Nonlinear Model Identification



Problem Nonlinear Identification 2

D’Amicis Salvatore – Romeo Matteo

875902 - 876029

**Description of the process:**

The process describes a heat exchanger where water is heated by pressurized saturated steam through a copper tube. The output variable is the outlet liquid temperature. The input variables are the liquid flow rate, the steam temperature, and the inlet liquid temperature. In this experiment, the steam temperature and the inlet liquid temperature are kept constant to their nominal values.

The provided dataset of 4000 samples has been divided in two parts:

* 3000 sample for the identification process
* 1000 sample to validate the identified model in term of prediction and simulation

**Description of the project:**

The goal is to perform a nonlinear model identification of the process using two different approaches:

* FROE (Forward regression orthogonal estimator) algorithm on a polynomial NARX
* Feedforward neural network

The identification has been performed in both cases based on different model assumption and complexity.

Python is the programming language used for this project.

Forward regression orthogonal estimator

Initially, as first assumption, the first model has been chosen with reference to the paper provided to us.

This is a nonlinear *ARX* model of the type

where 3 is the order of the AR part, 9 the order of the exogenous one and is a polynomial expansion of the second degree.

The algorithm, setting a threshold p = 0,05 for , provide the following result:

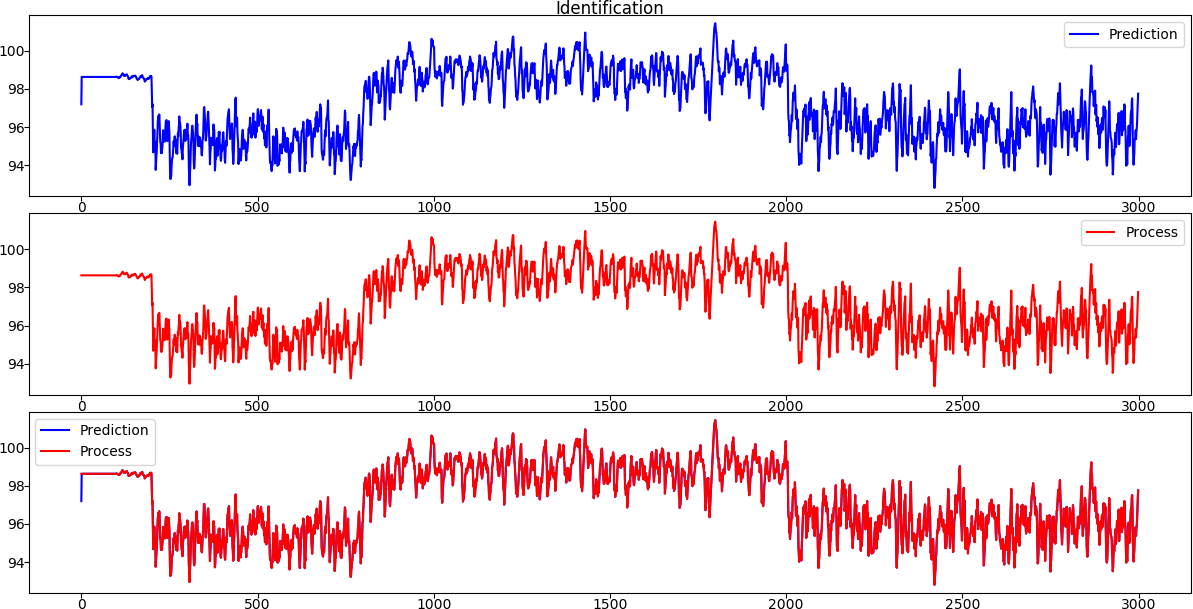
* The forward selection process stopped after just one iteration, choosing as first regressor because the related to it was 0.999. This fact is explained because the process is a slow varying one, and the FROE algorithm tend to prefer autoregressive terms in the first stages of the selection process.
* Identified Model: where

Now, let’s see the result in term of prediction and simulation:

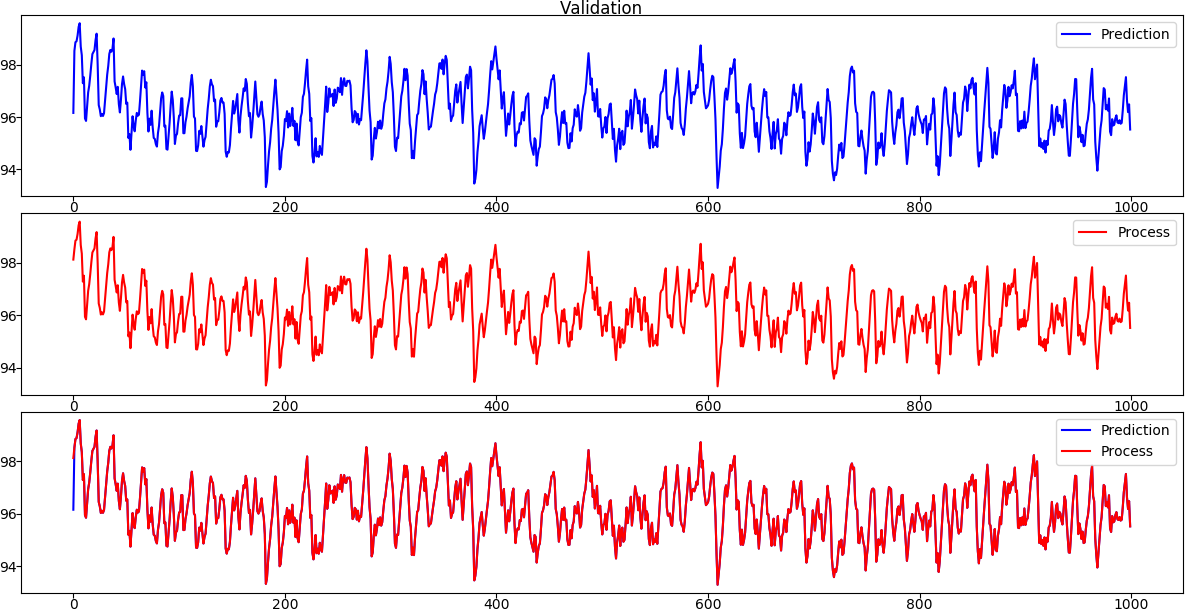
* Result in terms of prediction on the identification set [fig1] and validation set [fig 2] are good, but even if 1-step ahead prediction properties are good, simulation is extremely inaccurate[fig3].

