Model Identification And Adaptive Systems Project

PROBLEM NL5

Milani Federico - 876346

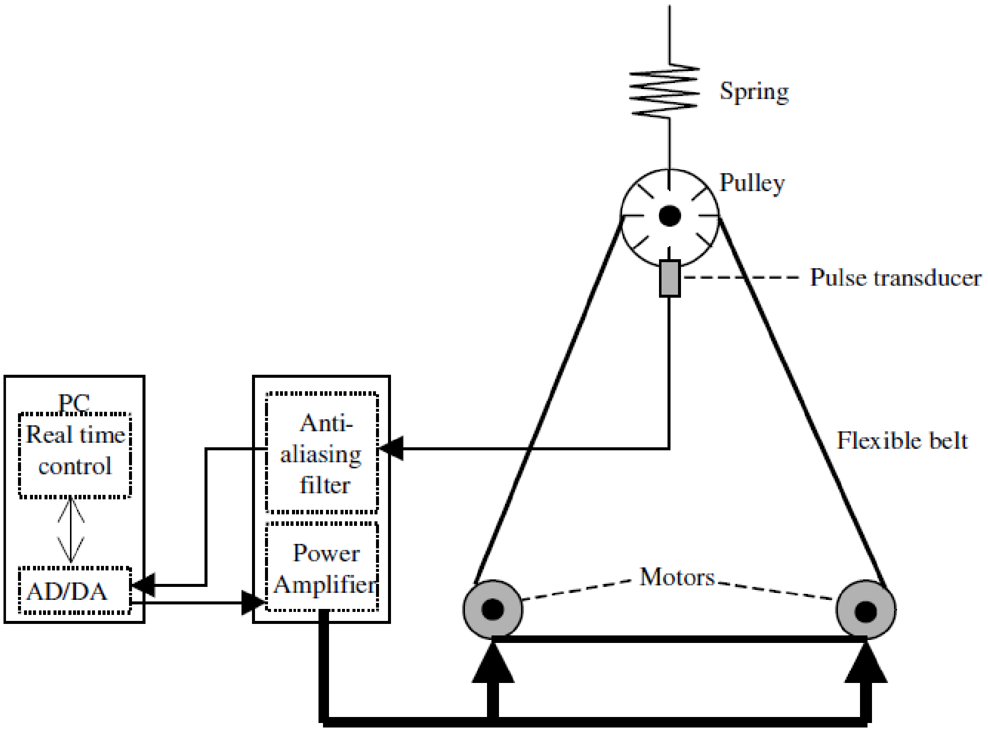
Urzino Davide - 875449

THE PROBLEM

The target of this work is to perform a nonlinear model identification of a SISO system with the FROE algorithm, using NARX or NARMAX models.

THE SYSTEM

The model obtained needs to describe the dynamics of this system:



The CE8 coupled electric drives (Wellstead, 1979) consists of two electric motors that drive a pulley using a flexible belt. The system is depicted in the figure. The pulley is held by a spring, resulting in a lightly damped mode. The electric drives can be individually controlled allowing the tension and the speed of the belt to be simultaneously controlled. The drive control is symmetric around zero, hence both clockwise and counter clockwise movement is possible.
Here the focus is only on the speed control system.

THE DATASET

The available data have been generated by feeding this system with two different kind of inputs

- **PRBS** data: the first input signal was a PRBS switching between $-u_{PRBS}$[V] and $+u_{PRBS}$[V], resulting in the process changing the belt rotation detection. The realizations were recorded for $u_{PRBS} =$ 0.5,1,1.5 [V]
- **UNIF** data: the second type of input signal was obtained from a PRBS, switching between -1.0 [V] and +1.5 [V] (first realization), as well as between -1.0 [V] and +3.0 [V] (second realization). The signal was then multiplied with a random number, uniformly distributed in amplitude between 0 and 1. The resulting input signal is then uniformly distributed in amplitude

IMPLEMENTATION

All the scripts are developed in Python with the help of some libraries to read the datasets, perform polynomial expansion, matrices operations and validation tests.

NARX IDENTIFICATION

The first goal of the implementation was to find an algorithm to identify NARX models using FROE and the ERR index.

Initially all the data are extracted from the provided files and separated into identification and validation sets (different combinations might be tried to find a good model), then the identification can start with the creation of Φ.

Φ is generated based on the parameters of the desired model: polynomial degree, biased or not and maximum lags for y and u. At first a basic Φ is filled with all first-degree terms and then it is expanded to include all the non-linear terms to reach the right polynomial degree. The data provided are very few (500 points) so the missing data are initialized each at the mean of its variable, instead of reducing further the data by using only the ones starting from the maximum lag.

All the needed matrices for FROE, such as W, A, G and Θ, are initialized as zeros matrices and then iteratively filled following the algorithm described in "Identification of MIMO non-linear system using a forward-regression orthogonal estimator" (Billings, Chen, Korenberg,1989):

- For every iteration:
  1. A_i, G_i, W_i, ERR_i of every non-chosen $regressor_i$ is calculated
  2. the regressor with the highest ERR_i is added to the model and W, A and G are updated accordingly
  3. the highest ERR_i is summed to the current "variance explained"
  4. if the stopping condition $(1 - Variance\ explained < 1 - Variance\ Threshold)$ is met the algorithm stops iterating

Θ is now calculated based on G and the chosen regressors, then an easily readable model is generated to be printed.

When the regressor selection is concluded, it's time to generate the prediction and simulation model. When dealing with prediction the model is easily generated as $\hat{Y} = \Phi\widehat{\Theta}$ where Φ is created as before and $\widehat{\Theta}$ is the one estimated by FROE. Instead, while dealing with simulation, there are more steps to do:

1. At first $\hat{Y}$ is initialized with zeros
2. A basic Φ without bias and with missing values set to zero is generated
3. Iteratively, every $\hat{y}_i$ is obtained as the polynomial expansion of the ith row of Φ multiplied by $\widehat{\Theta}$
4. Then the missing values in Φ are filled with $\hat{y}_i$ calculated at step 3.

For both identification and validation, predicted and simulated outputs are generated, evaluated in terms of MSE and plotted to show the differences with respect to the real input-output curve.

Finally, validation tests from "Billings and Voon, 1986" are calculated on the validation datasets and plotted. A model is considered valid if all the tests are satisfied with a 95% confidence interval.

At every step of the algorithm, all the values are checked to find divergent models or deal with not expected values causing errors.

NARMAX IDENTIFICATION

After coding a FROE algorithm working with NARX identification the next step was the extension to the NARMAX case. To accomplish this task, there were some changes to be done.

First of all, Φ must be created including the new MA terms; this happens only from the second iteration of the FROE algorithm because at the first one only the process model is estimated.

At every iteration, the FROE algorithm is performed like in the NARX case:

- at the first one MA terms are not present (i.e. only a NARX model is obtained)
- at the next ones MA terms are included in Φ and the "variance to explain" parameter changes

Once obtained the temporary model, these further operations are executed:

1. $\hat{Y}$ is calculated:
   - at the first iteration, as $\hat{Y} = WG$
   - at the next ones is generated like in the NARX simulation described above but with the addition of the MA terms estimated as $\varepsilon_i = y_i - \hat{y}_i$
2. The residuals are calculated and the convergence criterion for NARMAX is checked and considered met if:
   - the maximum number of iterations set in the parameters has been reached or
   - $\frac{\sum_{i=0}^{yDim-1} |\varepsilon_i^K - \varepsilon_i^{K-1}|}{yDim} \leq th$ i.e. the average difference between the residuals of the actual and the previous iterations is within the threshold set in the parameters.
3. If convergence is not satisfied, the FROE algorithm is launched for the next iteration with the new residuals obtained from the temporary model, otherwise $\widehat{\Theta}$ and the printable model are returned

The new variance to explain is used starting from the 2nd iteration and it has been introduced to allow the usage of the best NARX model found for the first residuals calculation.

The convergence condition chosen is the one suggested in the publication cited above with the addition of the iteration limit, since in the cases where the convergence is not met within few iterations, going forward with the iterations wouldn't provide better results because it has been noticed a periodic recurrence of the same models.

The prediction task is performed as before while simulation is performed differently because, since the MA terms are related to the noise, it has been decided to generate them from a GWN process with mean and standard deviation estimated from the prediction residuals.

SRR IMPLEMENTATION

Since the forward regression performed in FROE is based on the Error Reduction Ratio, the chosen regressors are the ones that improve prediction capabilities of the model, it has been decided to apply the theory about the forward regression exploiting the Simulation error Reduction Ratio described in "An identification algorithm for polynomial NARX models based on simulation error minimization" (Piroddi, Spinelli, 2003).

It has been tried to expand the algorithm also to NARMAX models.

The only difference with the described implementation is in the FROE algorithm; the new regressor to choose is not anymore the one with the highest ERR but the one with the highest SRR which is calculated in this way:

- $\hat{Y}_i$ in simulation is retrieved using the model with the addition of the current regressor$_i$
- The MSSE$_i$ is calculated and used to in the SRR$_i$ formula $SRR_i = \frac{MSSE(M_{j-1}) - MSSE_i}{\frac{1}{N}\sum_{t=1}^{N} y^2(t)}$ where M$_{j-1}$ is the model without the current regressor.

In the NARMAX case, for simulation, it's necessary to generate a GWN, as explained before, and its parameters are estimated from the residuals obtained by the $\hat{Y}_i$ in prediction.

Since the SRR implementation is computationally demanding, a script that performs the SRRs calculation in parallel has been implemented in order to speed up the identification.

EXPERIMENTS

All the NARX experiments parameters are the results of various tries following this pattern:

1. Choose the identification and validation sets
2. Set an initial model with small lags and polynomial degree
3. Set a small variance to explain
4. Launch the algorithm and analyse the results
5. Three cases:
    a. If the results are an improvement, increase the variance to include one more regressor and go back to step 4.
    b. If the results are not an improvement and the model is not satisfactory, increase one or more of the lags and/or the polynomial degree, include or exclude the bias and go back to step 3.
    c. If the results are not an improvement and the previous model was satisfactory, keep the prev. model.

In the NARMAX experiments it has been followed the same pattern at the beginning, in order to find the optimal NARX model and then again changing only the noise maximum lag and the convergence threshold at the point 5b and the second variance to explain at the point 5a.

