Multi-Scale Uncertainty Analysis for Data Driven Models of Oxyfuel Combustion in Bubbling Fluidized Bed Combustor

Ivo Bukovský*, Vladimir Malý
Pavel Skopec*

&

Kei Ichiji, Noriyasu Homma

Czech Technical Univ. in Prague,
Faculty of Mechanical Engineering,
**Research Centre for*
Low-Carbon Energy Technologies
Prague, Czech Republic

Tohoku University
Graduate School of Biom. Engineering
Dpt. of Intelligent Biom. Systems Eng.
Sendai, Japan

presented by **Jiří Juřík**
FME, CTU in Prague

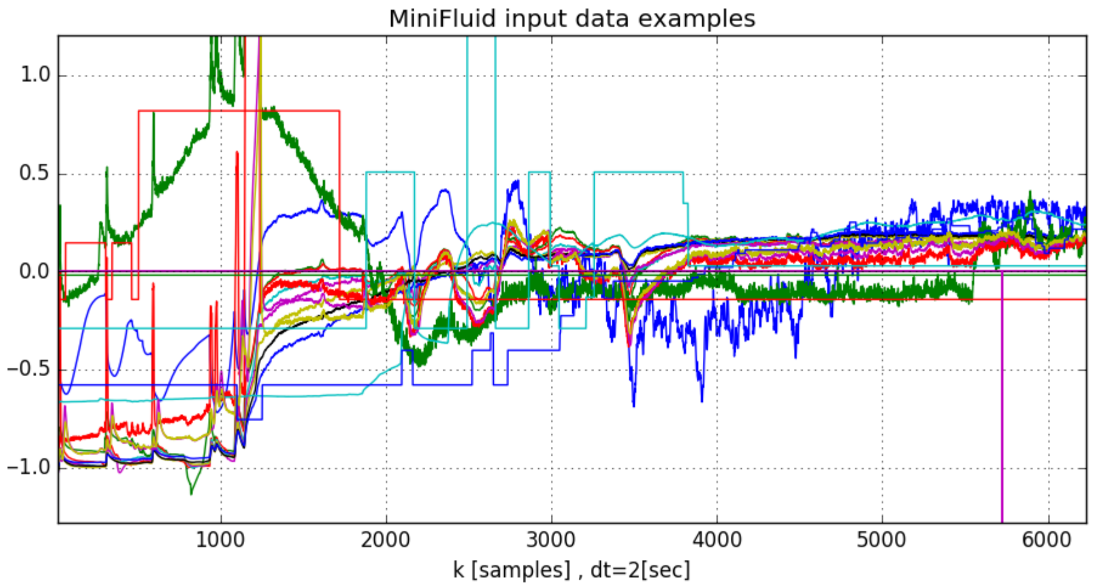
EUROPEAN UNION

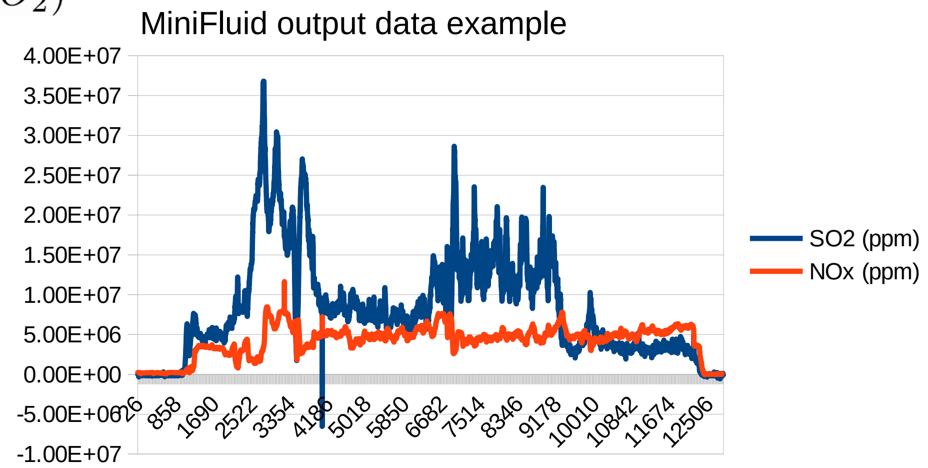
This work was supported by the Ministry of Education, Youth and Sports under OP RDE grant number CZ.02.1.01/0.0/0.0/16_019/0000753 "Research Centre for Low-Carbon Energy Technologies”