|  |  |  |
| --- | --- | --- |
| MSE on identification | MSE on validation | MSE on simulation |
| 0.0006836127158434082 | 0.003886567056177283 | 2.2016865512134025 |

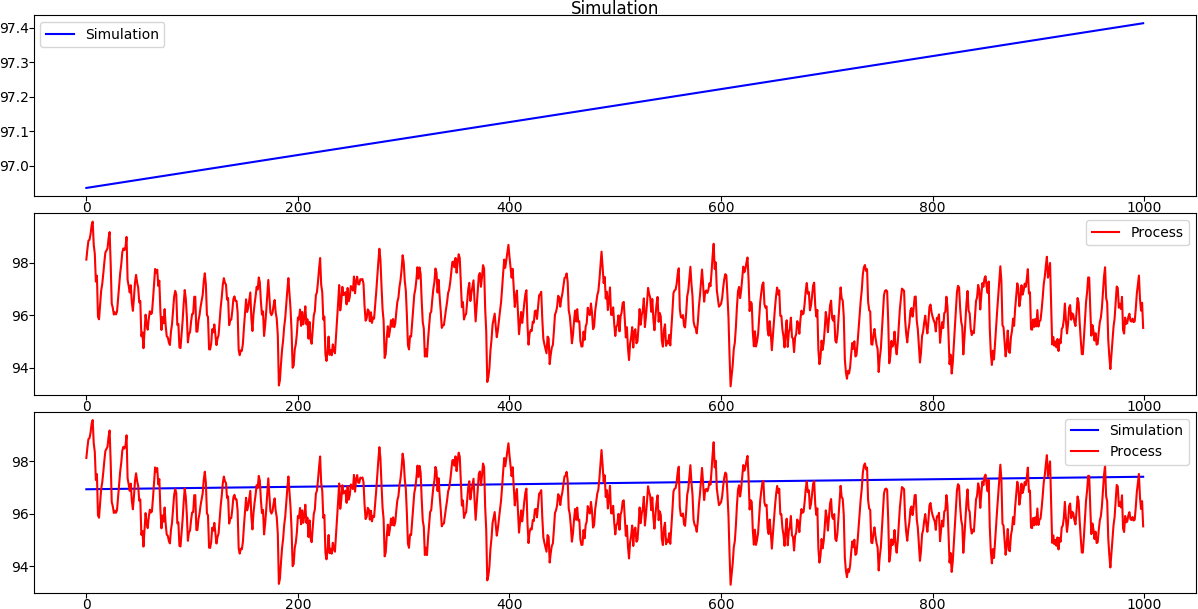
* Correlation tests performed on the model [fig4], confirm that the model with one regressor is a good model in term of prediction (all the correlation tests are satisfied, considering a 95% confidence interval)
* Bad simulation capabilities are justified by the selection of just one autoregressive term, which is not enough to describe the process in a long-time period. The identified model is a simple straight line in simulation, with slope



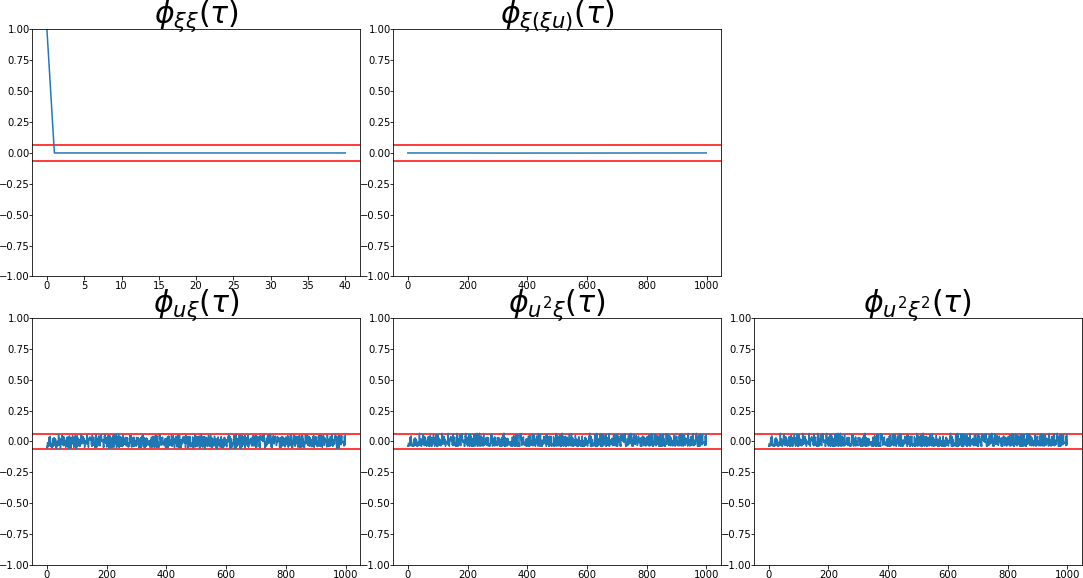
Prediction on identification set [fig1]



Prediction on validation set[fig2]



Simulation of the process [fig3]