EXPERIMENT 1

UNIF DATA AS IDENTIFICATION SET

Parameters:

- nU: 8
- nY: 5
- nE: 0
- polynomial degree: 3
- bias included: yes
- variance to explain 1: 99.44%
- variance to explain 2: -

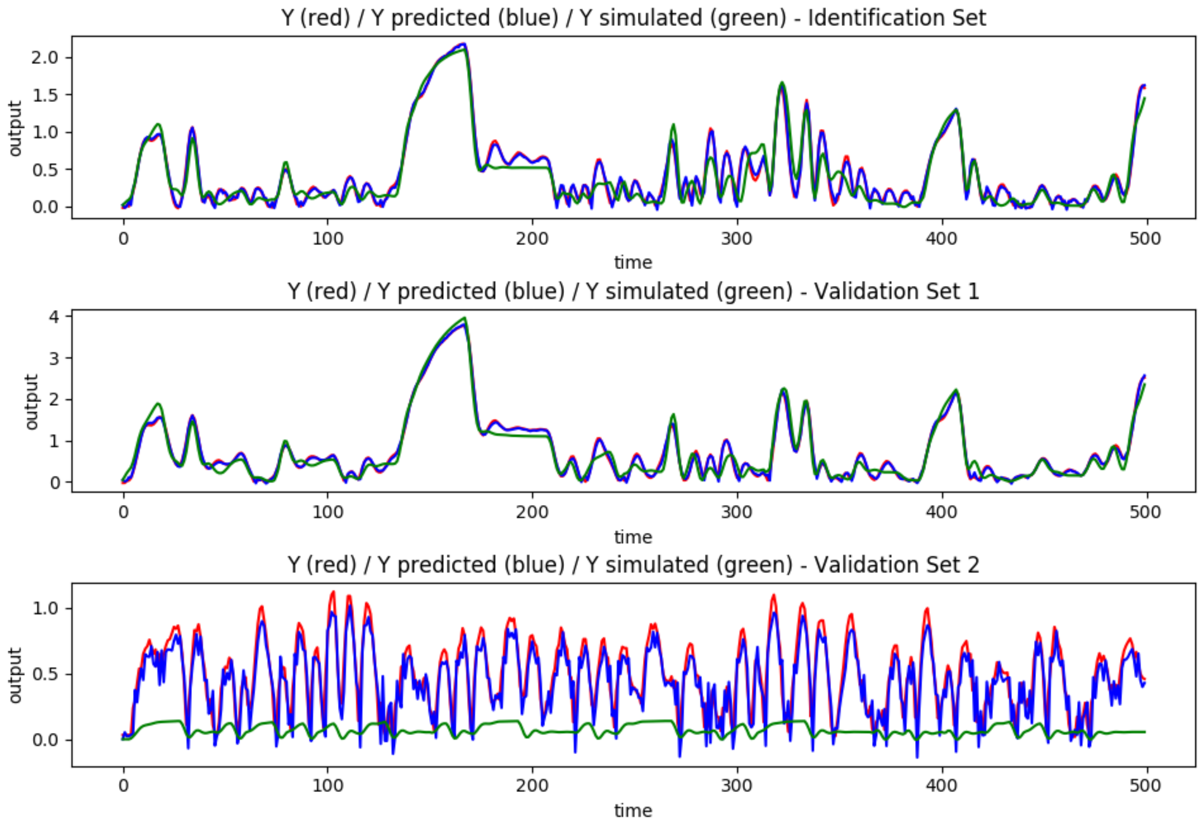
- NARMAX convergence threshold: -
- NARMAX max iterations: -
- Identification set [u11, z11]
- Validation set 1 [u12, z12]
- Validation set 2 [u3, z3]

Results:

Model: $y(t) = 0.12u(t-4) + 1.6y(t-1) - 0.97y(t-2) - 0.019y(t-4) - 0.02u(t-4)u(t-7) + 0.016u(t-4)u(t-5)$

$+0.017u(t-5)^2 - 0.1u(t-6)u(t-8) + 0.01u(t-3)u(t-8)y(t-3) - 0.007u(t-4)u(t-7)u(t-8)$

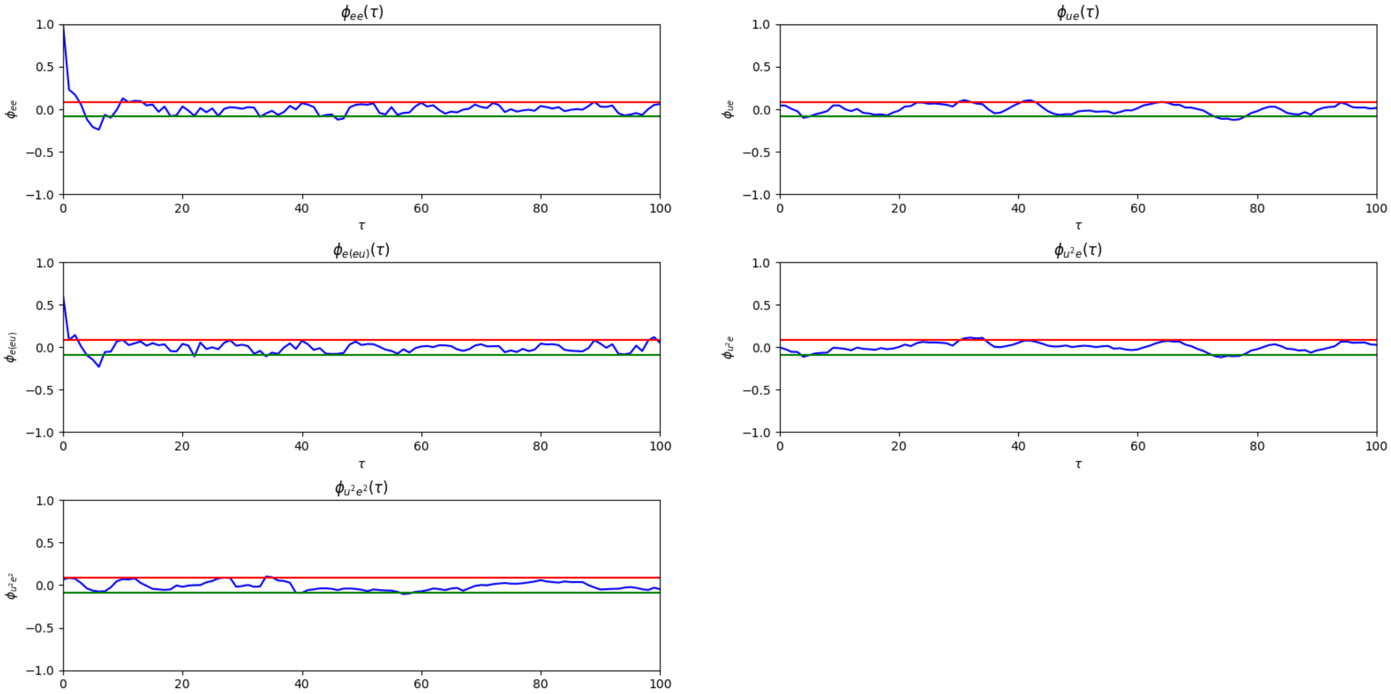
Variance Explained = 99.441%



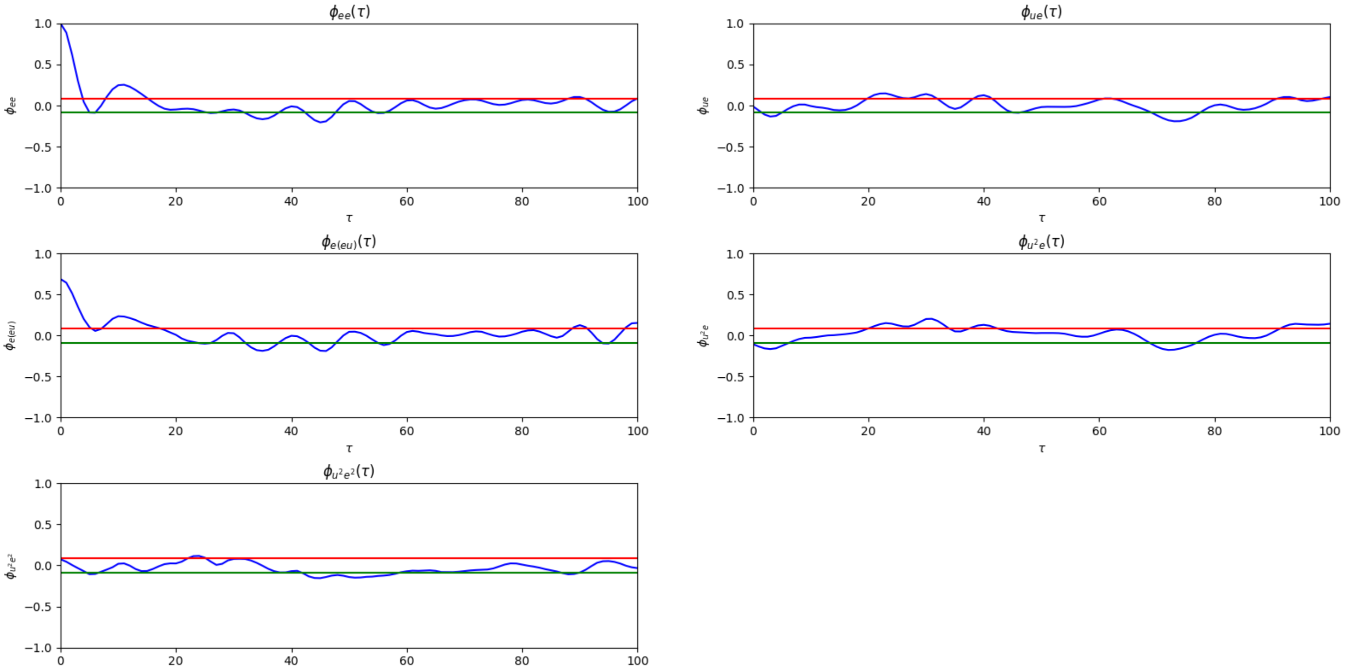
Identification set:	Validation set 1:	Validation set 2:
MSPE= 0.0018	MSPE= 0.0018	MSPE= 0.018
MSSE= 0.022	MSSE= 0.023	MSSE= 0.23

Model Validation:

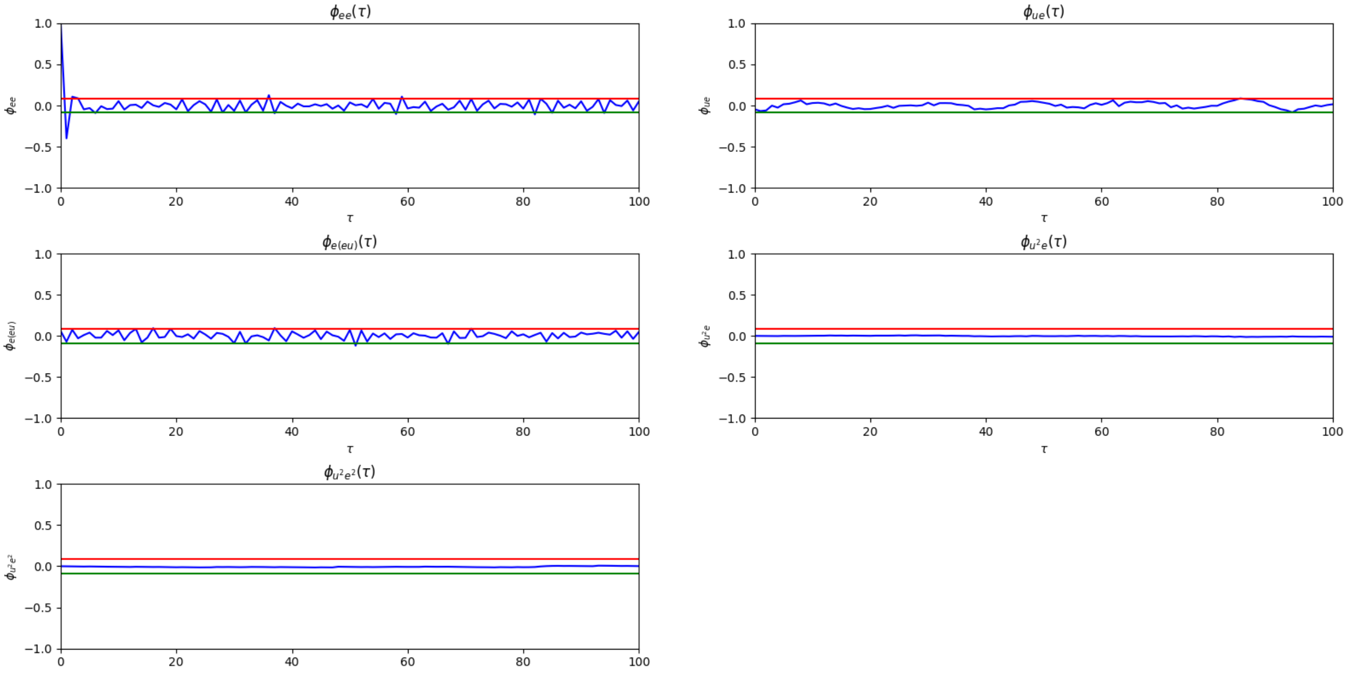
Prediction on validation set 1



Simulation on validation set 1



Prediction on validation set 2



Considerations:

The model found has shown very good predictive capabilities, confirmed by the validation tests.