MINISTRY OF EDUCATION, YOUTH AND SPORTS

Motivation

- The oxyfuel combustion in a bubbling fluidized bed combustor is relatively new technology, and the combustion process is a complex nonstationary MIMO process:
  - about 33 input variables (incl. temperatures), and
  - at least 5 output variables $(CO_2, CO, SO_2, NO_x, O_2)$





We wish to developed reliable data-driven dynamical model of the oxyfuel combustion process, and proper (or mutliple validated) feature selection method is crucial esp. in case of system nonlinearity.

Motivation

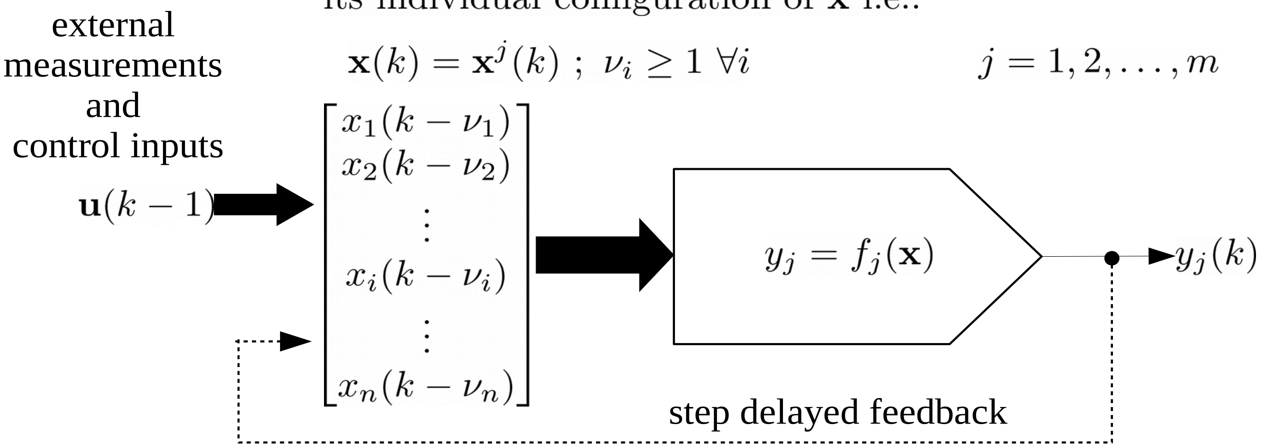
- The oxyfuel combustion in a bubbling fluidized bed combustor is relatively new technology, and the combustion process is a complex MIMO process:
  - about 33 input variables (incl. temperatures), and
  - at least 5 output variables $(CO_2, CO, SO_2, NO_x, O_2)$
- The accurate physical dynamical model of the process is unreachable and the process parameters (such as of the input fuel, etc) are not exactly known.
- We are researching the use of data-driven approaches (such as regression models, neural networks, Bayesian models, etc) to model the process with prospects of its optimization and control.
- Optimization techniques to achieve robust auto-regressive models are still a challenge.
- **At first here, we are focusing on data analysis for the feature vector of discrete-time dynamical computational models.**

Motivation

- A MIMO model can be implemented as multiple MISO models (multiple-input single output) $y_j = f_j(\mathbf{x}_j)$, where $f_j(.)$ denotes a computational model (regression model, neural network, Bayesian model) in general.
- It is easy to train a neural network to exactly fit the model for training data; however, this does not assure a correct model function for new input data (due to process nonlinearity, nonstatonarity, and variability in time <=> too many degrees of freedom)
- We need our computational model be causally correct, and it can hardly be achieved without a proper input data
- So the question is: **What is the proper configuration of input vector $\mathbf{x}$?** (i.e. What inputs and what feedbacks in $\mathbf{x}$? How long their history (how many step delays)?….)

