

# Interpretation and Analysis of Input Selection Approaches in Distance Space

Tim Oliver Heinz

Automatic Control, Mechatronics  
Department of Mechanical Engineering  
University of Siegen  
D-57068 Siegen, Germany  
Email: tim.heinz@uni-siegen.de

Oliver Nelles

Automatic Control, Mechatronics  
Department of Mechanical Engineering  
University of Siegen  
D-57068 Siegen, Germany

**Abstract**—In this paper five different model free input selection approaches are summarized. All of them rely on distance measurements of the data, which can be visualized in the here called 'Distance Space'. The five discussed approaches for input selection are interpreted and analyzed in the Distance Space scatter plot. The influence of noise and wrong chosen inputs to the Distance Space is outlined. With the chosen search strategy, the input space is changed. Afterward the best input combination is selected. Several dynamic systems are used to generate data sets, in order to compare the different approaches.

## I. INTRODUCTION

The modeling of static realworld problems is usually involved with a large number of potential model inputs. The identification of *dynamic* systems expands this amount of model inputs by the delayed versions of the physical inputs and outputs. In practice only a few of them are necessary for the representation of the process. So the number of the initial given potential inputs must be reduced. This reduction leads to a higher density of data in the input space and also to fewer parameters of the estimated model (e.g. in case of linear regression). The initial set of inputs can be defined as:

$$\mathcal{X} = \{x_1, x_2, x_3, \dots, x_d\}. \quad (1)$$

Note that for a dynamic model the potential inputs  $x_i$  are delayed inputs or outputs. The set of inputs in  $\mathcal{X}$  is chosen very generously to make sure that every necessary input is included. Out of this set different subsets can be extracted:

$$\mathcal{S}_i \subset \mathcal{X}. \quad (2)$$

The methods for determining the model order can generally be divided into two main groups: The modelbased and the modelfree methods. The modelbased methods predict the output with different subsets. To evaluate the selected input space, there are several criteria which can be used. In [1] the Akaike information criterion (AIC) is proposed, but also the root-mean squared error (MSE) on the training or an independent (validation) data set can be used. After all, the subset with the best evaluation is chosen.

These model-free approaches are only able to detect the order with respect to the overall relationship between input(s)

and output. Recent research indicates that substantial advantages may be obtained from distinguishing between two types of input variables: (i) scheduling or operating variables and (ii) local model or submodel variables. All scheduling type of approaches fall in this category like local model networks of Takagi-Sugeno fuzzy systems. In a fuzzy system context type (i) variables describe the rule premise space and type (ii) variables describe the rule consequents space [2]. Note that both types of variables can be dynamic in nature but typically scheduling or operating variables are chosen statically or with very simple dynamics. This allows for much easier modeling and strongly weakens the curse of dimensionality [3]. Clearly, for these more advanced strategies a model-free approach is not sufficient.

There are several model free criteria which try to determine the right subset. The following five are discussed: The Lipschitz Quotient approach [4], which tries to determine the right model order by approximating the Lipschitz-bounds of the underlying continuous function. The False Nearest Neighbor approach [5] classifies the nearest neighbors in the input by geometric properties of the data. The Proximal Correlation Coefficient approach [6] attempts to determine the model order with autocorrelation. The Gamma Test [7] estimates the variance of the process, which can be used to evaluate a chosen subset. The Delta Test [8], a simplification of the Gamma Test, averages the distances of output pairs defined by the two nearest neighbors in the selected subset.

With an increasing number of possible inputs  $d$ , the number of potential subsets increases with  $2^d - 1$ . To overcome this challenge a search strategy has to be chosen. In [9] several of them are summarized. Most popular are the forward selection or sequential growing and the backward elimination or sequential pruning. These algorithms start at an empty (full) input space and add (take away) the best (worst) of input variables. Other methods alternate between growing and pruning (e.g. plus  $l$ -take away  $r$ , or floating search strategies [9]). This increases the number of evaluated input spaces, which decreases the danger of local optimum. Afterwards the best combination is chosen to be the correct input space.

## II. FIVE APPROACHES FOR INPUT SELECTION

In contrast to a static model, a dynamic time-discrete NARX model depends on time delayed inputs and outputs. In this paper the focus lies on Single-Input Single-Output (SISO) Systems. So lets assume some nonlinear dynamic function:

$$y(k) = f(\vec{x}(k)), \quad (3)$$

$$\vec{x}(k) = [y(k - n_1), y(k - n_2), \dots, y(k - n_{no}), \dots, u(k - m_1), u(k - m_2), \dots, u(k - m_{ni})]. \quad (4)$$

Here  $y(k)$  is the physical output and  $u(k)$  is the physical input of the nonlinear process at time instance  $k$ .  $n_{no}$  and  $m_{ni}$  represent the number of delays for output and input respectively. A common choice is  $n_{j+1} = n_j + 1$ ,  $n_1 = 1$  and  $m_{j+1} = m_j + 1$ ,  $m_1 = d$  where  $dT_0$  is the dead Time of the model with  $T_0$  being the sampling time. For  $n_{no} = 0$  an NFIR model is obtained. For  $n_{no} \geq 1$  an NARX model of order  $\max(n_{no}, m_{ni})$  is the most common choice. However, it is more flexible, but seldom, to drop the restrictions  $n_{j+1} = n_j + 1$  and  $m_{j+1} = m_j + 1$  in order to obtain more flexibility. An extension to multiple physical inputs is straightforward.