The results are especially good on the validation set coming from the UNIF data (like the identification set) where the model is also able to perform a very good simulation of the output; the validation tests on the simulation is not as good as the ones performed on the prediction but can be accepted, since the expectations are lower.

It can be noticed a small loss of performances on the validation set coming from the PRBS data on the predictive side, the validation test is quite good but looking at the predictive curve it doesn’t follow very well the real one on peaks; on this dataset, the model in simulation is almost useless.

The results suggest that using the UNIF data for the identification, the algorithm catches very well the system’s dynamic fed by the specific kind of input but it doesn’t catch the real one, since the performances are worse on a dataset with a different kind of input.

EXPERIMENT 2

PRBS DATA AS IDENTIFICATION SET

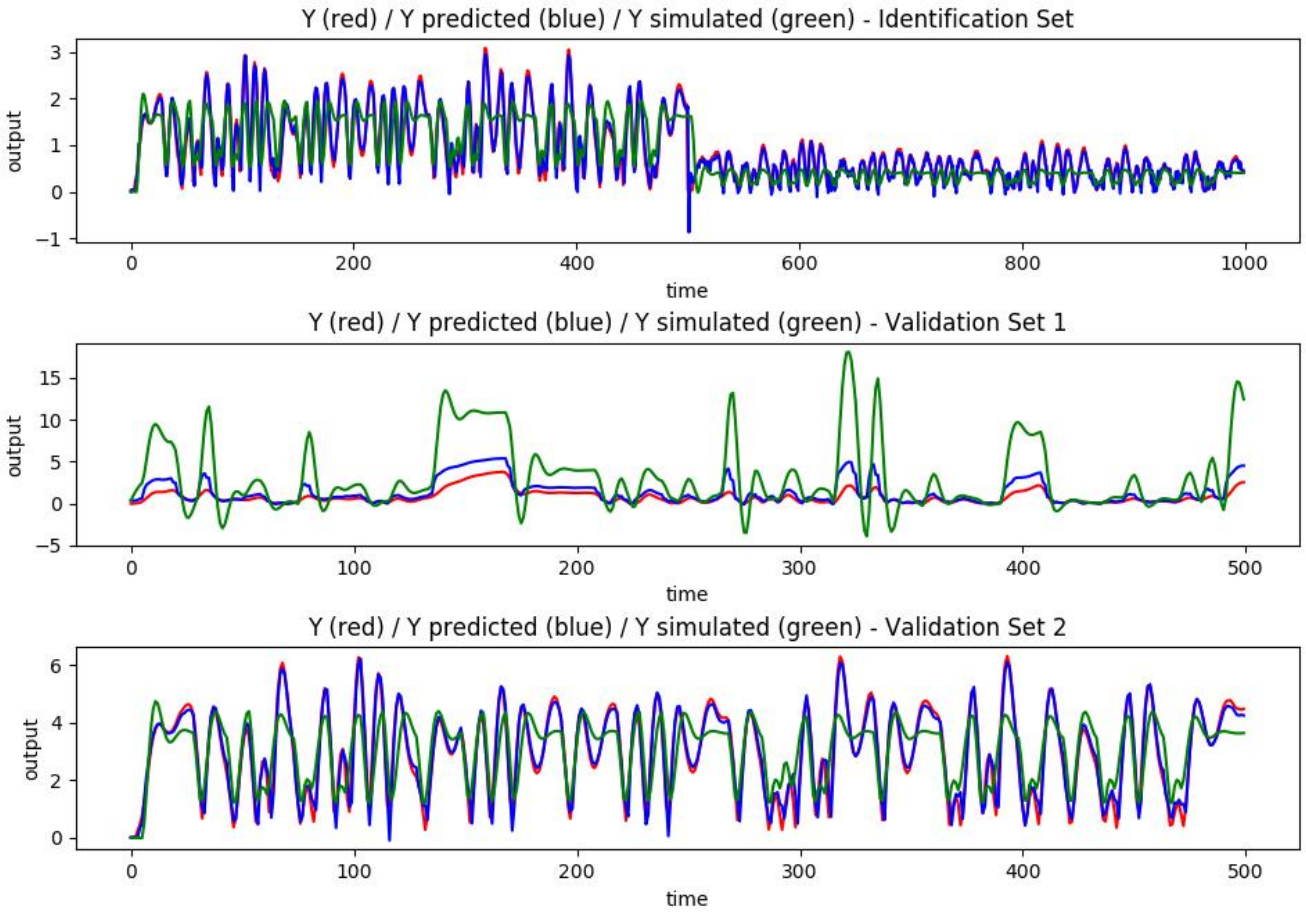
Parameters:

- nU: 8
- nY: 5
- nE: 0
- polynomial degree: 3
- bias included: yes
- variance to explain 1: 98%
- variance to explain 2: -
- NARMAX convergence threshold: -
- NARMAX max iterations: -
- Identification set [u1°u3, z1°z3]
- Validation set 1 [u12, z12]
- Validation set 2 [u2, z2]

Results:

Model: $y(t) = 1.4y(t-1) - 0.68y(t-2) + 0.13u(t-4)u(t-7) - 0.0007y(t-4) - 0.24u(t-6)^2$

Variance Explained = 98.16%


Y (red) / Y predicted (blue) / Y simulated (green) - Identification Set

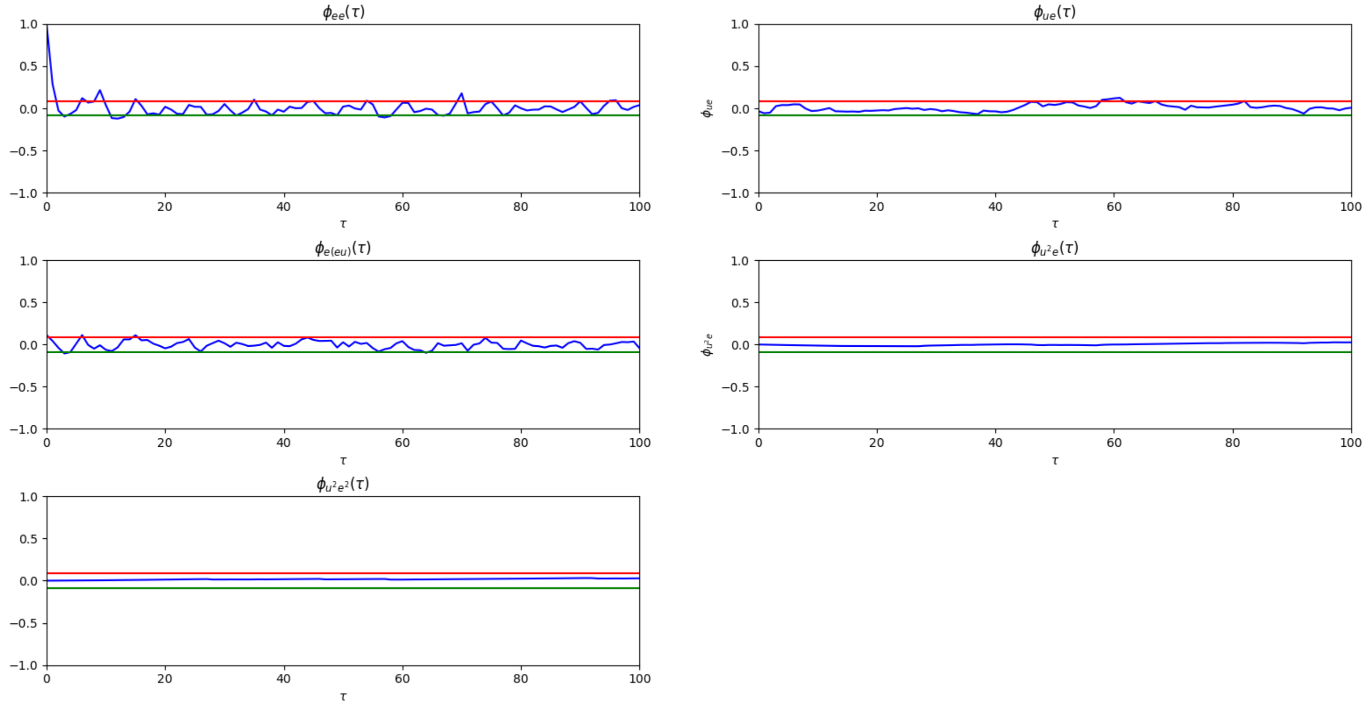
Y (red) / Y predicted (blue) / Y simulated (green) - Validation Set 1