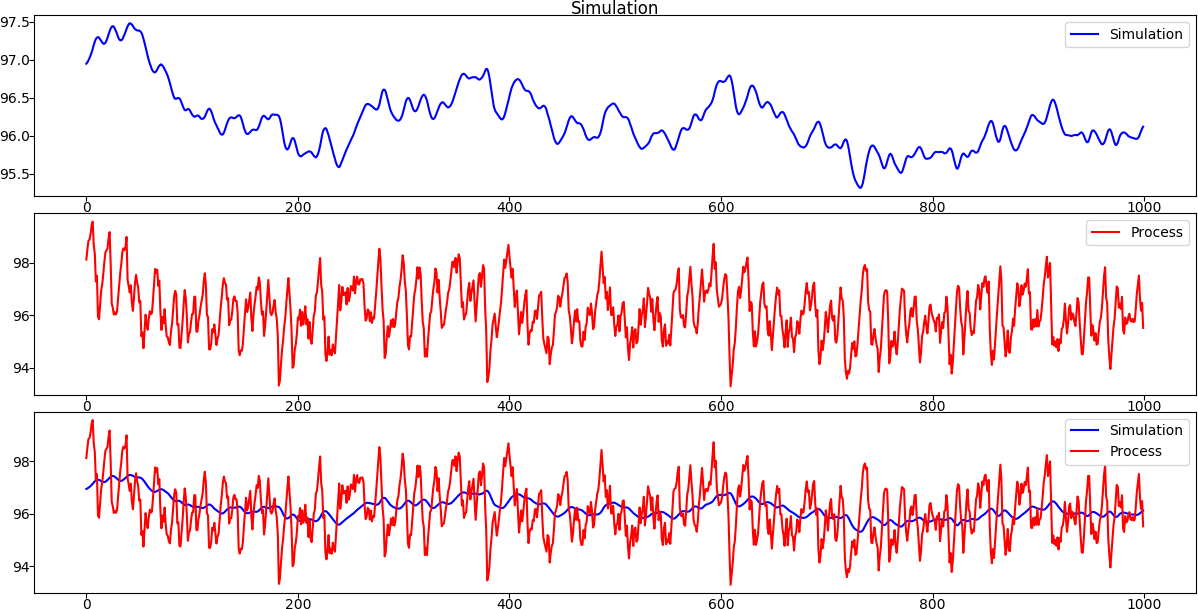


Correlation Tests [fig4]

After this, 3rd and 4th degree of the polynomial expansion has been tried, leaving the threshold at the same value. As expected, FROE produced the same model as in the previous case.

As consequence of the previous considerations, to find a model with good simulation properties, the model performance has been analysed adding one regressor at a time.

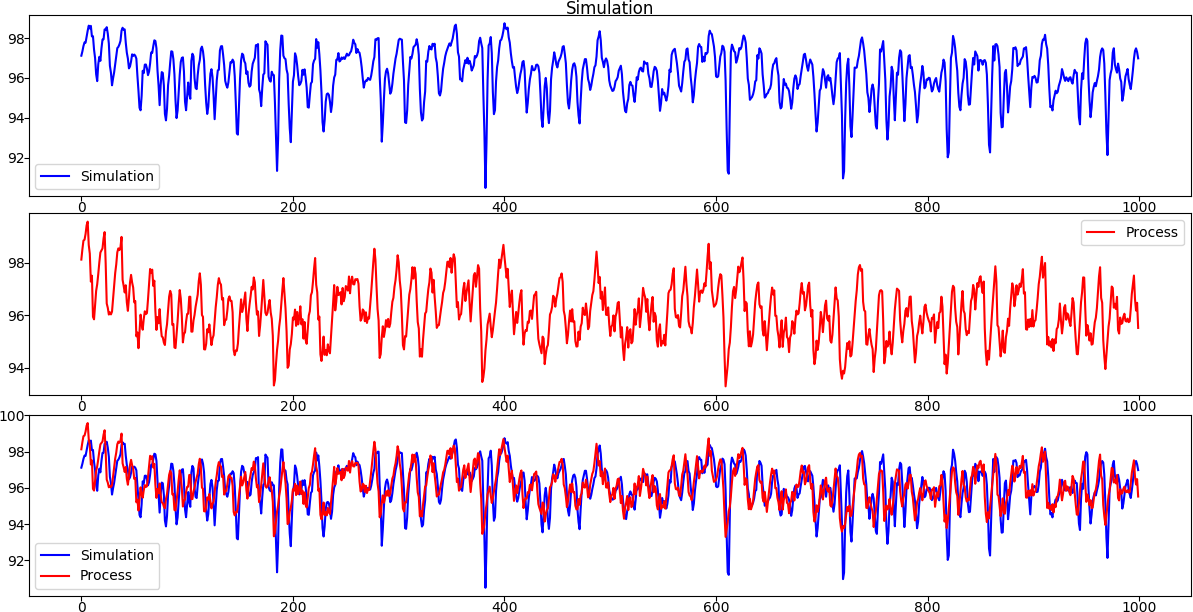
The best model in term of simulation (2nd degree polynomial expansion) has been found including all the regressors [fig5], but seems that the 2nd degree is not enough to describe the system dynamic.



2nd degree polynomial simulation including all the regressors [fig 5]

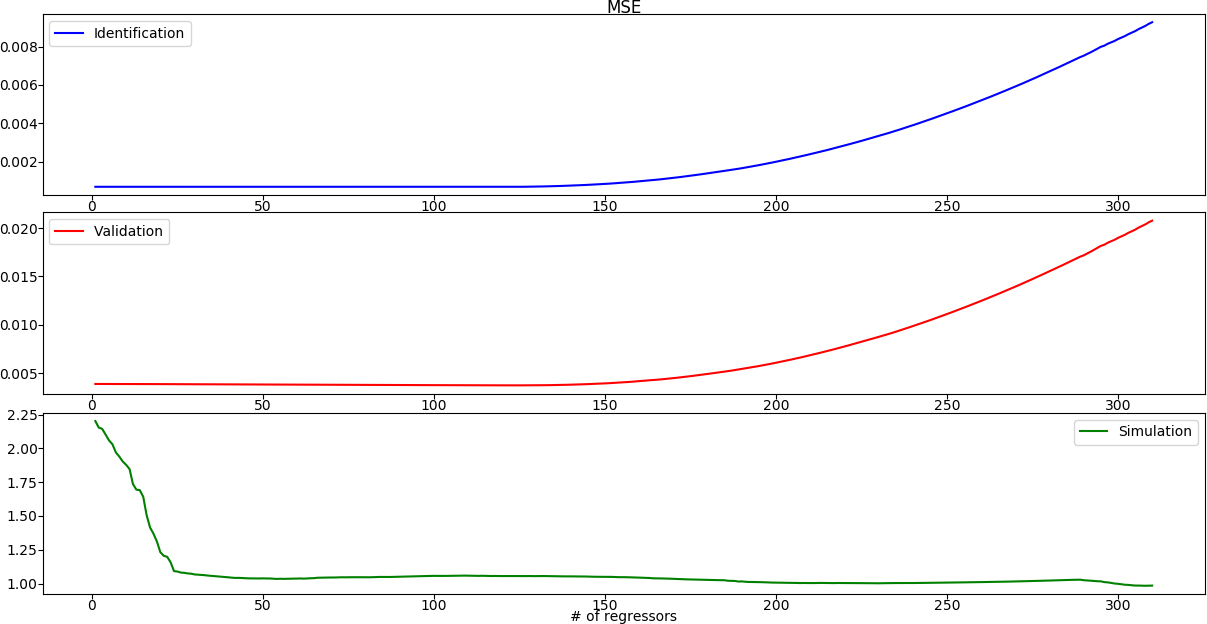
|  |  |  |
| --- | --- | --- |
| MSE on identification | MSE on validation | MSE on simulation |
| 0. 0006783057528772301 | 0. 0038079139673512816 | 1.041176337143496 |

Considering the 3rd and 4th degree, the best result has been achieved on the 3rd degree with 310 regressors [fig6], with a slightly worst result in prediction performances [fig7].