### A. Lipschitz Quotient Approach

With the assumption, that the underlying dynamic process is Lipschitz continuous (which applies for a great number of real world processes) the derivative of every secant of the unknown nonlinear function is bounded by some value  $L$ . He and Asada [4] showed that neglecting a necessary input leads to high derivatives of the secant. With the choice of the right subset the bounding value  $L$  can be approximate. Further increase of the input space dimension leads to similar or even higher values of the secant derivative, so that the minimum indicates the right subset.

For a Lipschitz bounded function the following expression holds:

$$L \geq \frac{|f(\vec{x}_1) - f(\vec{x}_2)|}{|\vec{x}_1 - \vec{x}_2|}. \quad (5)$$

The derivative of every secant between the points  $\vec{x}_i$  and  $\vec{x}_j$  can be defined by:

$$q_{i,j} = \frac{|f(\vec{x}_i) - f(\vec{x}_j)|}{|\vec{x}_i - \vec{x}_j|}. \quad (6)$$

The numerator can be rewritten with total differential and the Cauchy-Schwarz inequality:

$$\begin{aligned} q_{i,j} &= \frac{\left| \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_n} dx_n \right|}{\sqrt{(dx_1)^2 + \dots + (dx_n)^2}} \\ &\leq M \cdot \frac{|dx_1 + \dots + dx_n|}{\sqrt{(dx_1)^2 + \dots + (dx_n)^2}} \leq \sqrt{n} \cdot M. \end{aligned} \quad (7)$$

The number of inputs is  $n$  and  $M$  is a bounding value. The difference of the  $n$ -th component of  $x_i$  and  $x_j$  is represented by  $dx_n$ .

Because of (5) the value of  $q_{i,j}$  lies between zero and  $L$ . So the largest values of  $q_{i,j}$  are close to  $L$ . With this information an index is defined based on the  $p$  largest  $q_{i,j}$ . Because of the quotient which defines  $q_{i,j}$  the geometric mean is used to average:

$$q(\vec{x}) = \sqrt[n]{\left( \prod_{l=1}^p q_{i,j}(l) \right)^{\frac{1}{p}}}. \quad (8)$$

High values for  $q(\vec{x})$  indicate a derivative which violates the bounding value  $L$ . These huge  $q(\vec{x})$  only occur at small changes in  $\vec{x}$  while the corresponding output  $y$  shows a relatively big discrepancy. This leads to the assumption that an important input of the system is missing.

It was proposed to add successive delayed inputs and outputs until the value of  $q(\vec{x})$  becomes a minimum or runs into saturation.

### B. False Nearest Neighbor Approach

The proposed approach classifies data points in true or false nearest neighbors (TNN/ FNN). For every data point in the chosen subset the nearest neighbor (NN) can be computed (e.g. with the Euclidean distance). This is done by minimizing the distance  $d_x$ :

$$d_x = |\vec{x}_i - \vec{x}_{i,NN}|_2. \quad (9)$$

After that, the distance of the corresponding output variables  $y(\vec{x}_i)$  and  $y(\vec{x}_{i,NN})$  is computed  $d_y = |y(\vec{x}_i) - y(\vec{x}_{i,NN})|$ . To check if the nearest neighbors in the input are also near in the output the quotient of the distances is computed:

$$\frac{d_y}{d_x} = \frac{|y(\vec{x}_i) - y(\vec{x}_{i,NN})|}{|\vec{x}_i - \vec{x}_{i,NN}|_2}. \quad (10)$$

If an important input is neglected, there will be some huge distances in the output space, while in the input space the distance is small. In this case the quotient in (10) becomes very large. In order to classify these points as FNNs a threshold  $R$  is introduced:

$$\frac{|y(\vec{x}_i) - y(\vec{x}_{i,NN})|}{|\vec{x}_i - \vec{x}_{i,NN}|_2} \leq R. \quad (11)$$

If the expression in (11) is true the neighbor are true neighbors, otherwise they are false.