each output variable y_j can have its individual configuration of $\mathbf{x}$ i.e.:

$$\mathbf{x}(k) = \mathbf{x}^j(k) \; ; \; \nu_i \geq 1 \; \forall i \qquad j = 1, 2, \ldots, m$$



Presentation Outline

- We will briefly discuss some fundamentals of methods for feature selection for data driven dynamical models (such as regression models, neural networks, Bayesian approaches,…) for complex multiple-input multiple-output (MIMO) systems
  - linear cross-correlations,
  - Principal Component Analysis (PCA),
  - neural network based auto-encoders,
  - Mutual Information (probability/likelihood) based techniques
- and then

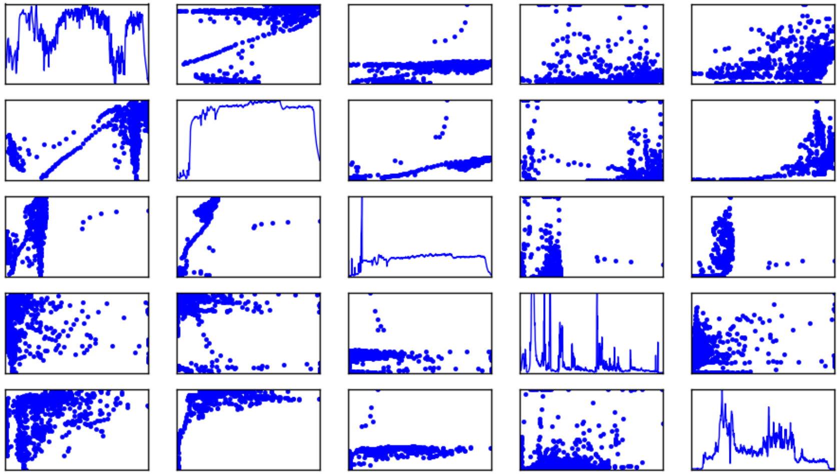
the presentation Objective

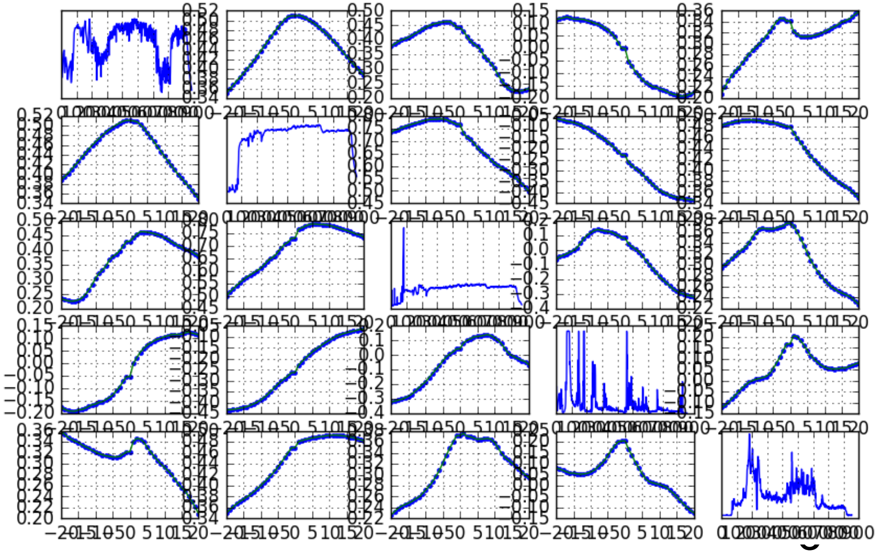
we discuss possible research direction of our original method [a] for
new research of feature selection via multi-scale false neighborhood minimization
(for optimizing the configuration of data driven discrete-time dynamical computational models)