3rd degree polynomial simulation including 310 regressors [fig 6]

|  |  |  |
| --- | --- | --- |
| MSE on identification | MSE on validation | MSE on simulation |
| 0. 009270671400389164 | 0. 02076928218235279 | 0.9849954100637331 |



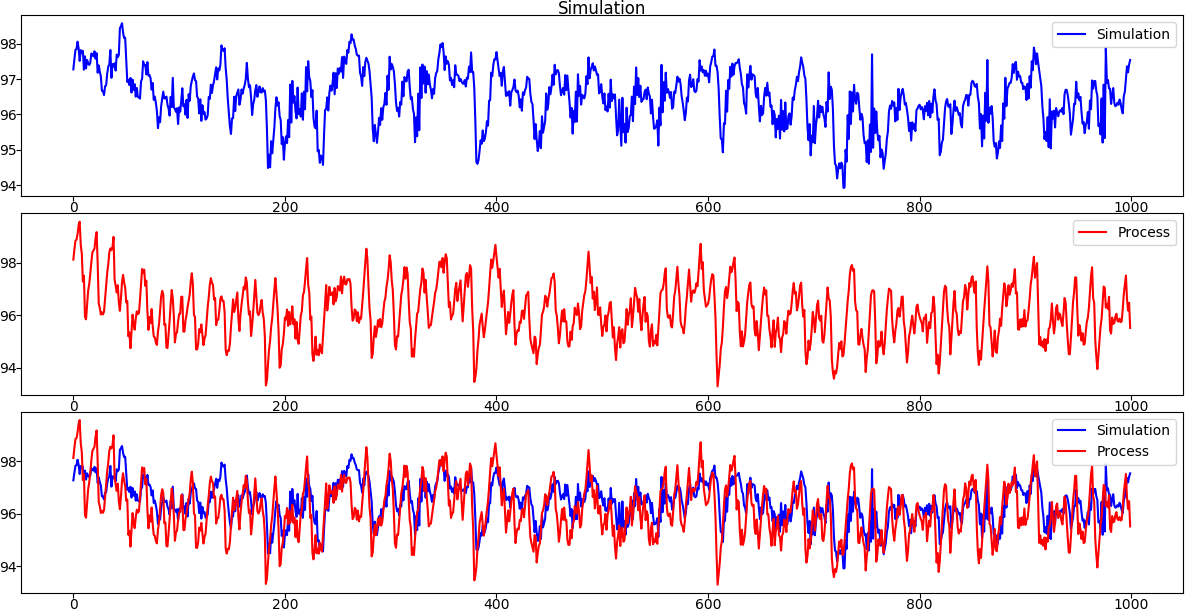
MSE on identification, validation and simulation [fig7]

Analysis considering different order of AR and X part

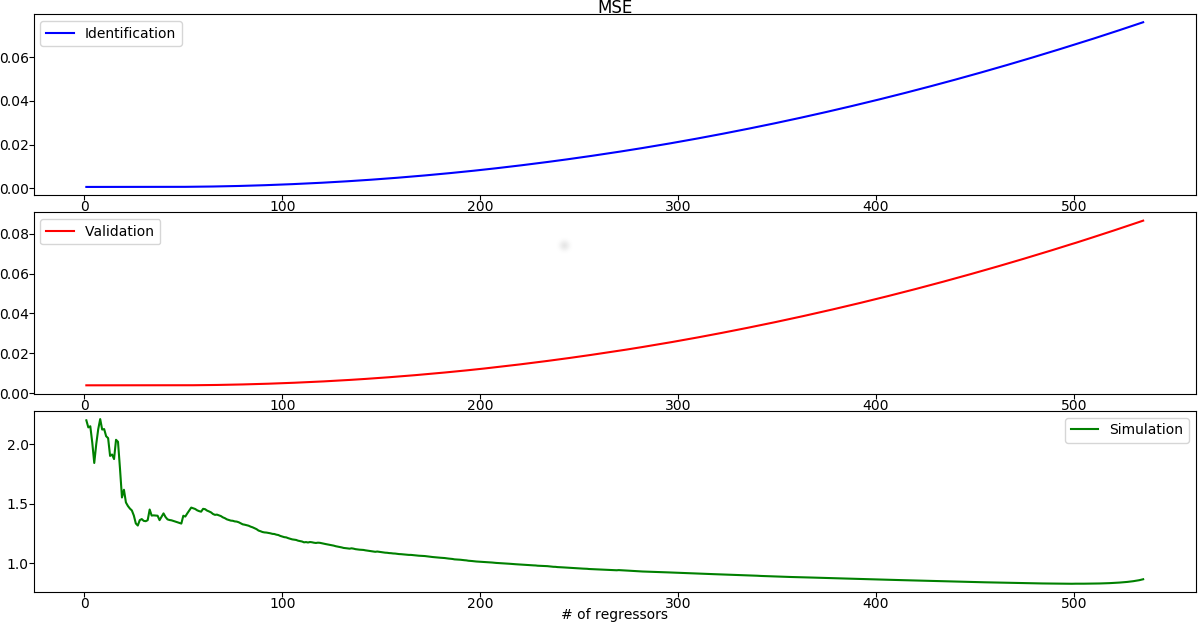
After the analysis of the model of the reference paper, an analogous procedure has been adopted starting from models with different order of AR and X parts.

