclusion

To reduce control complexity of model-based production control and optimisation, only inputs with the

Table 8: Optimal *Cost* values for all six operating modes.

Mode	*Cost* [$/h] (Ricker)	*Cost* [$/h] (HPC contr.)
1	114.31	116.02
2	181.09	174.01
3	43.93	45.12
4	243.92	224.49
5	194.59	190.46
6	49.23	51.03

strongest influence on production performance should be employed. Due to the extensive number of input candidates and a limited insight into the process, it is vital to point out the most relevant parameters, the importance of which could be unknown to process managers. A detailed analysis of historical process data is needed to extract knowledge about the relationships between process variables. For this purpose a combination of the IVS methodology and data-based controllability analysis has been presented.

None of the tested IVS methods stood out for all of the applied experiments. Therefore, it is suggested that a final selection of the important manipulative variables needs to be made with due consideration of the results from different IVS methods. A case study of the TE process has demonstrated how to combine scores from different IVS methods. By taking into account the scores from different methods, the IVS results could be validated, while the concurrence among the methods gave us an additional assurance to identify a specific input as relevant.

But even if the selected inputs have a strong influence on the controlled production variable, it is still possible that the physical limitations of the manipulative variable could severely limit control performance. Operating-space-based controllability analysis of the selected inputs is discussed in order to gain some additional information of their manipulative importance. In contrast to more complex controllability measures, which employ detailed first-principles models, the only option for our problem is to use an empirical model. Consequentially, the results should be considered as a supplementary insight into the process when considering the selection of manipulative variables or even as the verification or examination of the model's knowledge, which will be applied for model-based control.

The practical use of such controllability analysis on the TE case study has shown that it is possible to gain some useful insight into the process. An informative input was recognised as an input with almost no manipulative effect when its working boundaries were considered. On the other hand, the analysis is severely limited by the model's validity and, if the controllability analysis is further extended to more realistic problems, the optimisation complexity is expected to be enlarged. Moreover, a comparison of the spaces for several controlled variables should be performed with the controllability index as shown in Subramanian and Georgakis [48], where geometrical intersections of the spaces are considered.

Acknowledgements The work has been supported by the Competence Centre for Advanced Control Technologies which is partly financed by the Republic of Slovenia, Ministry of Higher Education, Science and Technology and European Union (EU) - European Regional Development Fund; and Slovenian Research Agency through the grant L2-2338.

References

1. Blum AL, Langley P (1997) Selection of relevant features and examples in machine learning. Artif Intell 97:245–271
2. Breiman L (1995) Better subset regression using the nonnegative garrote. Technometrics 37(4):373–384
3. Chen CC, Cheng WY (2007) Customer-focused and product-line-based manufacturing performance measurement. Int J Adv Manuf Tech 34:1236–1245
4. Chen R, Liu JS, Tsay RS (1995) Additivity tests for nonlinear autoregression. Biometrika 82(2):pp. 369–383
5. Chong IG, Jun CH (2005) Performance of some variable selection methods when multicollinearity is present. Chemometr Intell Lab 78(1-2):103 – 112
6. Choudhary A, Harding J, Tiwari M (2009) Data mining in manufacturing: a review based on the kind of knowledge. J Intell Manuf 20:501–521
7. Downs JJ, Skogestad S (2011) An industrial and academic perspective on plantwide control. Annu Rev Control 35(1):99–110
8. Downs JJ, Vogel EF (1993) A plant-wide industrial process control problem. Comput Chem Eng 17(3):245–255
9. Engell S (2007) Feedback control for optimal process operation. J Process Contr 17:3:203–219
10. Evans D, Jones AJ (2002) A proof of the gamma test. Proc R Soc Lond (series A) 458:1–41
11. Fernando T, Maier H, Dandy G (2009) Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach. J Hydrol 367:165–176
12. Frenzel S, Pompe B (2007) Partial mutual information for coupling analysis of multivariate time series. Phys Rev Lett 99
13. Fukuda Y, Patzke R (2010) Standardization of key performance indecator for manufacturing execution system. In: Proceedings of SICE Annual Conference 2010, pp 263 –265
14. Georgakis C, Uztürk D, Subramanian S, Vinson DR (2003) On the operability of continuous processes. Control Eng Pract 11(8):859 – 869
15. Glavan M, Gradišar D, Strmčnik S, Mušič G (2013) Production modelling for holistic production control. Simul Model Pract Th 30:1 – 20
16. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182