[a] BUKOVSKY, I., Witold KINSNER, V. MALY a K. KREHLIK. Multiscale Analysis of False Neighbors for state space reconstruction of complicated systems. In 2011 IEEE Workshop On: Merging Fields of Computational Intelligence and Sensor Technology (CompSens) . 2011,p. 65–72. doi:10.1109/MFCIST.2011.5949517

Linear Cross-Correlation for MIMO Systems

- The cross correlation is for bi-variate linear dependency (SISO input-output), i.e., it evaluates the strength of linear relationship between two scalar variables.
- For MIMO systems it serves to reveal co-linearity among input variables.
- Assuming real processes are not so heavily nonlinear, the linear cross-correlation may be a first-to-do technique to:
  - indicate correlations of some input and output variables (if it can be linearly captured)
  - obtain some notion about time-shift (lag) between input and output variables
- However, for nonlinear MIMO systems, we can hardly use linear cross-correlation for more advanced model design

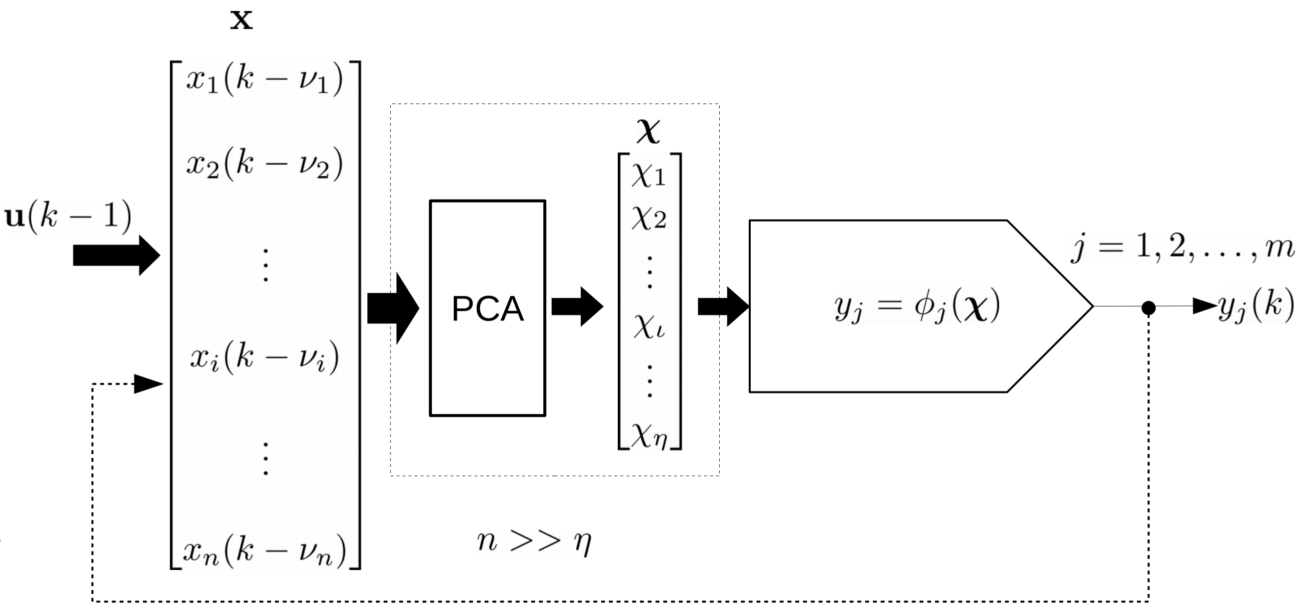




Principal Component Analysis (PCA) for (dynamical) Data Driven Models

- PCA is a classical linear transformation method that:
  - is based on linear cross-correlations, so
  - it reduces input data dimensionality via linear compression
    i.e. $\mathbf{X} \rightarrow \boldsymbol{\chi}$ (right fig.)

    while preserving the desired amount of information in data



- We applied PCA to dimensionality reduction for MIMO dynamical model of coal powder combustion process [b], where we decreased model complexity and suppressed outliers; however, the accuracy of the auto-regressive prediction was still limited (esp. for emissions of CO, CO_2)

- Furthermore, PCA as a dimensionality reduction technique is not suitable for direct feature selection (input $\mathbf{x}$ configuration, sensor selection), so additional techniques together with PCA has to be used.