### C. Gamma Test approach

The Gamma Test uses  $k$ -Nearest-Neighbors ( $k$ NN) in the input space to estimate the variance of the process, with  $1 < k < p$  and  $p \approx 10$ . For each  $k$  the averaged squared distance between  $\vec{x}_i$  and its  $k$ NN  $\vec{x}_{i,kNN}$  is calculated:

$$\delta(k) = \frac{1}{N} \sum_{i=1}^N (\vec{x}_i - \vec{x}_{i,kNN})^2. \quad (12)$$

For each distance between  $\vec{x}_i$  and  $\vec{x}_{i,kNN}$ , the squared output distance is calculated and averaged to get  $\gamma(k)$  (note that

$y(\vec{x}_{i,k\text{NN}})$  and  $y(\vec{x}_i)$  are not necessarily  $k\text{NN}$  in the output space, although  $\vec{x}_i$  and  $\vec{x}_{i,k\text{NN}}$  are  $k\text{NN}$  in the input space).

$$\gamma(k) = \frac{1}{2 \cdot N} \sum_{i=1}^N (y(\vec{x}_i) - y(\vec{x}_{i,k\text{NN}}))^2. \quad (13)$$

The values of  $\bar{\gamma}$  and  $\bar{\delta}$  are used to fit a regression line. On the one hand the vertical intercept of this line represents the variance of the process, on the other hand, the intercept can be interpreted as the squared distance of the  $0\text{th}$  Nearest Neighbor, which is a good measurement for the variance.

While in this paper it is not the task to determine the variance of the process, the Gamma Test can be used to evaluate the subset. This is possible, because to small subsets act like the true subset with high variance (see Sec. III).

#### D. Delta Test approach

The Delta Test is a simplification of the Gamma Test [7]. The Delta Test uses in contrast to the Gamma Test just the first-Nearest-Neighbor (NN) for this approach. Similar to FNN the distance in the output, is calculated by the nearest neighbor in the input. In contrast to FNN the aggregated distance of the output is used to evaluate the chosen subset. Note that  $\delta$  is the same value like  $\gamma(1)$  out of the Gamma Test (Eq. (13)):

$$\delta = \frac{1}{2 \cdot N} \sum_{i=1}^N (y(\vec{x}_i) - y(\vec{x}_{i,\text{NN}}))^2. \quad (14)$$

#### E. Proximal Correlation Coefficient approach

The main idea of the Proximal Correlation Coefficient is based on the relation of distances in the input to the output space. It can be shown, that a function  $h(\vec{x})$  correlated with itself where the input is disturbed by some small quantities  $h^*(\vec{x} + \Delta\vec{x}_0)$ , the correlation will be high:

$$r_{h,h^*} = \frac{1}{N} \sum_{i=1}^N \frac{(h(\vec{x}_i) - \mu_h)(h^*(\vec{x}_i) - \mu_{h^*})}{\sigma_h \sigma_{h^*}} \approx 1. \quad (15)$$

The mean and standard deviation of the function  $f$  is represented by  $\mu_f$  and  $\sigma_f$ . On the other hand, two different functions  $h(\vec{x})$  and  $g(\vec{x})$  are usually uncorrelated:

$$r_{h,g} = \frac{1}{N} \sum_{i=1}^N \frac{(h(\vec{x}_i) - \mu_h)(g(\vec{x}_i) - \mu_g)}{\sigma_h \sigma_g} \approx 0. \quad (16)$$

By reproducing the introduced functions  $f$  and  $f^*$  this can be used to determine the correct subset. Let  $\vec{x}_i$  be a data point in a chosen subset, then the nearest neighbor of  $\vec{x}_i$  can be represented by.

$$\vec{x}_{i,\text{NN}} = \vec{x}_i + \Delta\vec{x}. \quad (17)$$

By sorting the input space by the nearest neighbors (here the Euclidean distance is used) subsequent data points can be described by

$$x_{i+1} = x_i + \Delta\vec{x} \quad (18)$$

so the correlation of the related outputs of the nonlinear function can be used for rating the chosen subset:

$$r_{f,f} = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{(f(\vec{x}_i) - \mu_f)(f(\vec{x}_{i+1}) - \mu_f)}{\sigma_f^2}. \quad (19)$$

A high correlation near to one indicates the chosen subset to be the ‘true’ input combination. In the following section the differences of the introduced approaches are discussed.

### III. A DISCUSSION OF THE APPROACHES

We propose the ‘Distance Space’ to visualize distances in the output over the corresponding distances in the input. This approach is inspired by [7] and can be used to show the strategy of most approaches. In this section the following steps are investigated:

- 1) To introduce the Distance Space a static process is used. The input is defined by uniform distributed input data  $u$  and  $N = 100$  data points. The process is a sigmoid function: The effect of noise and a wrong selected subset is underlined.
- 2) In the next step the undisturbed process with correct input is transformed to the Distance Space and visualized with a scatter plot.
- 3) For the most approaches only the (lower) left corner of the whole Distance Space is important, which leads to a reduction of the whole Distance Space. The influence of noise and wrong chosen subsets will be illustrated.
- 4) Finally the different approaches are visualized in the Distance Space using the wrong subset  $\mathcal{S}_i = \{v\}$ .