17. Hassibi B, Stork DG, Com SCR (1993) Second order derivatives for network pruning: Optimal brain surgeon. In: Advances in Neural Information Processing Systems 5, Morgan Kaufmann, pp 164–171
18. Haykin S (1999) Neural Networks: A Comprehensive Foundation, druga edn. Prentice Hall, Upper Saddle River, NJ
19. Hu C, Wan F (2009) Input selection in learning systems: A brief review of some important issues and recent developments. In: IEEE International Conference on Fuzzy Systems, 2009, pp 530 –535
20. Kadam JV, Marquardt W, Schlegel M, Backx OHBT, Brouwer PJ, Dunnebier G, van Hessem D, Tiagounov A, de Wolf S (2003) Towards integrated dynamic real-time optimization and control of industrial processes. In: Grossmann IE, McDonald CM (eds) Proc. FOCAPO 2003, pp 593–596
21. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97(1):273–324
22. Kraskov A, Grassberger P (2008) Information Theory and Statistical Learning, Springer US, chap MIC: Mutual Information based hierarchical Clustering, pp 101 – 123
23. Larsson T, Skogestad S (2000) Plantwide control - A review and a new design procedure. Model Ident Control 21(4):209–240
24. Larsson T, Hestetun K, Hovland E, Skogestad S (2001) Self-Optimizing Control of a Large-Scale Plant: The Tennessee Eastman Process. Ind Eng Chem Res 40(22):4889–4901
25. Li K, Peng JX (2007) Neural input selection-a fast model-based approach. Neurocomputing 70:762–769
26. Lind I, Ljung L (2005) Regressor selection with the analysis of variance method. Automatica 41(4):693–700
27. Lind I, Ljung L (2008) Regressor and structure selection in narx models using a structured anova approach. Automatica 44:383–395
28. Lydon B (2012) Intelligence-based manufacturing: Moving to preventative & proactive manufacturing. In: Pharmaceutical Automation Roundtable (PAR) 2011, Part 6
29. Matlab (2012) Statistics toolbox – anova. `http://www.mathworks.com/help/toolbox/stats/anovan.html`, accessed 30th August 2012
30. May R, Dandy G, Maier H (2011) Review of input variable selection methods for artificial neural networks. In: Suzuki K (ed) Artificial Neural Networks - Methodological Advances and Biomedical Applications, InTech
31. May RJ, Maier HR, Dandy GC, Fernando TG (2008) Non-linear variable selection for artificial neural networks using partial mutual information. Environ Modell Softw 23:1312–1326
32. Narendra K, Parthasarathy K (1990) Identification and control of dynamical systems using neural networks. IEEE Transactions on Neural Networks 1(1):4 –27
33. Niska H, Heikkinen M, Kolehmainen M (2006) Genetic algorithms and sensitivity analysis applied to select inputs of a multi-layer perceptron for the prediction of air pollutant time-series. In: Intelligent Data Engineering and Automated Learning IDEAL 2006, Lecture Notes in Computer Science, vol 4224, Springer, pp 224–231
34. Nørgaard M (2000) Neural network based system identification toolbox. Tech. Report 00-E-891, Department of Automation, Technical University of Denmark
35. Nudurupati S, Bititci U, Kumar V, Chan F (2011) State of the art literature review on performance measurement. Comput Ind Eng 60(2):279 – 290
36. Rawlings J, Amrit R (2009) Optimizing process economic performance using model predictive control. In: Magni L, Raimondo D, Allgöwer F (eds) Nonlinear Model Predictive Control, Lecture Notes in Control and Information Sciences, vol 384, Springer Berlin / Heidelberg, pp 119–138
37. Ricker N (1995) Optimal steady-state operation of the tennessee eastman challenge process. Comput Chem Eng 19(9):949 – 959
38. Rutanen K (2011) Tim 1.2.0. `http://www.cs.tut.fi/~timhome/tim/tim.htm`, accessed 30th August 2012
39. Sadoyan H, Zakarian A, Mohanty P (2006) Data mining algorithm for manufacturing process control. Int J Adv Manuf Tech 28
40. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423 & 623–656
41. Sharma A (2000) Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 – a strategy for system predictor identification. J Hydrol 239(1-4):232 – 239
42. Sindelár R, Babuska R (2004) Input selection for nonlinear regression models. IEEE T Fuzzy Syst 12(5):688–696
43. Skogestad S, Postlethwaite I (2005) Multivariable Feedback Control: Analysis and Design. John Wiley & Sons
44. Smits G, Kordon A, Vladislavleva K, Jordaan E, Kotanchek M (2006) Variable selection in industrial datasets using pareto genetic programming. In: Yu T, Riolo R, Worzel B (eds) Genetic Programming Theory and Practice III, The Kluwer International Series in Engineering and Computer Science, vol 9, Springer, pp 79–92
45. Stefánsson A, Koncar N, Jones AJ (1997) A note on the gamma test. Neural Comput Appl 5(3):131–133
46. Stephanopoulos G, Ng C (2000) Perspectives on the synthesis of plant-wide control structures. J Process Contr 10(23):97 – 111
47. Su M, Rhinehart R (2009) Recursive estimation on spatially rearranged data for dynamic order determination and nonlinear component detection. In: American Control Conference, 2009. ACC '09., pp 4398 –4403
48. Subramanian S, Georgakis C (2005) Methodology for the steady-state operability analysis of plantwide systems. Ind Eng Chem Res 44(20):7770–7786
49. Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and Testing Dependence by Correlation of Distances. Ann Stat 35(6)
50. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc (Series B) 58:267–288
51. Vinson DR, Georgakis C (2000) A new measure of process output controllability. J Process Contr 10(2-3):185 – 194
52. van de Wal M, de Jager A (2001) A review of methods for input/output selection. Automatica 37(4):487–510
53. Yuan Z, Chen B, Zhao J (2011) An overview on controllability analysis of chemical processes. AIChE Journal 57(5):1185–1201
54. Zorzut S, Gradišar D, Jovan V, Mušič G (2009) Use of a procedural model in the design of production control for a polymerization plant. Int J Adv Manuf Tech 44:1051–1062
55. Zorzut S, Jovan V, Gradišar D, Mušič G (2009) Closed-loop control of a polymerisation plant using production performance indicators (PIs). Int J Comput Integ M 22:12:1128–1143