[b] BUKOVSKÝ, Ivo a Michal KOLOVRATNÍK. A Neural Network Model for Predicting NOx at the Mělník 1. Acta Polytechnica [online]. 2012, 52(3) [vid. 2015-01-26]. ISSN 1805-2363. doi:10.14311/1538

Neural Network based Auto-Encoders for Data Driven Dynamical Models

- Auto-encoders (AE) are nonlinear mappings that encode (compress) and then decode (decompress) input data, so they perform nonlinear dimensionality reduction via some customized optimization algorithm.

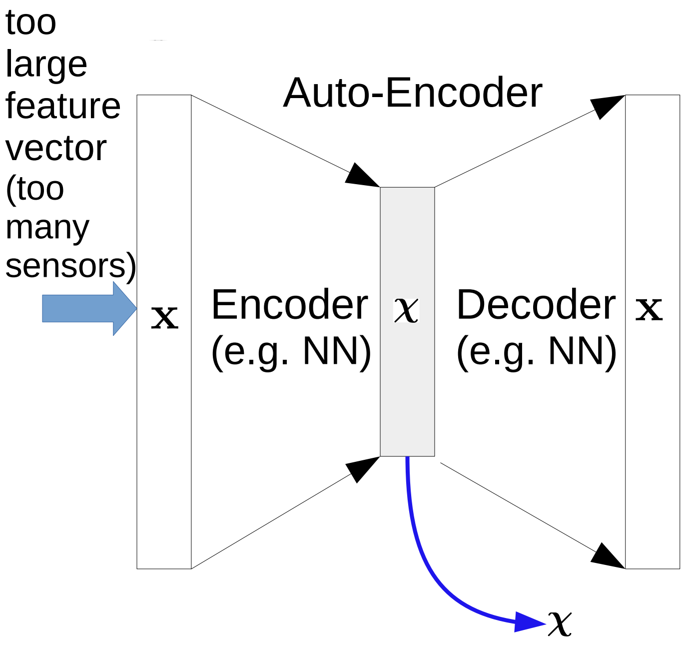
- AE has been widely studied and applied in Deep Learning that excels esp. in image and speech recognition nowadays.

- The application of Deep Learning in complicated tasks such as medical imaging or prediction of complex nonstationary dynamical MIMO systems is still an actual challenge.

- Because of customizable structure, the AE can combine linear and nonlinear dimensionality reduction.

- Extending studies of using AE for feature extraction for dynamical systems has been appearing in literature just since recently.

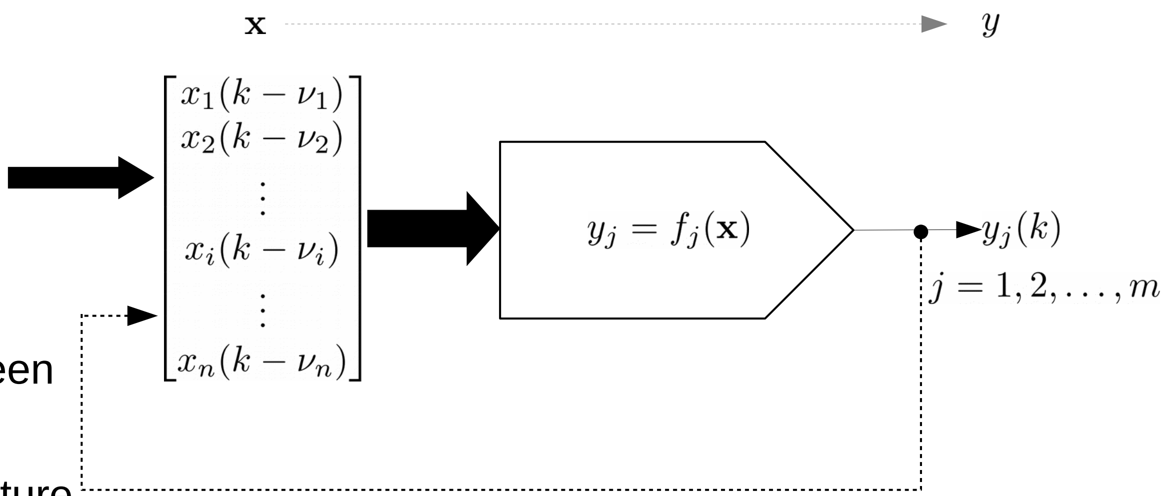
- The performance of AE can be excellent; however, it depends on:
  - their custom design,
  - selected (optimization) algorithms and its setups,
  - quality and amount of training data.