Y (red) / Y predicted (blue) / Y simulated (green) - Validation Set 2

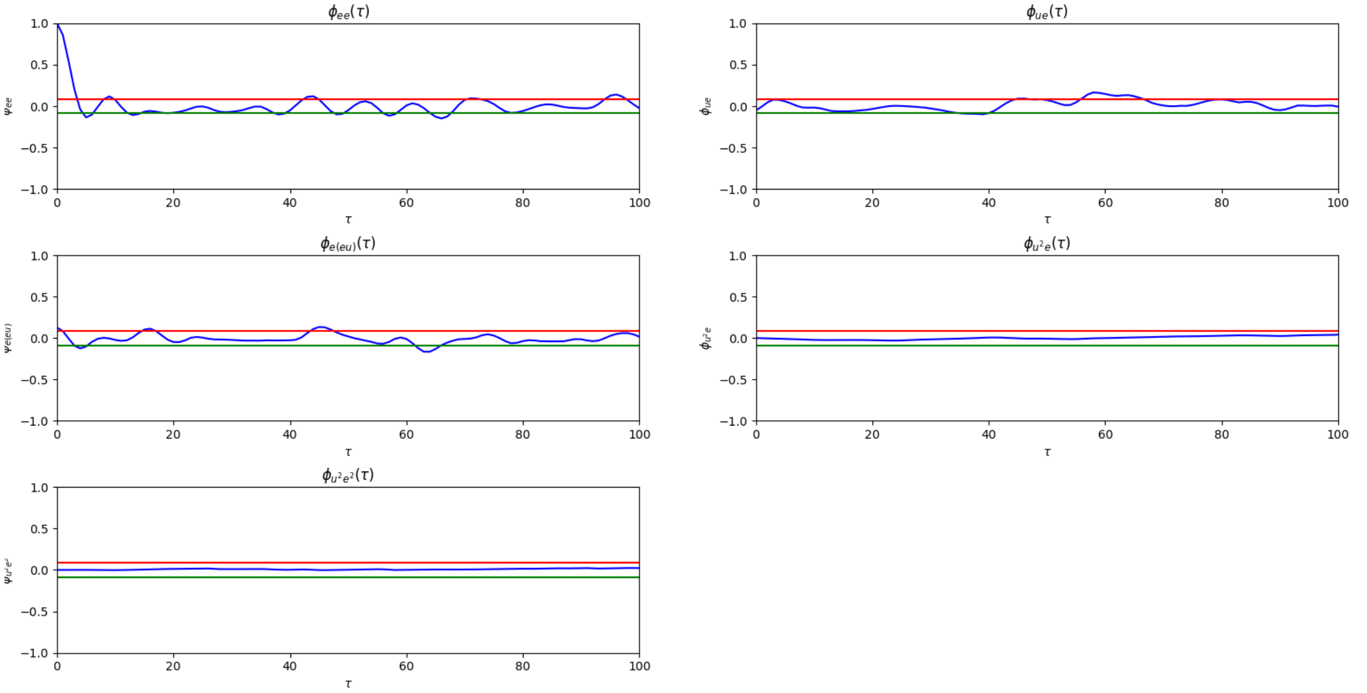
Identification set:	Validation set 1:	Validation set 2:
MSPE= 0.022	MSPE= 0.78	MSPE= 0.087
MSSE= 0.18	MSSE= 17.31	MSSE= 0.8

Model Validation:

Prediction on validation set 2



Simulation on validation set 2



Considerations:

Using a single PRBS dataset leaded only to overfitting models, working very well on the dataset chosen for the identification ([u1, z1] [u2, z2] [u3, z3] independently) and very badly on the others; for this reason, it has been necessary to concatenate two datasets, in order to better understand the dynamics.

This dataset is [np.append(u1, u3), np.append(z1, z3)] (python syntax).

The result is quite similar to the one obtained in the first experiment on the opposite side: the model obtained works well for PRBS data and poorly for UNIF data, however predictive performances are not as good as the ones obtained previously but there has been a significant improvement in the simulation ones on PRBS data.

The results of these first two experiments suggest that it might be useful to use data coming from both UNIF and PRBS datasets to catch a more robust model.

EXPERIMENT 3

PRBS AND UNIF DATA AS IDENTIFICATION SET

Parameters:

- nU: 8
- nY: 7
- nE: 3
- polynomial degree: 3
- bias included: yes
- variance to explain 1: 99.6%
- variance to explain 2: 99.38%
-

- NARMAX convergence threshold: 0.1
- NARMAX max iterations: 10
- Identification set [u12 + u1, z12 + z1]
- Validation set 1 [u11, z11]
- Validation set 2 [u3, z3]

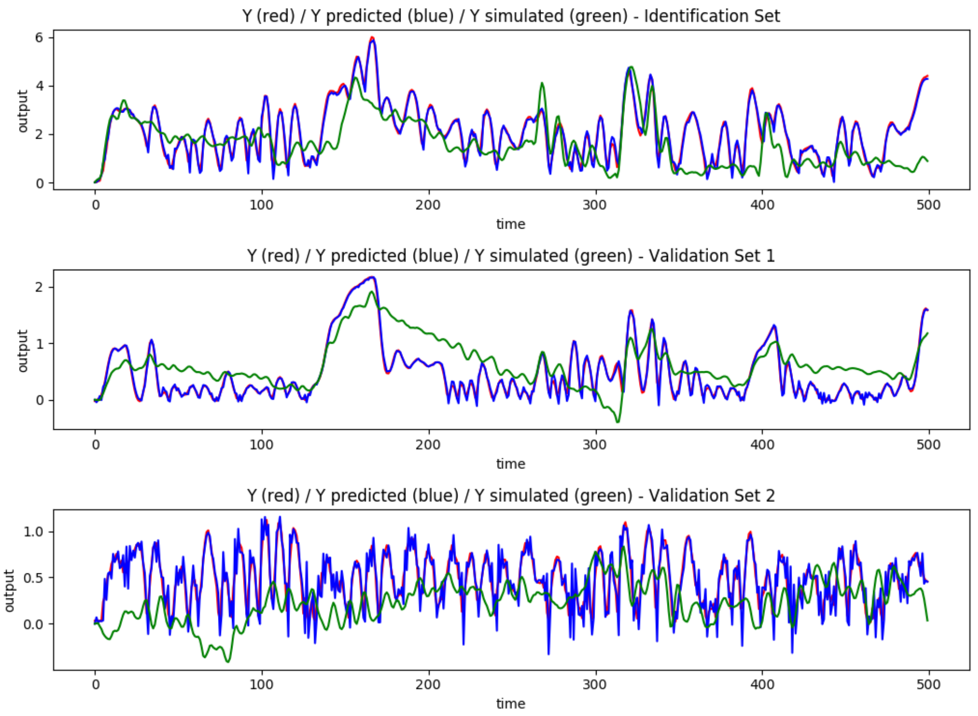
Results:

Model: $y(t) = 2.18y(t-1) - 1.76y(t-2) + 0.68y(t-3) - 0.42y(t-3) - 0.14y(t-4) + 0.15e(t-3)$

$+0.05u(t-1)^2e(t-1) + 0.005u(t-2)^2u(t-4) - 0.0045u(t-2)u(t-3)y(t-6) + 0.33u(t-4)e(t-1)^2$

$-0.13y(t-6)e(t-2)^2$

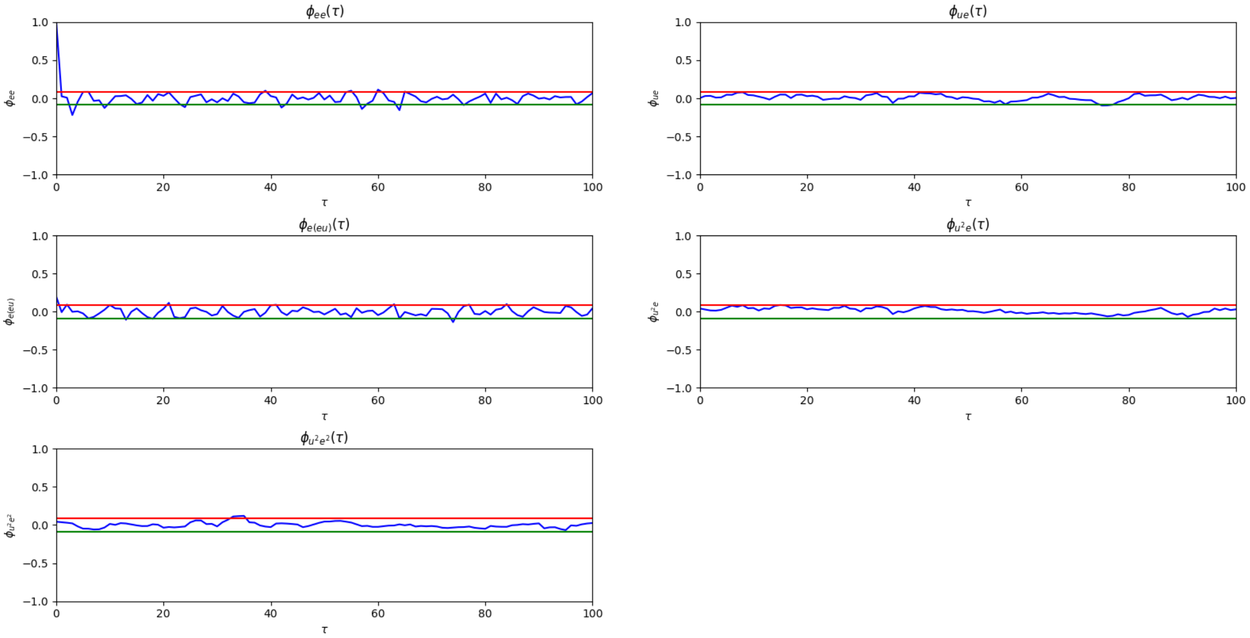
Variance Explained = 99.385%



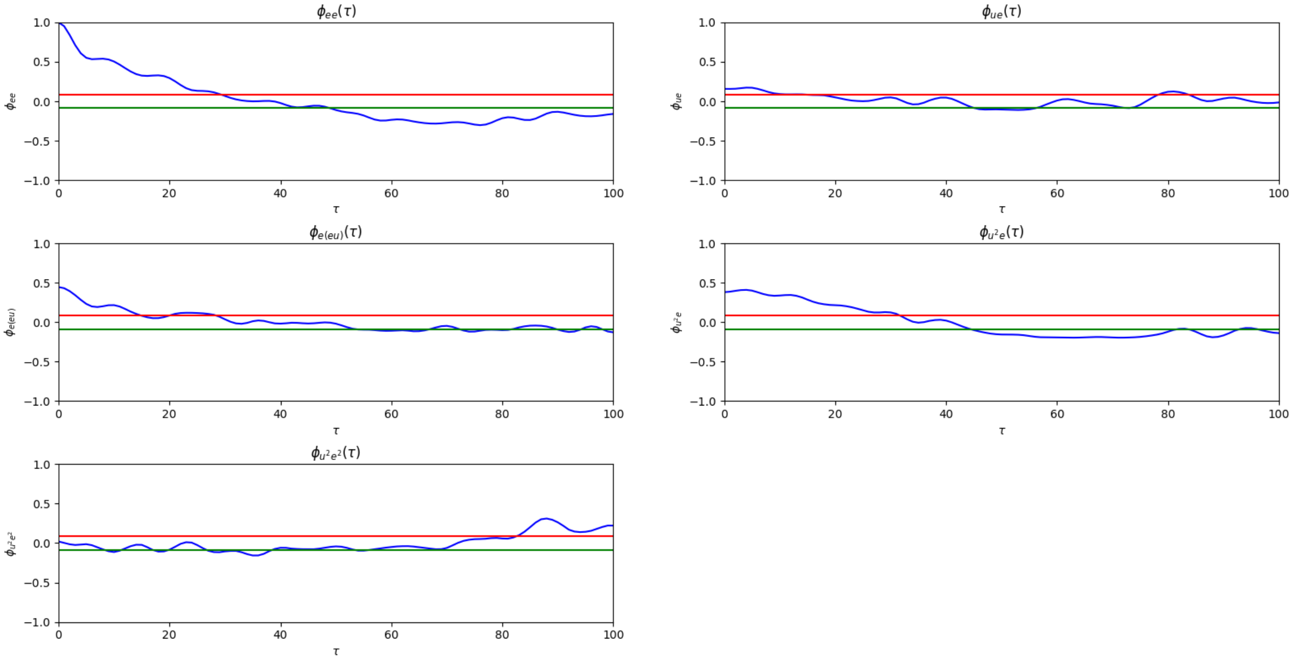
Identification set:	Validation set 1:	Validation set 2:
MSPE= 0.028	MSPE= 0.003	MSPE= 0.027
MSSE= 1.17	MSSE= 0.16	MSSE= 0.17