It is obvious, that all of the introduced approaches are based on a geometrical approach. Especially the Lipschitz quotient approach, the FNN approach, the Gamma Test and the Delta Test evaluate the distances in the input and output space directly. In the first step the PCC uses also the geometric characteristics of the input, but in the second step the output is used to compare with some theoretic assumptions.

While the distances of the input and output space are used for determination of the correct subset, these two values can be plotted, using a scatter plot. In Fig. 1a) the undisturbed sigmoid function output is plotted over the correct input. In the following the influence of noise and wrongly chosen subsets are investigated. Fig. 1b) outlines the influence of white gaussian noise. For a wrong subset  $\mathcal{S}_i = \{v\}$  where  $v$  is uniform distributed data, Fig. 1c) clarifies, that there is no relation between  $v$  and  $y$ . By expanding the input space to  $\mathcal{S}_i = \{u, v\}$ , Fig. 1d) shows that only  $u$  influences the output. So  $v$  is not necessary to reconstruct the output.

The whole Distance Space for the correct subset without noise is plotted in Fig. 2.

Now the examples shown in Fig. 1 are transformed to the Distance Space. For the transformation only the ten nearest neighbors in the input are used. Because of that qualitatively only the left hand side of the whole Distance Space is plotted. In Fig. 3 the reduced Distance Space is plotted for the examples from Fig. 1.

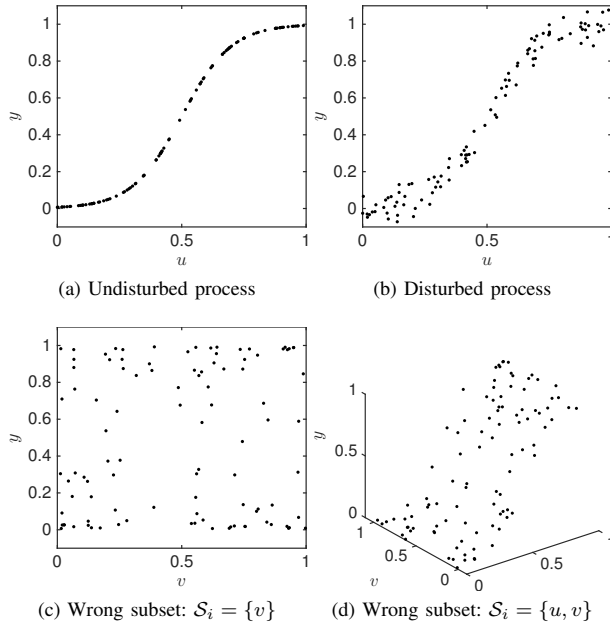


Fig. 1. For representation of the Distance Space a sigmoid function with uniform distributed input data  $N = 100$  is used. Upper left: Undisturbed process. Upper right: Process disturbed with gaussian noise ( $\sigma = 0.05$ ). Lower left: Wrong subset  $\mathcal{S}_i = \{v\}$  without noise. Lower right: Wrong subset  $\mathcal{S}_i = \{u, v\}$  without noise.

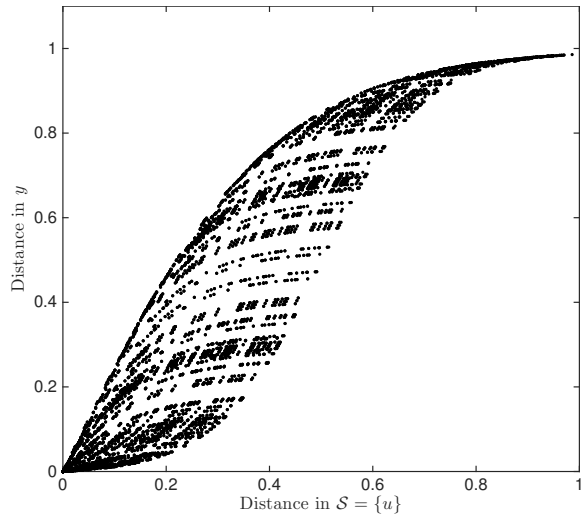


Fig. 2. All distances in the output plotted over the input distances for an undisturbed process with correct input

Figure 3a demonstrates how an ideal process dataset looks like in the Distance Space. A sharp separation line, crossing the origin of the coordinate system, can be observed. Neither Fig. 3b nor Fig. 3c show this clear separation. This indicates, that the process output is noisy, or the wrong/ to small input space is used to represent the output. Only Fig 3d delivers similar results like the correct input without noise.