Mutual Information (probability/likelihood) Techniques for Data Driven Models

- Mutual Information (MI, Shannon 1948) is a probabilistic measure that quantifies the strength of nonlinear relationships.

$$MI(\mathbf{x}, y) = \sum_{y \in Y} \sum_{\mathbf{x} \in X} p(\mathbf{x}, y) \log \left(\frac{p(\mathbf{x}, y)}{p(\mathbf{x})\, p(y)} \right)$$



- Furthermore, MI can quantify the strength of nonlinear relationship not only between two scalar variables, but also between two vectors; thus,

- MI concept is more suitable for nonlinear feature selection, i.e. for configuration of input vector $\mathbf{x}$ via optimizing a suitable criterion Q that utilizes maximizing the strength of nonlinear relationship between $\mathbf{x}$ and y , i.e.

$$Q\Big(arg\big(max\{MI(\mathbf{x}, y)\}\big), \ldots\Big) \implies \mathbf{x}$$

- More recent achievements on information theoretic nonlinear feature selection via MI and recommendations on particular criteria selection and on suitability for small data set can be found in [c].

[c] Brown, G., Pocock, A., Zhao, M.J. and Luján, M. "Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection." Journal of Machine Learning Research 13 (2012): 27--66.

Here, finally the our method and its extension…

Multi-Scale Uncertainty Analysis for Data Driven Models...

- By the ***Uncertainty*** (in the design of input vector $\mathbf{x}$ for a given training data), we mean the existence of ***False Neighbors*** , i.e. if for two distinct times k_i , k_j:

$$\text{if } ||\mathbf{x}(k_i) - \mathbf{x}(k_j)|| < R_x \ AND \ ||y(k_i) - y(k_j)|| > R_y$$

$$\text{then } \mathbf{x}(k_i) \text{ and } \mathbf{x}(k_j) \text{ are false neighbors in } \mathbf{R}^n$$

- Nowadays, there exists published methods on feature selection via false neighborhood concept; however, usually with scalar thresholds Rx and Ry...
- We consider the user-defined fixed scalar thresholds Rx and Ry be still a challenge for further improvement of feature selection methods
- In [a] based on cumulative-sum approximation of power-law measure for multi-fractals, we have proposed to evaluate false neighbors in data over the whole range of setups of thresholds (to cope with the issue of a priory unknown optimal values of the thresholds Rx and Ry).
- Next slide recalls the core of this multi-scale method, and then the current research extension is presented

Multi-Scale False Neighborhood Analysis for Data Driven Models...

if $||\mathbf{x}(k_i) - \mathbf{x}(k_j)|| < R_x \; AND \; ||y(k_i) - y(k_j)|| > R_y$ then $\mathbf{x}(k_i)$ and $\mathbf{x}(k_j)$ are false neighbors in $\mathbf{R}^n$

- The method is based on cumulative-count approximation of power-law measure for multi-fractals ,
- so instead of detecting and counting False Neighbors (FN) for constant thresholds *Rx* and *Ry*,
- we detect and count FN for the range of thresholds $Rx=\{Rx_1,Rx_2,Rx_3,\ldots\}$, $Ry=\{Ry_1,Ry_2,Ry_3,\ldots\}$ via ***Areal Cumulative FN*** count or ***Diagonal Cumulative FN*** count as:

$$ACFN = \sum_{\forall i} \sum_{\forall j} FN(Rx_i, Ry_j) \qquad DCFN = \sum_{\forall i} FN(Rx_i, Ry_i)$$

where *FN*(.) is the count of false neighbors in the all dataset for given thresholds.

please see [a] for more details on *False Neighbor Matrix* and on power-law cumulative sum approximation.