After the first group of experiment (different combination with maximum 4 delays for each part), some consideration came out:

* Models with only exogenous parts does not provide good results both in prediction and simulation
* As expected, all the models with at least one autoregressive term provide good prediction properties, because of the presence of the regressor in all of them.  
  About simulation, what has been found out is that models with less than 3 delays in AR and 4 delays in X parts aren’t able to produce at least acceptable result [fig8] (also in this case, with slightly worst performance in prediction [fig9])



Simulation on 4th degree polynomial, 510 regressors, AR = 3, X = 4 [fig8]



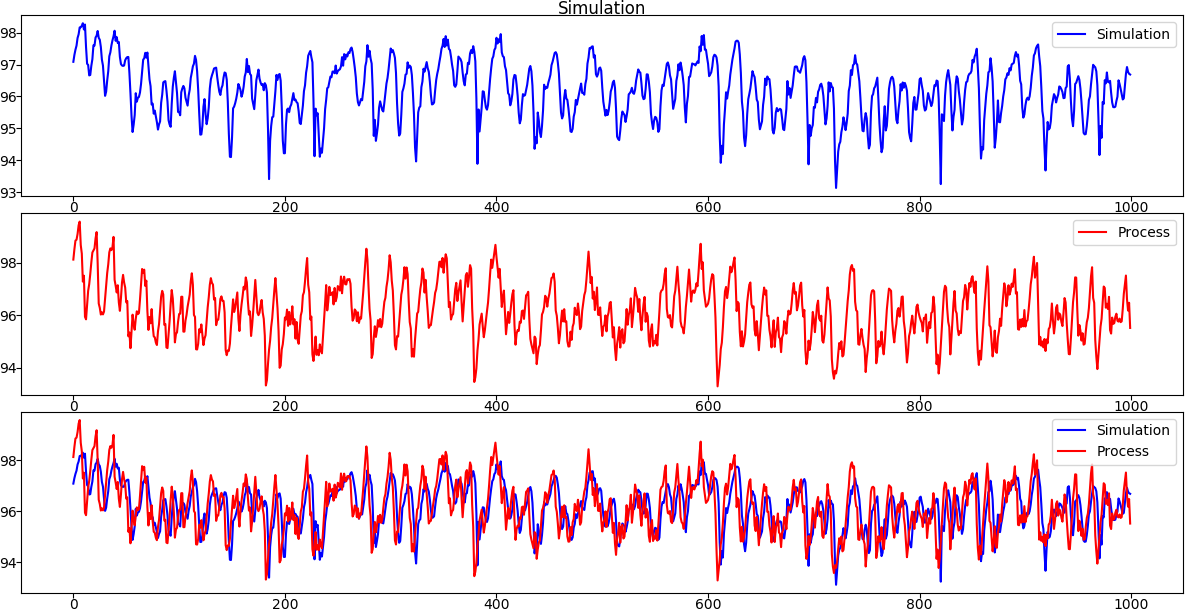
MSE on 4th degree polynomial, 510 regressors, AR = 3, X = 4 [fig9]

Starting from the latter shown model, different experiment has been tried adding more autoregressive terms to the model, without any significant improvement of the performance in simulation.

This mean that with this kind of nonlinearity, the added autoregressive terms are not useful to describe the real system dynamic.

After that, starting again from the latter model (AR = 3, X = 4), exogenous terms have been added:

* 2nd order degree of the polynomial expansion, as expected, is not able to describe well the real system dynamic
* The addition of exogenous terms has provided a slightly improvement on the simulation performances, considering at least a 3rd degree of the polynomial expansion, with a high number of regressors included in the model (example in fig10)



Simulation of a AR=3, X=6 model, 4th degree of polynomial expansion and 250 regressors[fig10]

|  |  |  |
| --- | --- | --- |
| MSE on identification | MSE on validation | MSE on simulation |
| 0.00964563878653221 | 0.03295860849205921 | 0.8423823937450876 |

Final Considerations

All the model with at least one autoregressive terms can produce really accurate result with a low number of regressor in prediction (due to the slow-varying nature of the process/oversampled process).

Simulation is far more complicated, since the optimal is in the NOE form:

Process output could not be used, and the model should be able to simulate the system dynamics in the long term. Due to the nonlinearity of the process, to obtain good result in simulation, a high order of the polynomial expansion is needed (at least 3rd degree) combined with a high number of regressors in the model.

Feedforward Neural Network

For the definition of the Neural Network, sklearn.neural\_network.MLPRegressor(Multi-layer Perceptron regressor) library has been used:

* This model optimizes the squared-loss using LBFGS, an optimization algorithm in the family of quasi-Newton methods.
* Activation function used in each neuron: Sigmoid (Logistic)
* Number of epoch = 200

The same approach as FROE has been adopted, that is: the first model has been chosen with reference to the paper provided to us.

Where, is a Multi-layer perceptron with the characteristics described above.

Another parameter which has not been cited yet, is the structure of the network (number of hidden layer, number of units per each layer), which has been set (just initially) as described in the reference paper: 2 hidden layers, 14 unit in the first one, 7 in the second one.

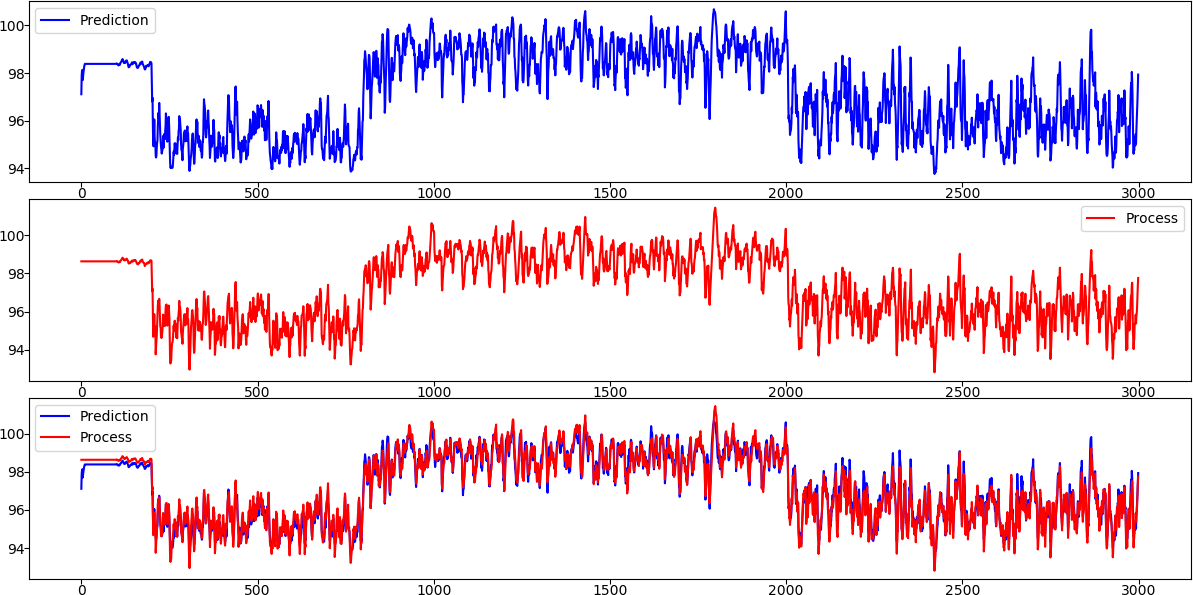
During the training, it has been noticed that, as expected, the result where different depending on the initialization: some of them were stuck in a local minima, producing a bad model both in prediction and simulation.