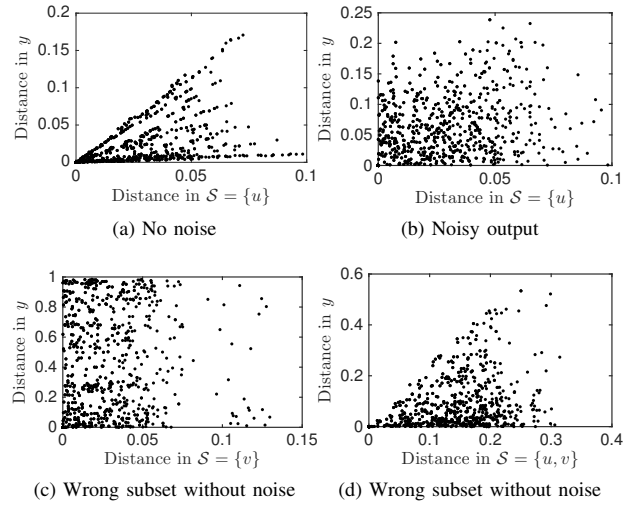


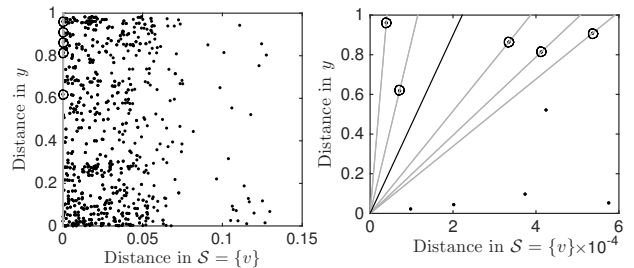
Fig. 3. Distribution of the first ten nearest neighbors in the Distance Space. Upper left: correct subset without noise. Upper right: correct subset with gaussian noise. Lower left: Wrong subset  $\mathcal{S}_i = \{v\}$  without noise. Lower right: Wrong subset  $\mathcal{S}_i = \{u, v\}$  without noise.

This underlines, that it is hard to separate the correct input from an expanded input space.

In the following the five discussed approaches for input selection are interpreted and analyzed in the Distance Space scatter plot. Because of the similarity of the disturbed process with correct input and the undisturbed process with the wrong subset  $\mathcal{S}_i = \{v\}$ , only the wrong subset will be further investigated.

### 1) Lipschitz quotient approach in the Distance Space:

The Lipschitz quotient approach estimates a regression line without intercept in the Distance Space. To calculate the regression line, the points with the highest resulting slopes are selected. These slopes are geometrical averaged to build the quotient  $q$ . This value  $q$  rates the subset. When  $q$  is very large, this indicates a wrong subset, while a low  $q$  refers to the right input space.



(a) View including all  $k$ NN distances (b) Zoomed view to the critical points  
Fig. 4. Representation of the Lipschitz Criterion in the Space of input and output distances

### 2) FNN approach in the Distance Space:

The threshold  $R$  of FNN can be represented by a line in the Distance Space with the slope  $R$  and no intercept

(see Figure 5). The points above this line are classified as false nearest neighbors. The ratio of FNN to TNN is used to rate the subset. For the correct subset the value for FNN should tend to zero.

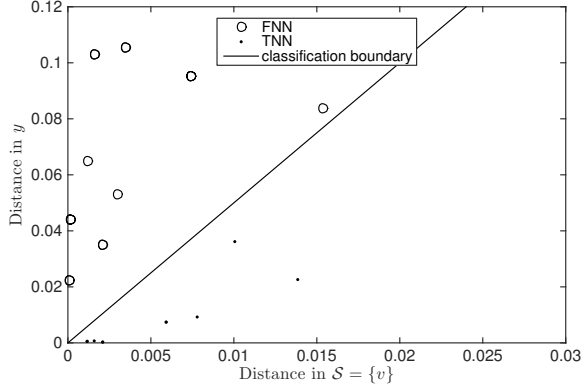


Fig. 5. Classification of FNN in the Distance Space

### 3) Gamma Test in the Distance Space:

Figure 6 illustrates the Gamma Test. The Gamma Test fits a regression line in the Distance Space similar to the Lipschitz quotient approach. In contrast to Lipschitz, the regression line has an intercept and is calculated with the averaged distances in the output corresponding to the 1st to  $k$ th nearest neighbor in the input. Through these resulting points (circles in Fig. 6) in the Distance Space a regression line is fitted. The intercept with the vertical axis of this line represents the variance of the underlying data set. This value can be interpreted as the '0-th Nearest Neighbor'. An incomplete subset shows similar properties in the Distance Space like noise in the output. So a high variance can be interpreted as a wrong subset.