Model Validation:

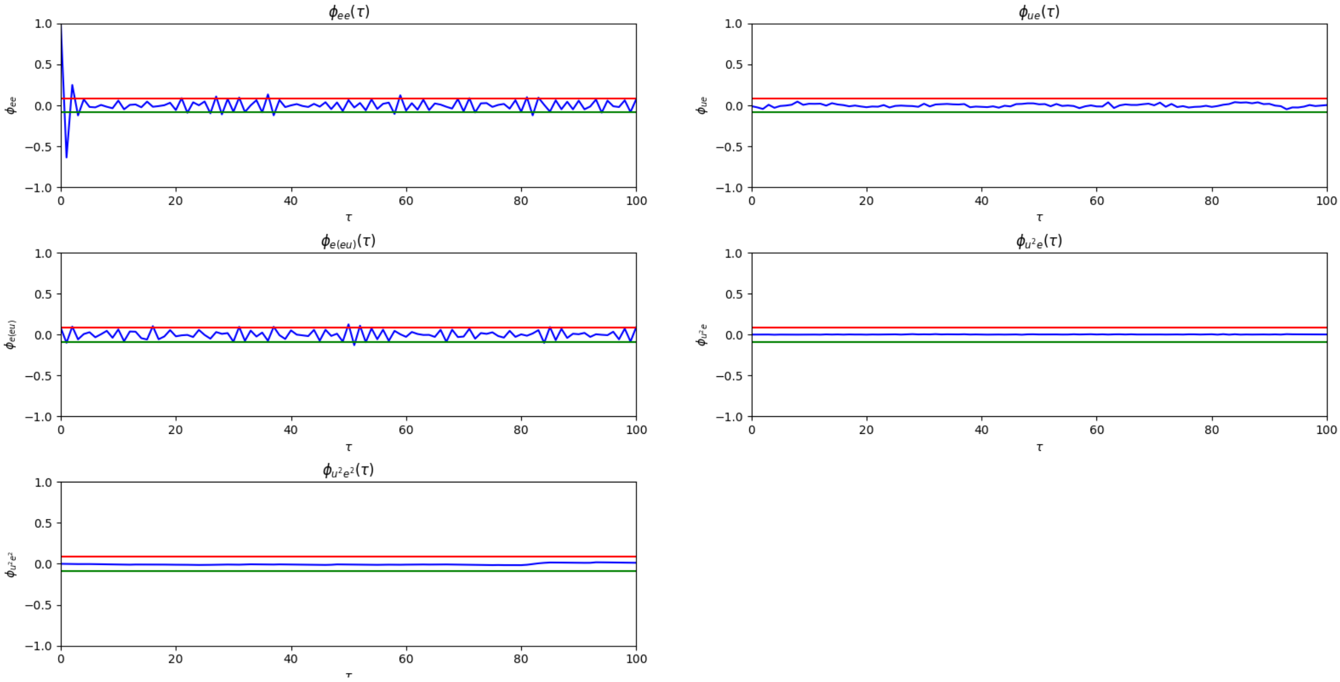
Prediction on validation set 1



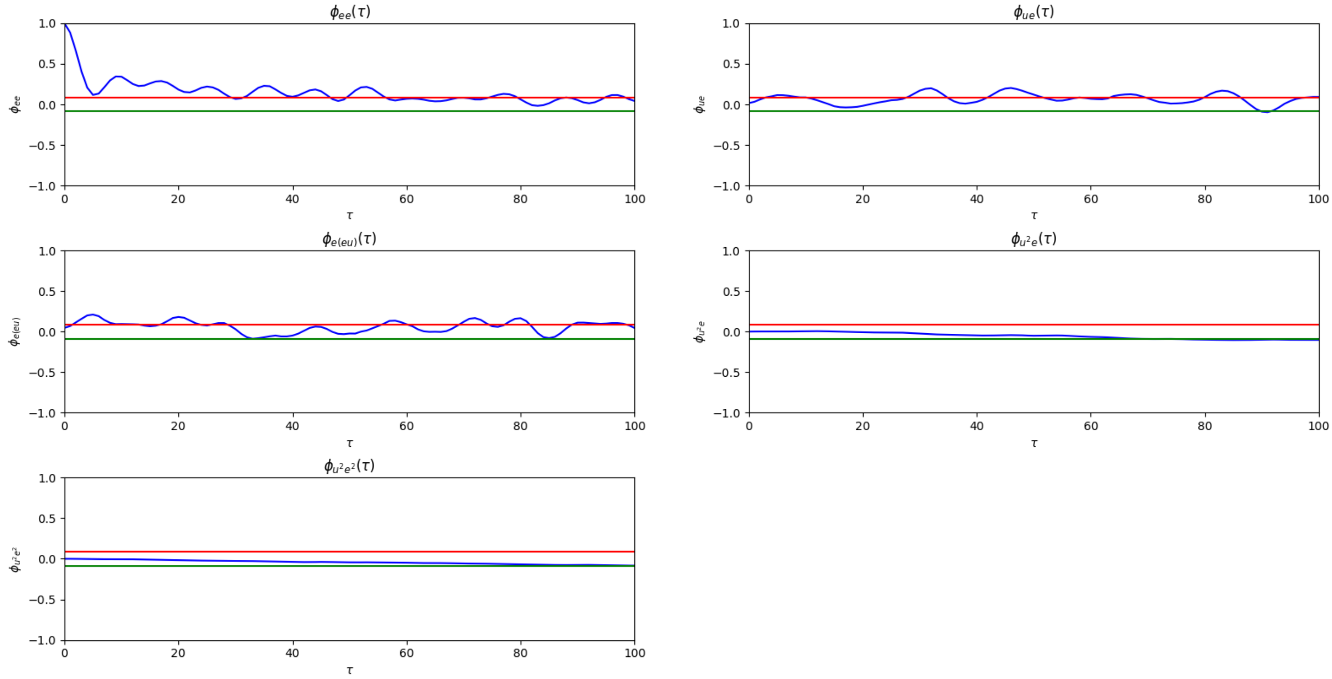
Simulation on validation set 1



Prediction on validation set 2



Simulation on validation set 2



Considerations:

For this experiment it has been used, as identification set, two datasets summed: one from UNIF data and one from PRBS in order to try to catch the dynamics of both.

The results are good in terms of prediction, confirmed also by the validation tests.

For simulation purposes, the resulting model could be seen as a sort of trade-off between the ones resulting from the first two experiments:

- Comparison to the experiment 1:
    - the PRBS validation set simulation curve follows better the dynamics
    - the UNIF validation set simulation curve loses performances
- Comparison to the experiment 2:
    - the PRBS validation set simulation curve loses performances
    - the UNIF validation set simulation curve follows better the dynamics (also in prediction)

Both simulation validation tests suggest that some regressors are missing from the model but by increasing the variance to explain the algorithm didn’t caught them.

EXPERIMENT 4

PRBS AND UNIF DATA AS IDENTIFICATION SET

Parameters:

- nU: 8
- nY: 6
- nE: 4
- polynomial degree: 2
- bias included: yes
- variance to explain 1: 99%
- variance to explain 2: 98.46%
-

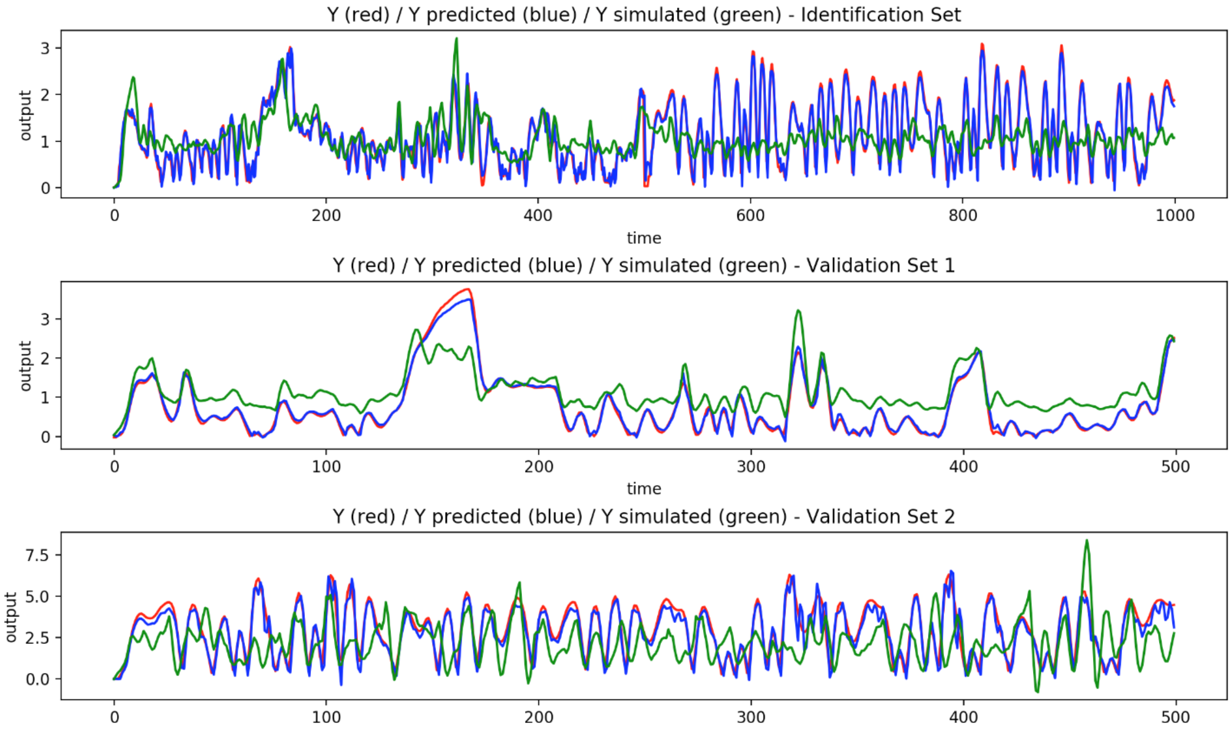
- NARMAX convergence threshold: 0.1
- NARMAX max iterations: 10
- Identification set [(u11+u3)°u1, (z11+z3)°z1]
- Validation set 1 [u12, z12]
- Validation set 2 [u2, z2]

Results:

Model: $y(t) = 1.64y(t-1) - 0.79y(t-2) + 0.17y(t-5) + 0.036u(t-4)u(t-7) - 0.069y(t-1)y(t-5)$