Consequently, we will report only model found with initialization leading to good result at least in term of prediction.

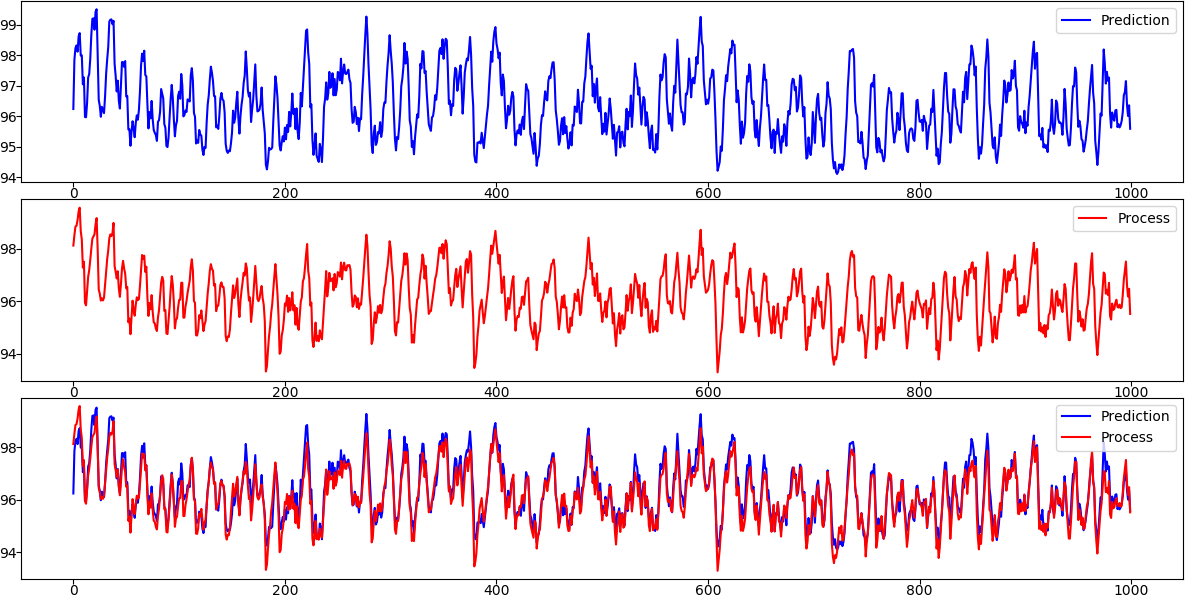
The training of the algorithm on the defined structure provided the following result:

|  |  |  |
| --- | --- | --- |
| MSE on identification | MSE on validation | MSE on simulation |
| 0.08679324327064053 | 0.12140612063786728 | 0.8913868669647105 |

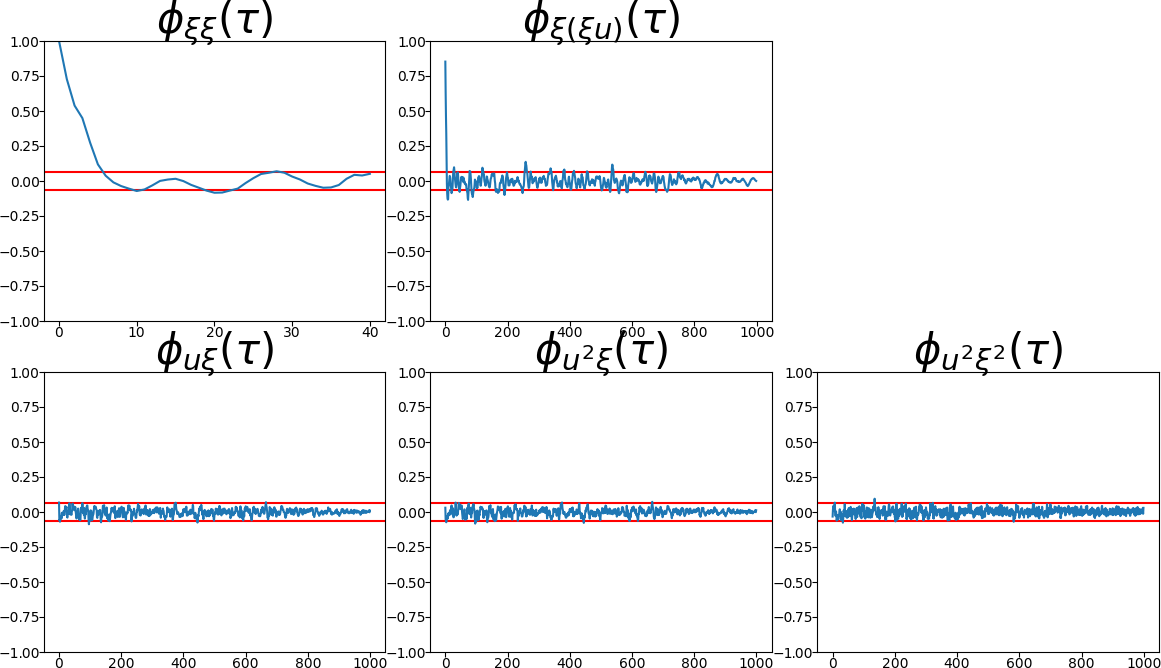
* With this structure, the NN model provide good result in term of prediction, as shown in fig11 (identification set), fig12 (validation set), fig13 (correlation tests)
* In simulation, the complexity of the nonlinear function defined by the NN seem able to describe the system dynamic pretty well (with some imprecision on the peaks) [fig14]