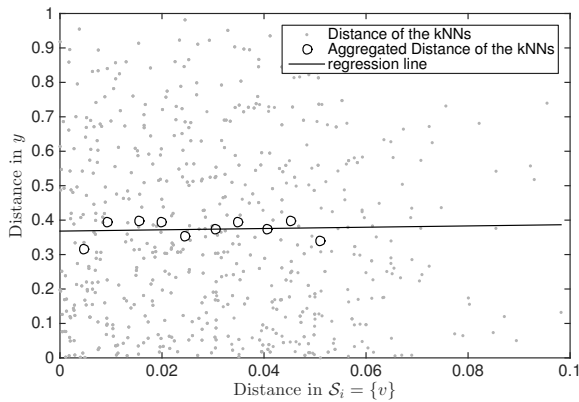


Fig. 6. Data distribution in the scaled  $u$ - $y$  input space

### 4) Delta Test in the Distance Space:

The Delta Test is a simplification of the Gamma Test. While the Gamma Test needs the 1st to  $k$ th nearest

neighbor in the input to determine the variance, the Delta Test approximates the variance just by the 1st nearest neighbor. So the value of the most left circle in Fig 6 represents the value of the Delta Test.

Note that the values for the Gamma Test and Delta Test (see Fig. 6) where linear transformed from the original equations in (12) to (14) to visualize these approaches in the introduced Distance Space.

Here were only the wrong subset  $S_i = \{v\}$  investigated. For a perfect dataset without noise and the correct input, the approaches would bring even better results. For example: The critical points of the Lipschitz quotient approach would lie on the sharp separation in Fig. 3a. The FNN would classify all neighbors as true, with an proper adjusted  $R$ . Even the Gamma Test would deliver an variance close to zero, which is just a little bit smaller than the Delta Test value.

## IV. EXAMPLE PROCESSES

For illustration and comparison, some synthetic dynamic systems are used. Different chosen subsets are evaluated and classified according to the five approaches.

1) Nonlinear dynamic function from [10]:

$$y(k) = (0.6 - 0.1 \cdot a(k-1)) \cdot y(k-1) + a(k-1) \cdot u(k-1) \quad (20)$$

$$\text{with: } a(k) = \frac{0.6 - 0.06 \cdot y(k)}{1 + 0.2 \cdot y(k)}.$$

2) Nonlinear dynamic function from [10]:

$$y(k) = 0.95 \cdot y(k-1) + \sin(0.8 \cdot \pi \cdot u(k-1)) \cdot (1.5 - u(k-1)). \quad (21)$$

3) Nonlinear dynamic function from [4]:

$$y(k) = \frac{u(k-1) + a(k) \cdot (y(k-3) - 1)}{1 + y(k-2)^2 + y(k-3)^2}$$

with:

$$a(k) = y(k-1) \cdot y(k-2) \cdot y(k-3) \cdot u(k-2). \quad (22)$$

4) Nonlinear time series from [6]:

$$y(k) = 3.57 \cdot y(k-1) \cdot (1 - y(k-1)). \quad (23)$$

5) Nonlinear time series from [11]:

$$y(k) = (0.8 - 0.5 \cdot \exp(-(y(k-1))^2)) \cdot y(k-1) - (0.3 + 0.9 \cdot \exp(-(y(k-1))^2)) \cdot y(k-2) + 0.1 \cdot \sin(\pi \cdot y(k-1)). \quad (24)$$

These dynamic systems are excited by a chirp signal.

For a given data size the input signals are equal which would lead to similar results. To modify these signals a small random variation is added to the original input. One input signal is shown in Figure 7. The output of the systems are left undisturbed, to exclude this influence.

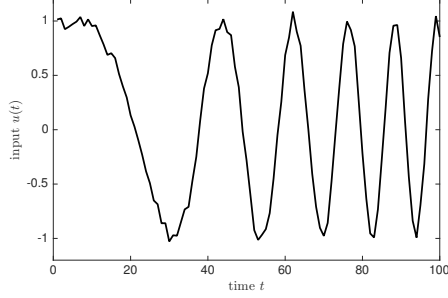


Fig. 7. One of the evaluated input signals for  $N = 100$  data points

While the proposed approaches deliver a value to rate the chosen subset, it is the task of the user to decide on a strategy to generate these subsets. Here we take a forward selection approach, which adds successively the inputs with the best evaluation criterion and the best subset is chosen.

The potential inputs for the function are chosen as  $u(k-1), \dots, u(k-5)$  and  $y(k-1), \dots, y(k-5)$ . To determine the effect of the amount of data 20 different data sizes are used. For each signal length 20 datasets are generated for statistical evaluation.

## V. RESULTS

The results can be separated into two parts: First in Sec. V-A the delayed inputs and outputs were chosen free. Sec. V-B presents the results with an additional restriction: The delayed inputs and outputs have to be chosen successively, which means, that an delayed input  $u(k-n)$  or  $y(k-n)$  can only be chosen, if  $u(k-n+1)$  or  $y(k-n+1)$  was chosen before.