$+0.26y(t-1)e(t-2) + 0.33e(t-1)^2$

Variance Explained = 98.51%


Y (red) / Y predicted (blue) / Y simulated (green) - Identification Set

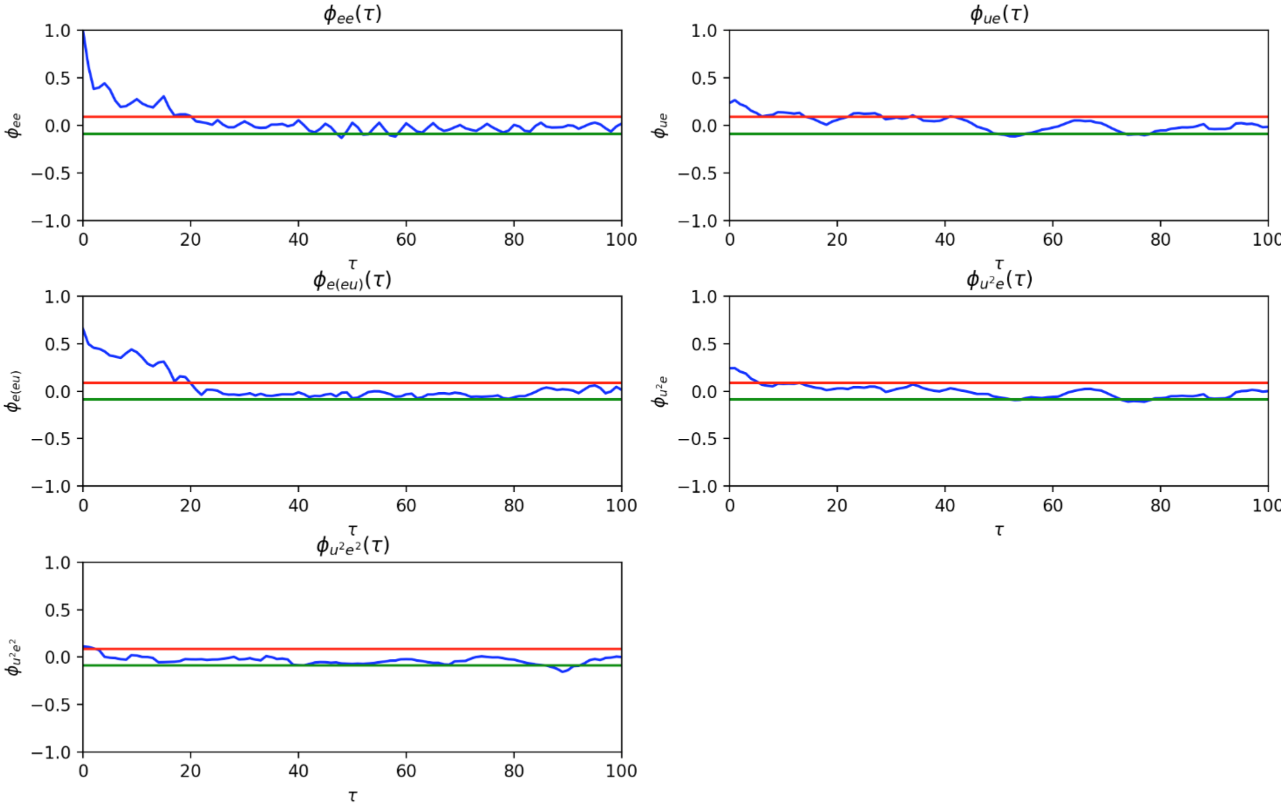
Y (red) / Y predicted (blue) / Y simulated (green) - Validation Set 1

Y (red) / Y predicted (blue) / Y simulated (green) - Validation Set 2

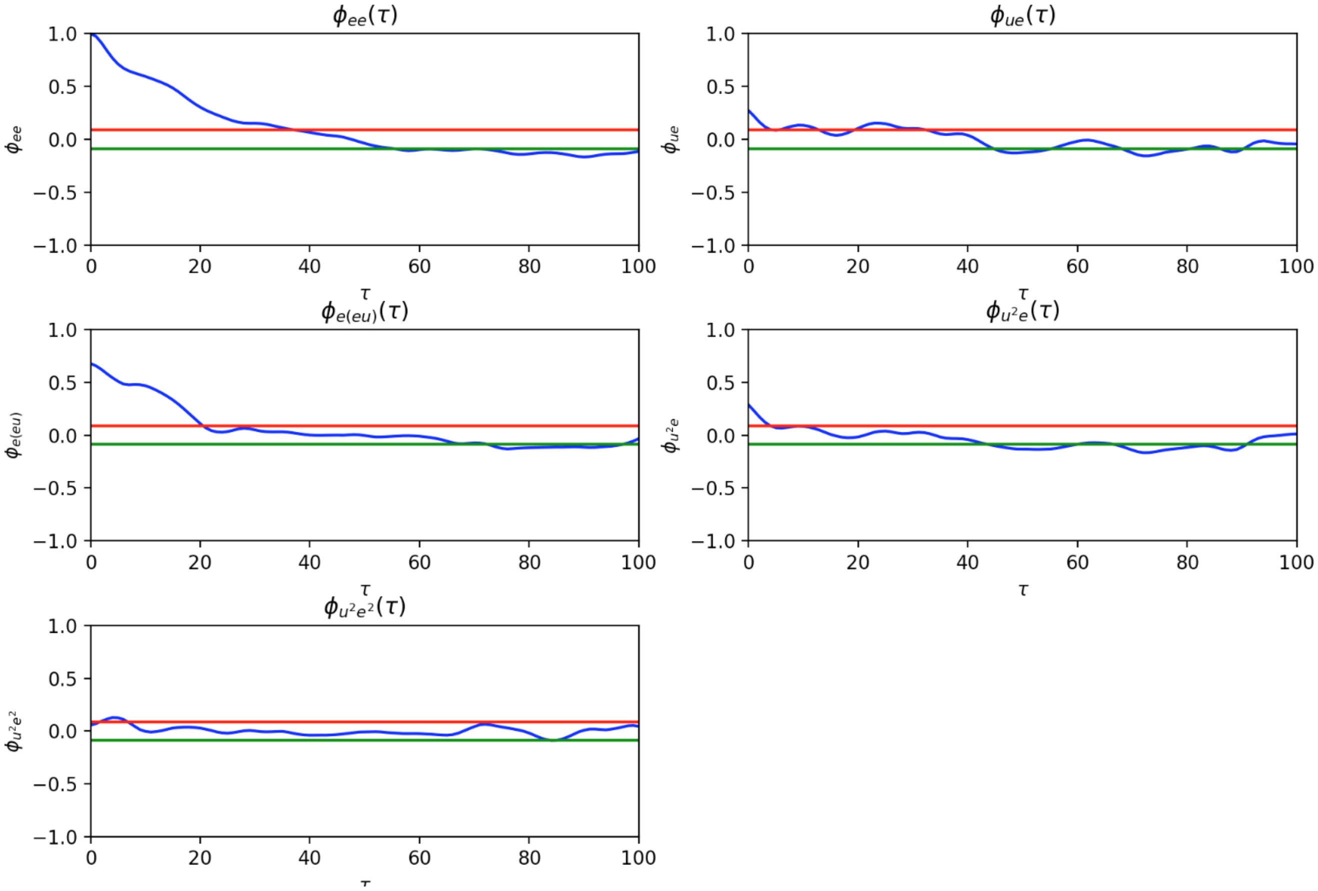
Identification set:	Validation set 1:	Validation set 2:
MSPE= 0.025	MSPE= 0.0065	MSPE= 0.23
MSSE= 0.39	MSSE= 0.35	MSSE= 3

Model Validation:

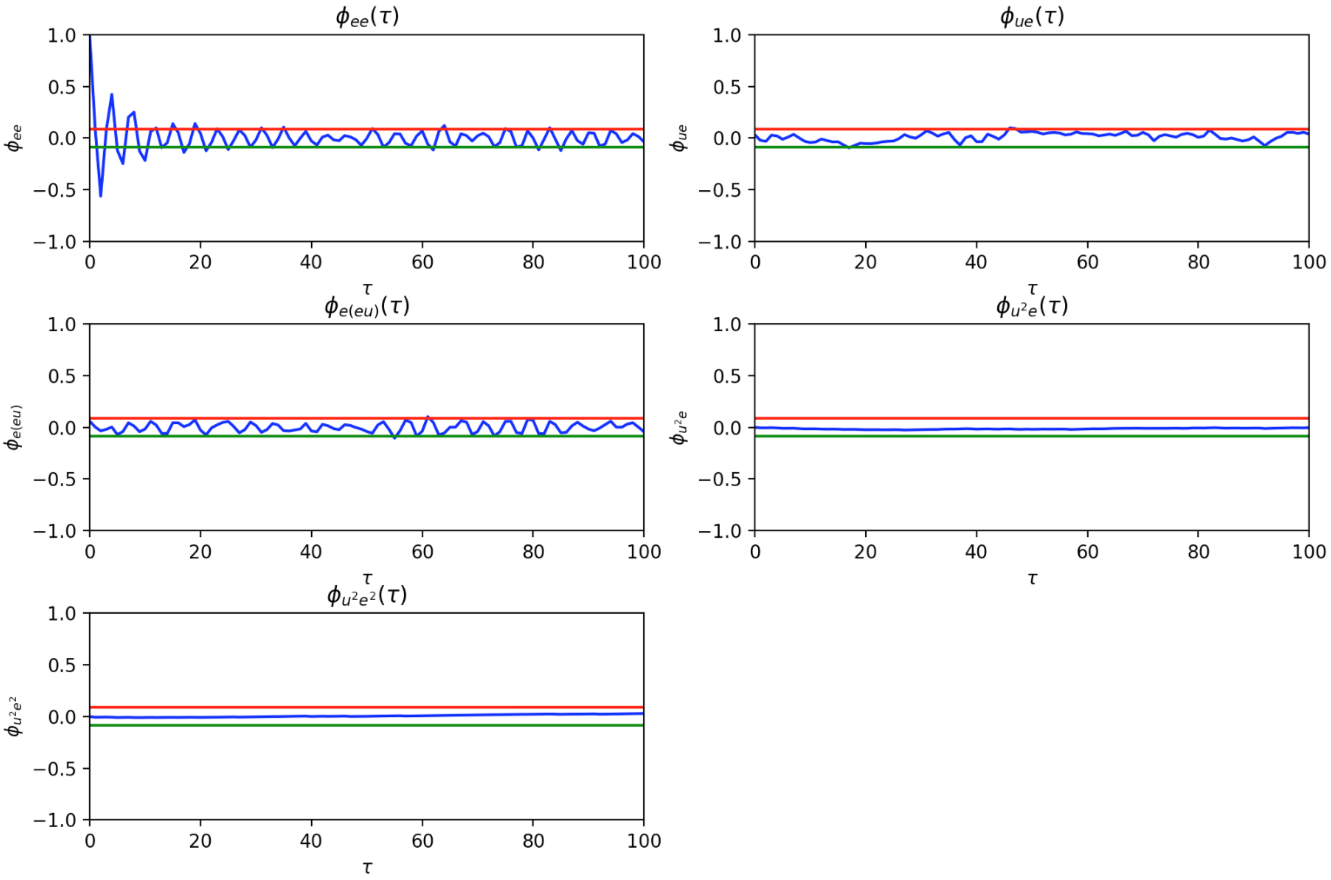
Prediction on validation set 1



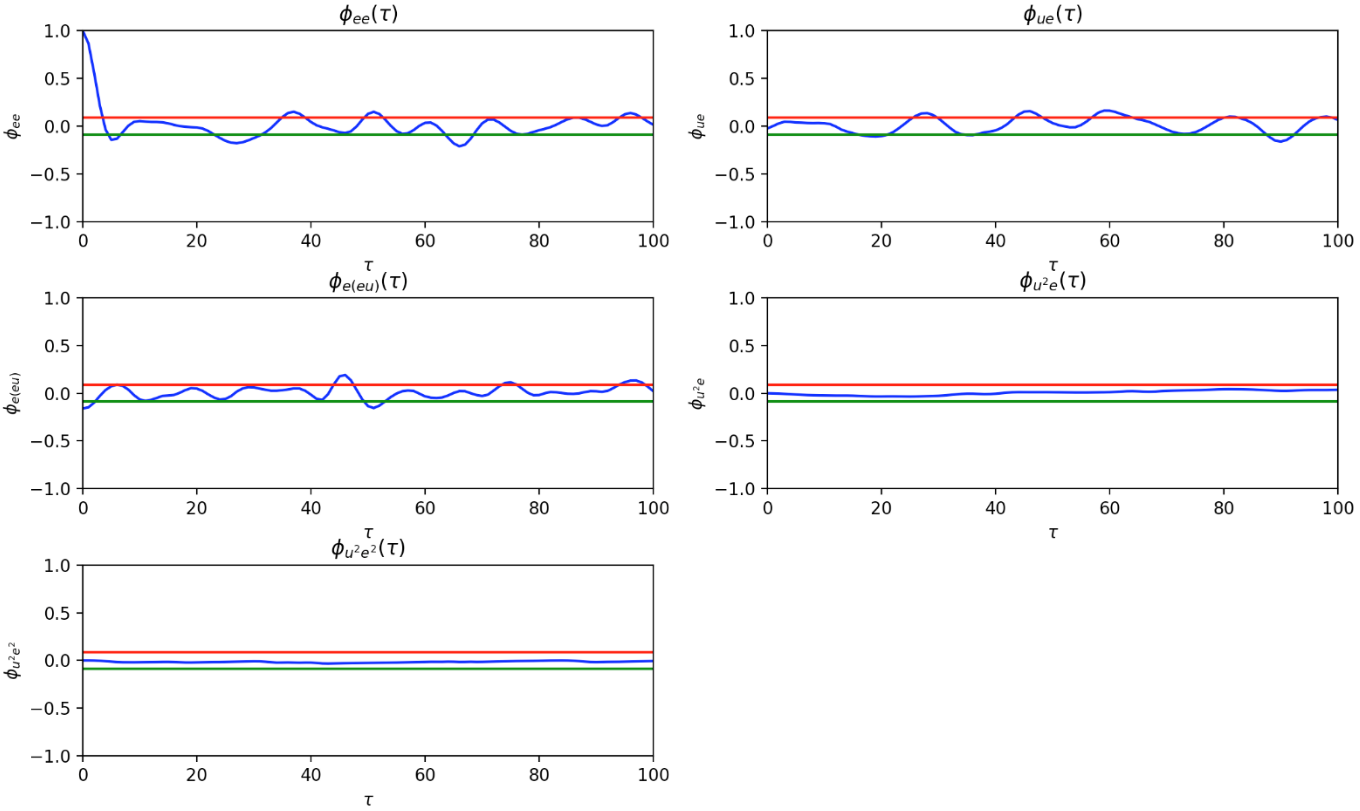
Simulation on validation set 1



Prediction on validation set 2



Simulation on validation set 2



Considerations:

In this experiment, it has been tried to use both techniques used in experiments 2 and 3 (sum and concatenation) for the identification set composition.

Comparing the results with the ones of the third experiment it can be noticed a small loss of performances in terms of prediction but, in terms of simulations there is a good improvement for both type of datasets since the simulation output curves follow better the real ones.

Also in this case the validation tests, especially the simulation ones, suggests a lack of terms for small lags but the correct parameters to allow the algorithm to find them haven't been found.

CONCLUSIONS

After all the experiments, it has been noticed that most models found with FROE were good for prediction purposes but poorly performant in simulation so, the great part of the work was dedicated to find a good simulation model.

At first, a very good model for prediction and simulation has been found, the one in the first experiment, but the real problem has arisen noticing that its simulation performances on PRBS data were far away from good, suggesting that the real dynamic of the process was not found. The same problem, but reversed, arises in the second experiment, that's why it has been tried to combine the two different datasets to find a more general and robust model.

A good trade-off has been found in the last experiment, mixing all the techniques used in the previous ones.

From the experiments results is evident that NARX models are enough for prediction purposes on this dataset, but for an acceptable generical simulation model it's necessary to include the MA part.

It must be remarked that the simulation part, when dealing with NARMAX, is highly dependent to the prediction one, since the generated noise will have mean and variance estimated from the prediction residuals and, if they are wrong, the simulation will deteriorate in performances.

Most of the times, the lack of terms highlighted by the validation tests, was not possible to be compensated by increasing the variance to explain and consequently the number of regressors since this procedure leaded to overfitting models with poor performances on the validation sets; the alternative was to modify the initial model parameters, increasing the lag limits or the polynomial degree.

From the experiments, it has been noticed that regressors with a degree superior to the fourth were rarely chosen and with a small importance, suggesting that the correct polynomial degree to choose is "4"; however, it has not been tried to run the algorithm with polynomial degrees superior than 6 since, because the "course of dimensionality", the available hardware was not enough to run the algorithm in a reasonable time.

Overall the FROE algorithm has been found very efficient in computational terms and very performant for prediction but, since the forward selection criteria is the Error Reduction Ratio and not the Simulation error Reduction Ratio, obtaining a good model for simulation is very hard; that's why the necessity to implement a version exploiting the SRR has arisen.

SRR REMARKS

For simplicity, the terms and the variable “variance to explain”/“variance explained” are kept but they are interpreted just as thresholds.

Only few experiments have been performed because of the computational load, the reported one is the best obtained.

EXPERIMENT SRR

PRBS DATA AS IDENTIFICATION SET

Parameters:

- nU: 7
- nY: 5
- nE: 0
- polynomial degree: 2
- bias included: yes
- variance to explain 1: 90,51%
- variance to explain 2: -
- NARMAX convergence threshold: -
- NARMAX max iterations: -
- Identification set [u1, z1]
- Validation set 1 [u12, z12]
- Validation set 2 [u2, z2]

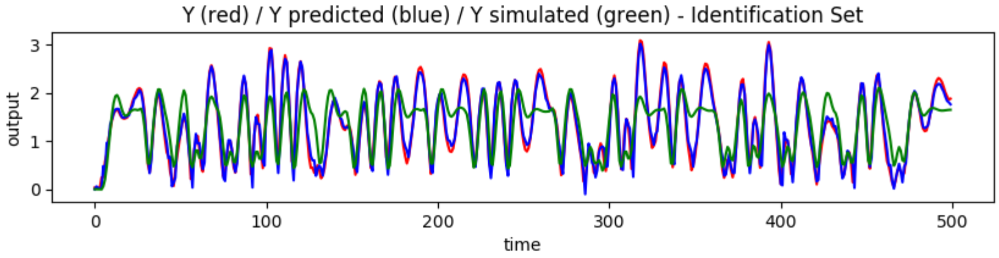
Results:

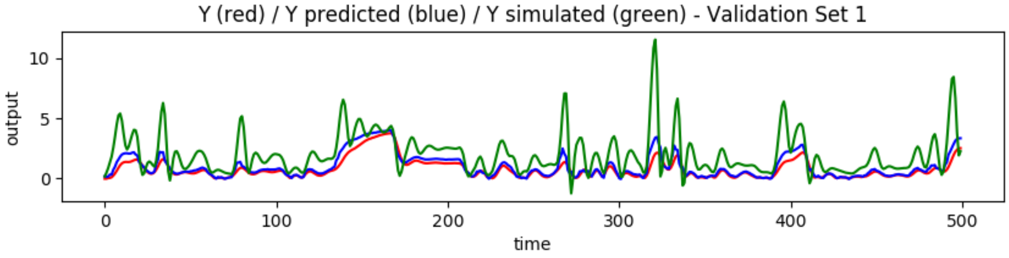
Model: $y(t) = 1.8y(t-1) - 1.05y(t-2) + 0.22y(t-4) - 0.008u(t-1)y(t-1) + 0.007u(t-2)u(t-5)$

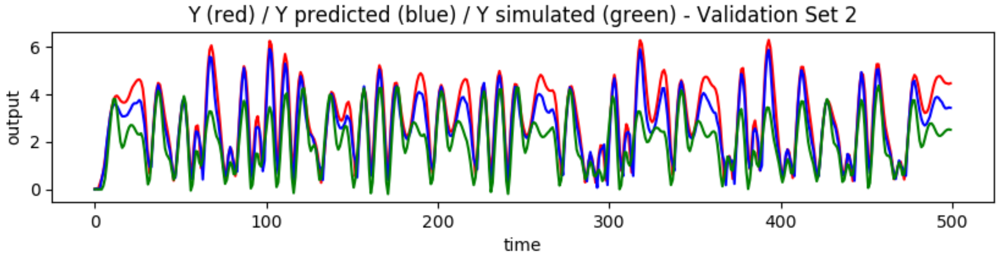
$+0.06u(t-3)u(t-5) + 0.02u(t-4)u(t-6) + 0.006u(t-4)y(t-2) + 0.11u(t-5)u(t-7)$

$- \ 0.03\mathrm{u}(\mathrm{t}-7)^2 - 0.03\mathrm{y}(\mathrm{t}-1)\mathrm{y}(\mathrm{t}-4) - 0.08\mathrm{y}(\mathrm{t}-1)\mathrm{y}(\mathrm{t}-5) + 0.05\mathrm{y}(\mathrm{t}-2)\mathrm{y}(\mathrm{t}-5)$