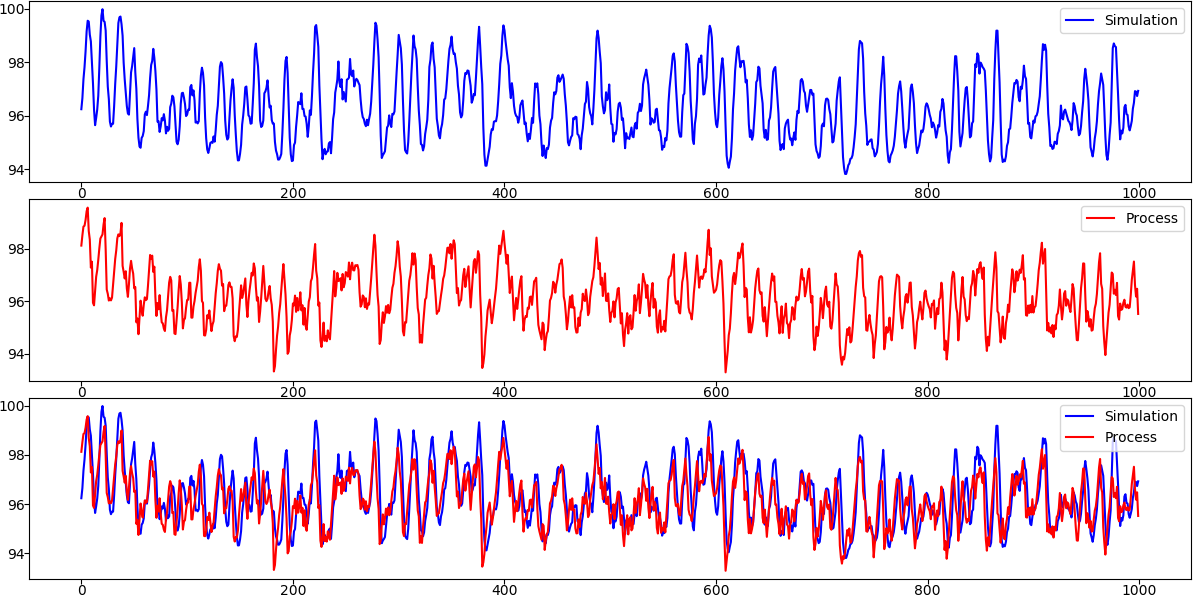
Prediction of the NN on identification set [fig11]



Prediction of the NN on validation set [fig12]



Correlation tests [fig13]



Simulation of the NN on validation set [fig14]

Changing the complexity of the network, but maintaining the same inputs of the NN, even with simple models (1 hidden layer and low number of units) is possible to obtain similar (or even better) results.

|  |  |  |
| --- | --- | --- |
| Structure | MSE on validation | MSE on simulation |
| 3 units | 0. 13167975263834955 | 1.1620545179118897 |
| 7 units | 0.13067908744772172 | 0.5600075486428612 |
| 10 units | 0.12641506575635966 | 0.8000548368212885 |
| 14 units | 0.13689108030405953 | 0.8790660951684601 |
| 15 units | 0.12264832284906814 | 0.771198586167046 |

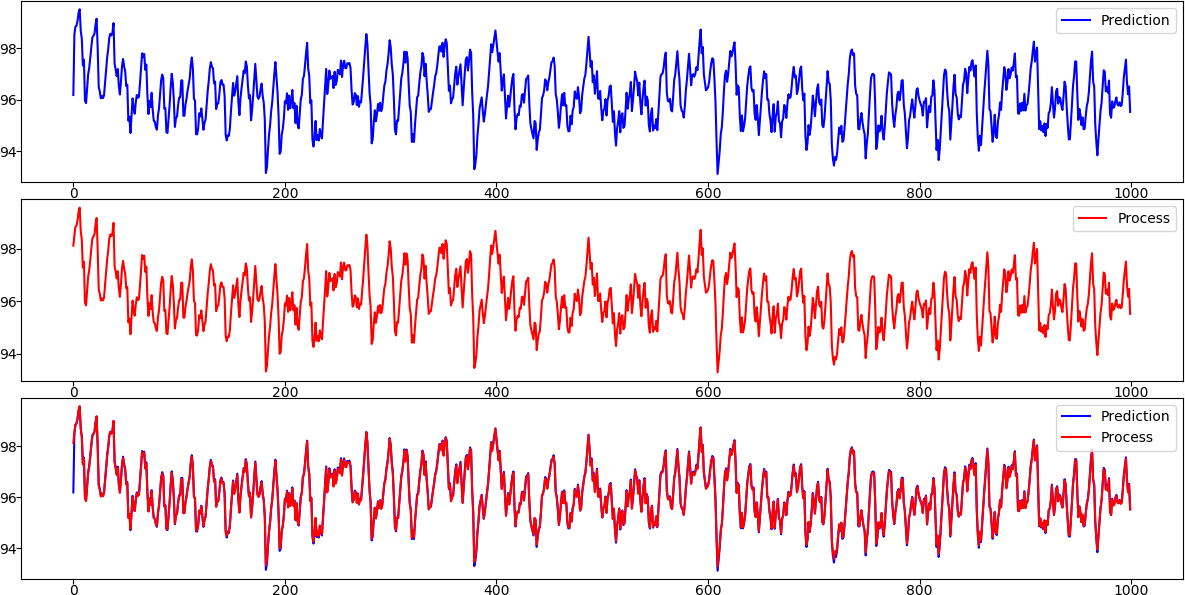
Analysis considering different order of AR and X part

In this section, the same analysis done with FROE has been repeated, considering different orders of the autoregressive and exogenous parts, and for each of them different complexity of the network structure.