### A. Choosing the inputs free

To rate the results the chosen subset is compared to the right subset. The figures 8a-9b show the fraction of correctly chosen inputs. Only the FNN approach has a tuning parameter which is set to  $R = 0.1$  for all computations.

- 1) For the first function the correct input space is spanned by  $\{u(k-1), y(k-1)\}$ . As we can see, the Delta Test and FNN approach find the right subset in many cases. The median indicates, that also the Lipschitz quotient approach choses the right inputs relatively often but with much higher variance. The PCC and the Gamma Test perform significantly worse.
- 2) The second function posses the identical correct subset like the first function. The results are comparable to the previous ones.
- 3) In Fig. 8c the results for function 3 with an input space spanned by  $\{u(k-1), u(k-2), y(k-1), y(k-2), y(k-3)\}$  are shown. This subset size brings the approaches to their limits. FNN and PCC find most of the right inputs in many cases. The Gamma Test, and its approximation the Gamma Test, are close to PCC and FNN. Here

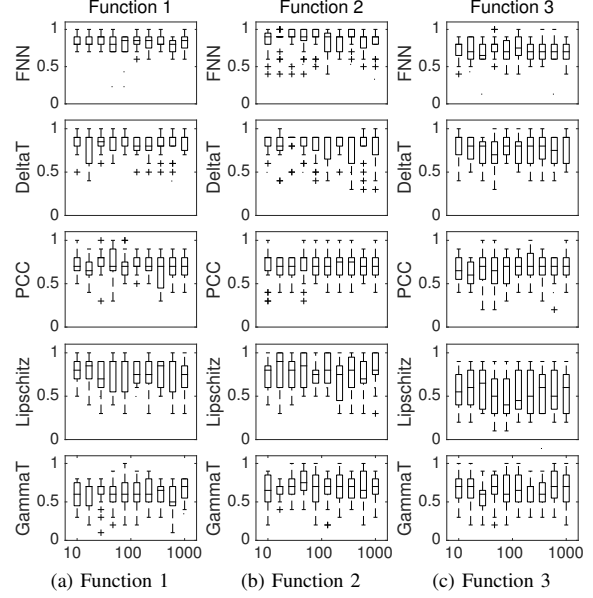


Fig. 8. Results of the input output processes

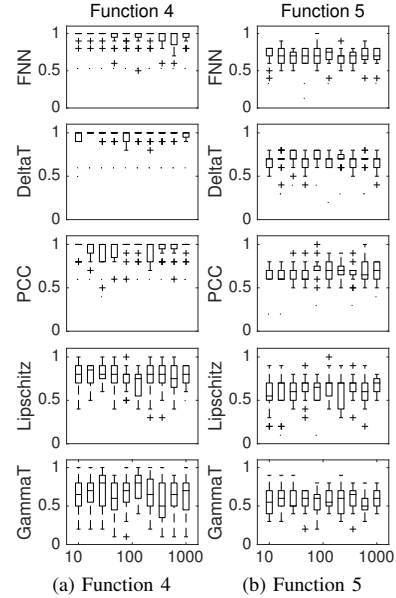


Fig. 9. Results of the time series processes

- 4) In Fig. 9a a time series is reviewed. To deal with the time series function in a fair manner compared to the input-output systems, the approaches are allowed to select delayed inputs, like the previous functions. Here the differences between the approaches are more pronounced. While FNN, Delta Test and even PCC perform well with

correct subset selection ratios close to one (with a few outlier), the Gamma Test and Lipschitz fail in many cases. Note that missing boxes are concentrated on the median value.

- 5) The last time series function 5 with a strong nonlinearity is harder to determine than the other time series function. Here all the approaches perform unsatisfactory. The Lipschitz quotient and the Gamma Test often determine one out of two inputs correct. The other approaches operate slightly better, but also unsuitable.

### B. Choosing the inputs successively

Figure 10 visualizes the results of the successively chosen inputs and outputs according to [4]. Equal to Sec. V-A for each of the 20 different data sizes, 20 variations of the input signal are used, so that each approach-function combination represents 400 evaluations. The color in fig. 10 shows how many of the evaluated data reached the input order (identified by the axes). A dark color indicates a high concentration of results.