Variance Explained = 90,53%



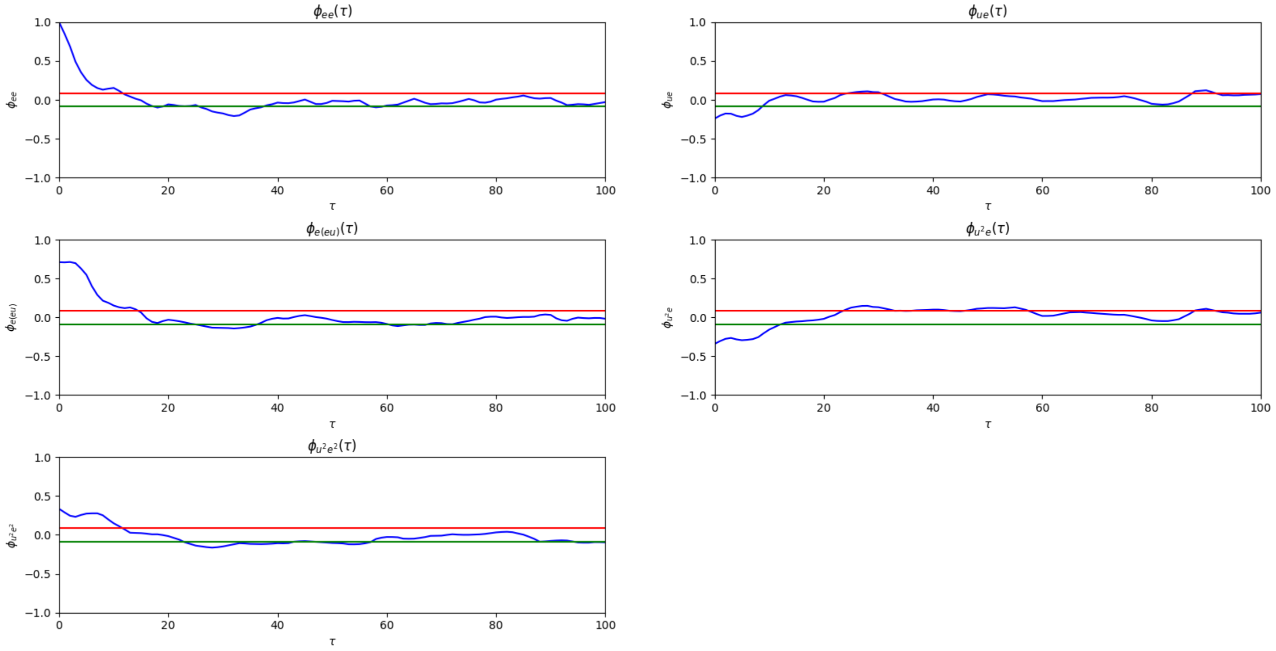




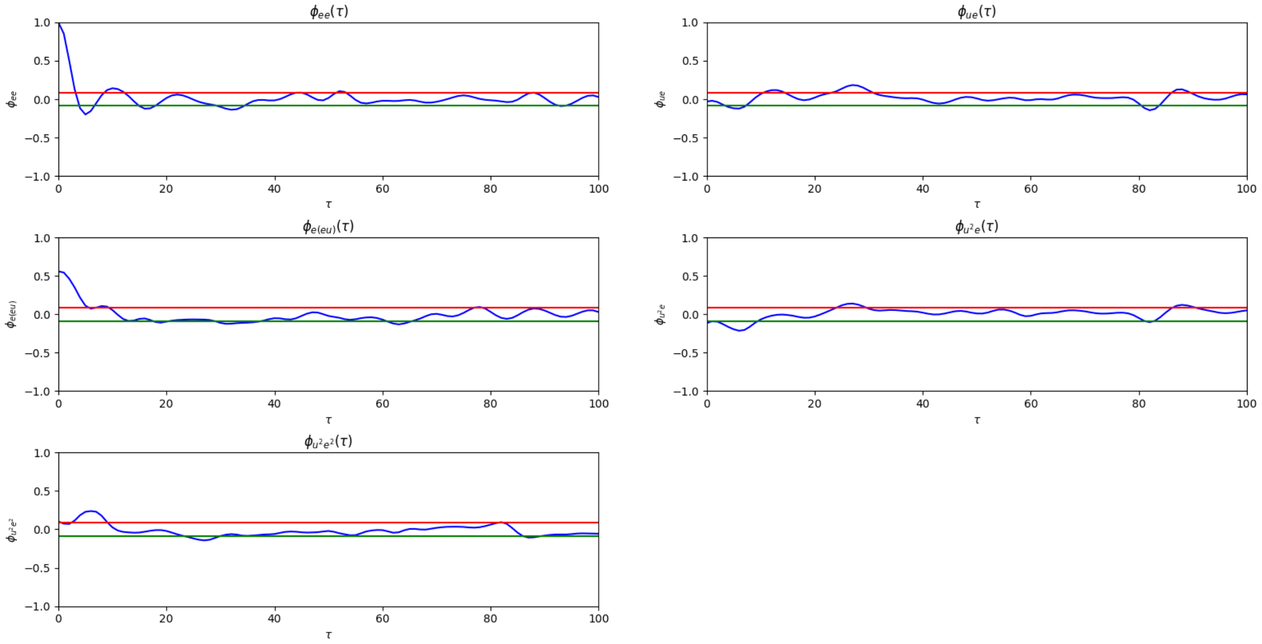
Identification set:	Validation set 1:	Validation set 2:
MSPE= 0.016	MSPE= 0.13	MSPE= 0.24
MSSE= 0.21	MSSE= 3.3	MSSE=1.5

Model Validation:

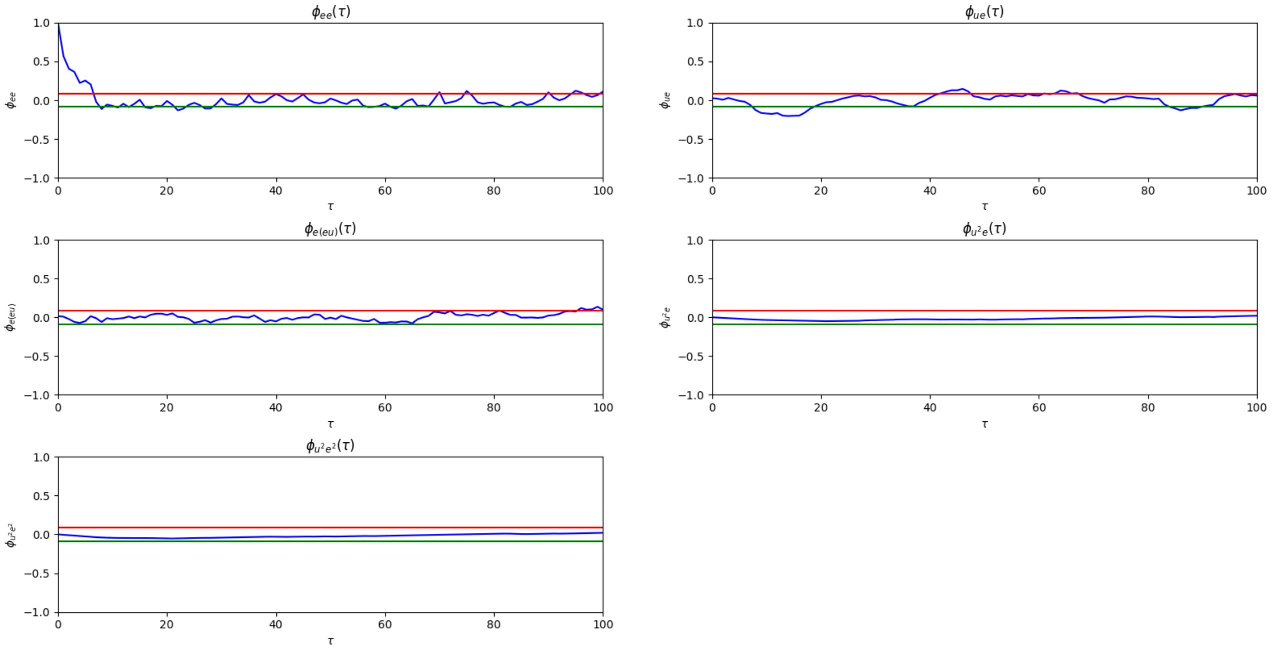
Prediction on validation set 1



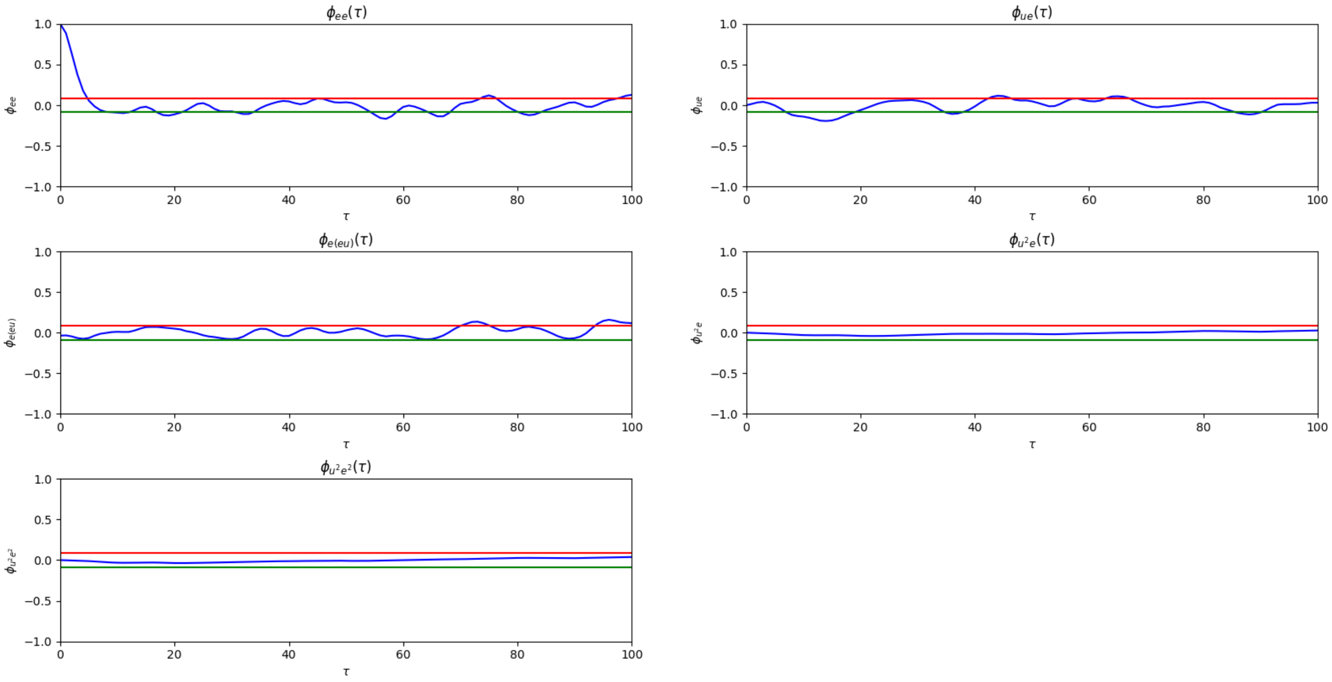
Simulation on validation set 1



Prediction on validation set 2



Simulation on validation set 2



Considerations:

There is an evident lack of regressors, but it’s still noticeable that:

- It has been obtained a result comparable with the one of the experiment 2 with only [u1,z1] as identification set, with even better performances on the UNIF validation sets; as mentioned before, the results obtained with this dataset alone, using the ERR criterion, leaded only to overfitting models, proving the validity of the SRR criterion, the results can be improved but to test models with higher polynomial degree and lags requires a better hardware.