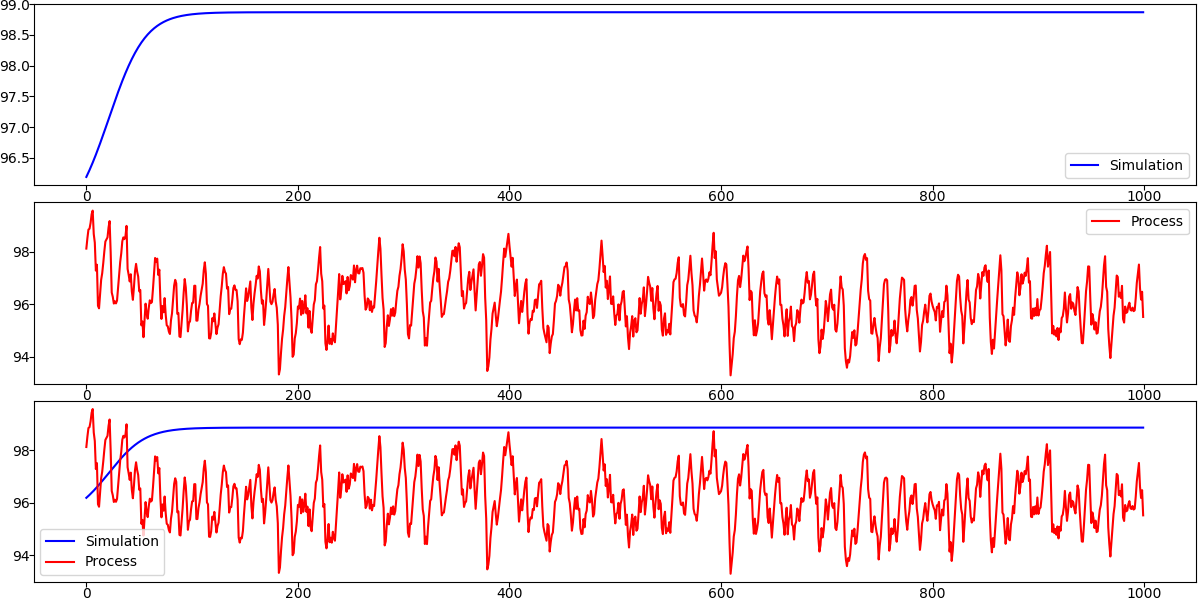
After some experiments on different model structure, some considerations came out:

* Models with exogenous parts does not provide good result both in term of prediction and simulation
* Models containing just one autoregressive and no exogenous part can perform really accurate result in term of prediction even with a simple structure of the network, but very bad result in simulation [fig15, fig16]

|  |  |  |
| --- | --- | --- |
| MSE on identification | MSE on validation | MSE on simulation |
| 0.0027038651198139025 | 0.005402811135141685 | 8.355283005584166 |



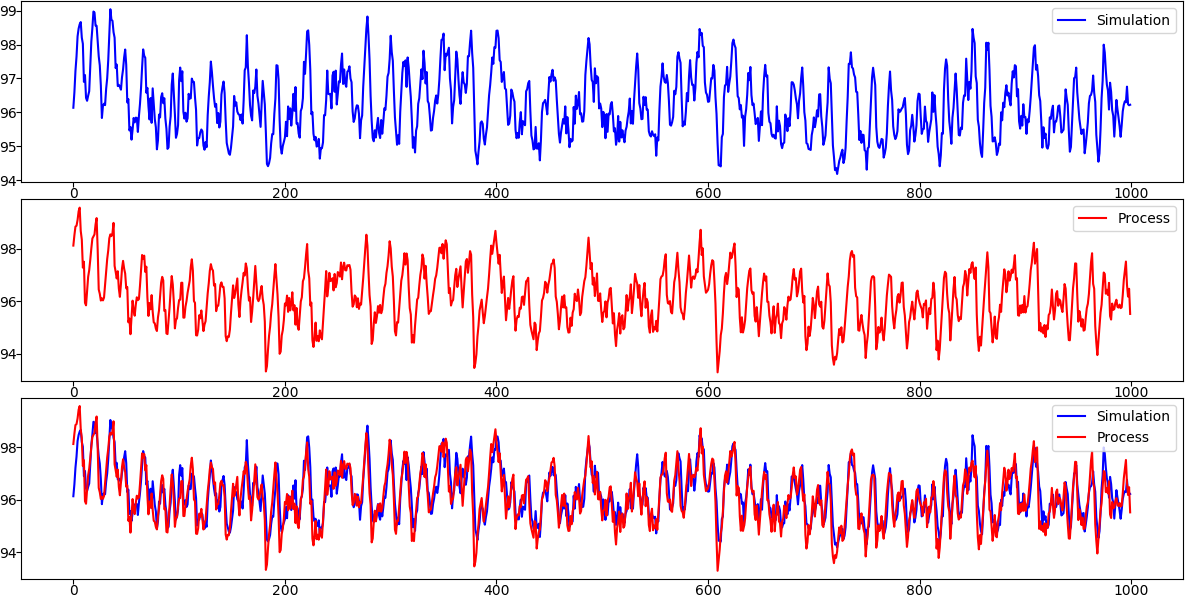
Prediction on validation set, , 1 hidden layer with 2 units [fig15]



Simulation on validation set, , 1 hidden layer with 2 units [fig16]

* Good result in simulation are obtained with low order of AR and X parts on some network structure, for example, one of the best result in simulation has been reached on a 2 layers NN, with 1 unit in the first layer, 12 units in the second layer, AR = 1, X = 3 (In this case, a more accurate result in simulation [fig17] has led to a slightly worst result in prediction with respect to the previously described model)

|  |  |  |
| --- | --- | --- |
| MSE on identification | MSE on validation | MSE on simulation |
| 0.09599881771183875 | 0.12690061111103054 | 0.3414172222318222 |



Simulation on validation set, AR=1 and X=3, 2 hidden layers with (1, 12) units [fig17]

From now on, increasing the number of regressors considering different complexity of the network structure, one can notice that (provided a correct initialization) good result in prediction are achieved.

Some of them produce also good result in simulation.

During the experiment, a strange behaviour on the MSE of the identification has been noticed: instead of decreasing with the complexity (as expected) it was increasing. After some considerations, we discovered that the behaviour was explained by the maximum number of epochs through which the NN was trained.

In fact, the defined threshold of 200 iterations was not enough to reach convergence on more complex network structures.

Despite this behaviour the analysis has been done keeping fixed the n° of iteration because the low computational power of the machine running the algorithm was not enough to examine also different iterations in a reasonable amount of time.

Comparison FROE – Feedforward NN

­­­­The application of two different approaches in identification of nonlinear systems has naturally carried out some consideration about them, regarding the identified process:

* FROE identification with polynomial NARX produce a model which is surely more interpretable with respect to the NN model.

Due to its interpretability, the FROE model allows a better classification about relevant and non-relevant regressor.

* Prediction performance are generally good with both approaches.

In particular, FROE models seem to be slightly better (especially compared to complex NN) but, as mentioned before, parameter setting is crucial in NN (max number of epochs, initialization of the weights) and with a good one, NN are also capable of produce the same results or even better.

* In simulation, where the correctness of the identified model become essential, NN are capable to perform better than FROE (at least considering the performed experiments, with maximum degree of the polynomial set to 4), due to the high degree of nonlinearity that NN can represent even with simple structures of the network.