- In the very principle, the above cumulative measures helps to avoid the risk of inappropriate threshold selection if FN shall be found via merely single valued thresholds *Rx* and *Ry*.
- The next slide shows (and discuss) ideas of our new research as we wish to developed fully automated multi-scale method for feature selection to achieve more reliable data driven model of the oxyfuel combustion process.

Multi-Scale False Neighborhood Analysis for Data Driven Models…
Current Research Extension

- We are developing a method that will design (or will be a supportive method for) the automated input vector configuration (feature selection) for complex data-driven dynamical MIMO model of (oxyfuel combustion) via the multi-scale false neighborhood concept ,i.e. via minimizing false neighbors in model configuration and its training data
- Currently several directions are investigated:

$$\begin{array}{ccccc} \boldsymbol{\theta} & & \mathbf{x} & & \\ \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_i \\ \vdots \\ \theta_n \end{bmatrix} & \times & \begin{bmatrix} x_1(k-\nu_1) \\ x_2(k-\nu_2) \\ \vdots \\ x_i(k-\nu_i) \\ \vdots \\ x_n(k-\nu_n) \end{bmatrix} & = \mathbf{x}_{\boldsymbol{\theta}} \end{array}$$

$$\theta_j = \begin{cases} 0 \ldots & \text{feature is not selected} \\ 1 \ldots & \text{feature is selected} \end{cases} \qquad \forall j$$

1) Being inspired by other feature optimization concepts, e.g. via the informatic theoretic feature selection [c] and optimization techniques therein, we may apply some optimization technique to a binary feature selection vector $\boldsymbol{\theta}$ (right top), so we can *search for minimum configuration of* $\mathbf{x}$ *with minimum number of multiscale false neighbors.*

2) The *sampling period of a model can be optimized via minimizing the multi-scale false neighbors*

3) Other optimization techniques to find a minimal input vector $\mathbf{x}$ with minimal multi-scale false neighbor measures … (?)

Summary

- We briefly reviewed some principles and discussed their potentials for feature selection for data-driven MIMO dynamical systems, i.e.:
  - Linear cross-corelations as the most fundamental method for getting first notion about possible relationships between data; however, that is rigorously unsuitable for design of nonlinear MIMO systems
  - PCA that reduces dimensionality of a feature vector; however, that is limited by its linearity and needs other extensions to perform feature extraction
  - Auto-encoders that perform powerful nonlinear dimensionality reduction; however, their use for nonlinear data-driven dynamical models, i.e. long term auto-regressive predictors is still not much explored
  - Information theoretic nonlinear feature selection, i.e. based on Mutual Information (probability/likelihood), as rather a well explored concept and popular domain
- Then, we recalled our original method on multiscale false neighbors analysis that is robust against the inappropriate threshold selection for false neighborhood, and
- we indicated its potentials for development of a new type of computationally effective automated feature selection method to support the design of complex data-driven dynamical systems.

- We hope, you could find something interesting in our presentation,
- we hope, we might initiate discussion and get your very valuable feedback (now or later)
- we hope, you are enjoying the Workshop and your stay in Prague

Thank you very much for your attention!!!

EUROPEAN UNION

This work was supported by the Ministry of Education, Youth and Sports under OP RDE grant number CZ.02.1.01/0.0/0.0/16_019/0000753 "Research Centre for Low-Carbon Energy Technologies”

MINISTRY OF EDUCATION, YOUTH AND SPORTS