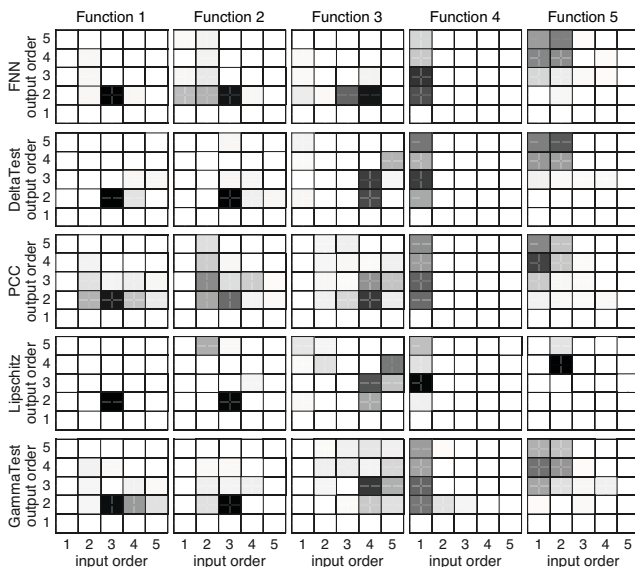


Fig. 10. Results of successively chosen inputs and outputs. A dark shaded color indicates a high concentration of results.

Similar to the free chosen results, it is shown, that the certainty of the estimation decreases with the complexity of the underlying function. By comparing the different approaches it becomes clear, that the PCC delivers uncertain results for each function. The Lipschitz quotient in contrast has a high certainty for each function, followed by the Delta Test. The Gamma Test and the FNN approach have similar reliable results in between of PCC and Delta Test.

It is obvious, that all the reviewed approaches identify a to complex model structure. One reason for this could be the difficult task, to recognition of the saturation.

Beside the discussed results, the systems were also evaluated with a different input signal. Even an amplitude modu-

lated pseudo random binary signal (APRBS) shows qualitatively the same results as the presented chirp signal. But the APRBS needs to get a variation like the presented chirp signal. That is because the APRBS delivers a constant zero signal for the first  $T$  steps  $u(k=0) = u(k=1) = \dots = u(k=T) = 0$ . So the output of the investigated systems become zero too  $y(k=0) = y(k=1) = \dots = y(k=T) = 0$ . This leads to a distance equal to zero in the input and the corresponding output. So the fraction of these distances is not defined. Another possibility is to cut the first steps of the signal, to get rid of this problem.

## VI. CONCLUSION AND OUTLOOK

In this work several model free input selection approaches are reviewed. It has been shown, that the most of them rely of the geometric in the Distance Space. The right subset with undisturbed output shows a sharp separation between feasible and unfeasible regions in the Distance Space, which the proposed approaches try to detect. While a to small input space acts similar to the true subset with noise, an input set with to many inputs does not differ that much from the true subset. So the automatically determination for each approach is a difficult task.

The free and the successively selection of inputs can be summarized as followed: For simple processes, all of the proposed approaches determine the model structure with a high certainty, but less quality. And the quality decreases even more with more complex process structures. Remember, that the development of the results are under optimal conditions. There was no noise in the input and in the output and the input has a high dynamic range, so that the whole input space is covered with data.

In contrast to the size of the right input space, there is no visible effect of the number of data samples to the result. Also the successively adding has to consider the structure (e.g. time series or input output system) of the underlying process. By disregarding, at least one unnecessary input is chosen by each approach.

Obviously it is hard to determine the correct model order of an complex process with uncomplex approaches. The next step is to compare the results of the model free approaches with model based ones. While today's modeling techniques are possible to identify nonlinearities, there should be noticeable differences between these two classes of model order determination approaches.

## REFERENCES

- [1] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning*. Springer, 2009, vol. 2, no. 1.
- [2] O. Nelles, "Nonlinear system identification," 2002.
- [3] J. Belz and O. Nelles, "Input selection using local model network trees," 2014.
- [4] X. He and H. Asada, "A new method for identifying orders of input-output models for nonlinear dynamic systems," in *American Control Conference, 1993*. IEEE, 1993, pp. 2520–2523.
- [5] C. Rhodes and M. Morari, "False-nearest-neighbors algorithm and noise-corrupted time series," *Physical Review E*, vol. 55, no. 5, p. 6162, 1997.

- [6] V. Krishnaswami, "determination of influencing inputs and their orders for nonlinear dynamic systems-application to ic engine modeling," in *American Control Conference*, 1997.
- [7] A. J. Jones, "New tools in non-linear modelling and prediction," *Computational Management Science*, vol. 1, no. 2, pp. 109–149, 2004.
- [8] F. Mateo and A. Lendasse, "A variable selection approach based on the delta test for extreme learning machine models," 2008.
- [9] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, "Feature extraction," *Foundations and applications*, 2006.
- [10] R. Orjuela, B. Marx, J. Ragot, and D. Maquin, "Nonlinear system identification using heterogeneous multiple models," *International Journal of Applied Mathematics and Computer Science*, vol. 23, no. 1, pp. 103–115, 2013.
- [11] M. Brown and C. J. Harris, "Neurofuzzy adaptive modelling and control," 1994.