Simple Heuristic Approach for training of Type-2 NEO-Fuzzy Neural Network

Yancho Todorov, Margarita Terziyska

yancho.todorov@iit.bas.bg, mterziyska@bas.bg

- **The idea for a NEO-fuzzy neuron has been introduced in late 90's by Yamakawa, but it has not received a wide attention by the scientific community.**

- **The main advantages of the NEO-fuzzy concept are:**

  - the low-level integration of fuzzy rules into a single neuron model and larger neural network structures.

  - tightly coupling of the learning and fuzzy reasoning rules into connectionists structures.

  - among the most important advantages of the NFN are: the high approximation properties, the computational simplicity and the possibility of finding the global minimum of the learning criterion in real time .

- **The NEO-Fuzzy Neuron concept enables the possibility to model complex dynamics with less computational effort, compared to classical Fuzzy-Neural Networks.**

- **Unfortunately, its application in purpose to process modeling and control under uncertainties/ data variations, have not been studied yet.**

- **In the presented approach, the conventional concept is extended with Type-2 Interval Fuzzy Logic in order to be achieved overall robustness of the proposed model**

- **Thus, introducing Type-2 Fuzzy Logic in purpose of handling uncertain variations is beneficial for modeling different plant processes with complex dynamics.**

- **To overcome some deficiencies in the classical gradient learning approach, a simple heuristic approach is introduced.**

✓The NEO-Fuzzy neuron is similar to a 0-th order Sugeno fuzzy system, in which only one input is included in each fuzzy rule, and to a radial basis function network (RBFN) with scalar arguments of basis functions

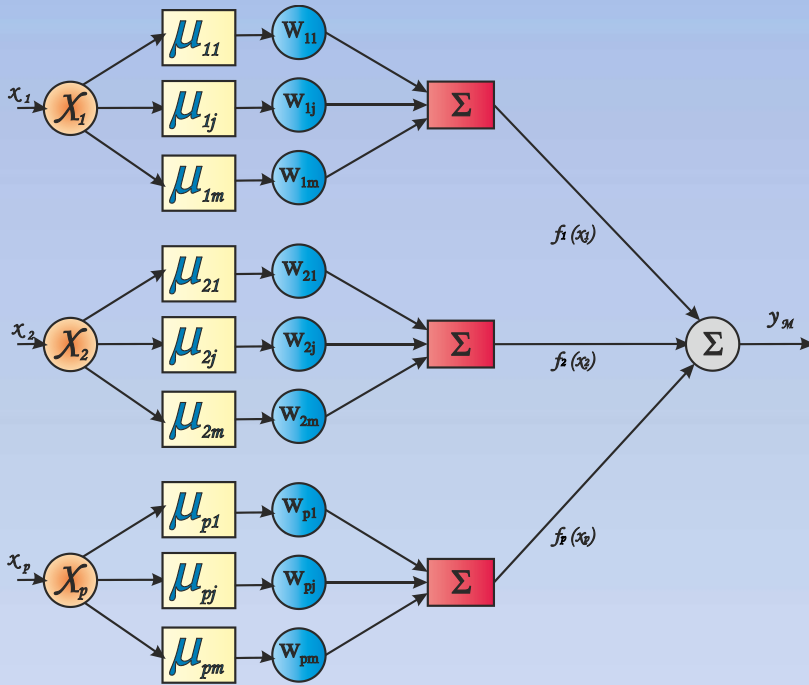✓In fact the NFN network is a multi-input single-output system – MISO !

✓The NEO-Fuzzy neuron has a nonlinear synaptic transfer characteristic.

✓The nonlinear synapse is realized by a set of fuzzy implication rules.

✓The output of the NEO-Fuzzy neuron is obtained by the following equation:

$$f(x) = \sum_{j=1}^{m} \mu_j(x(k)) w_j$$
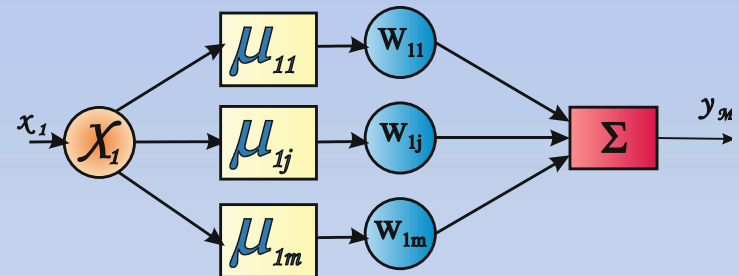
Type-2 Neo-fuzzy network

- The MISO NEO-fuzzy neural network topology can be represented as:

$$\hat{y}(k) = f(x(k))$$
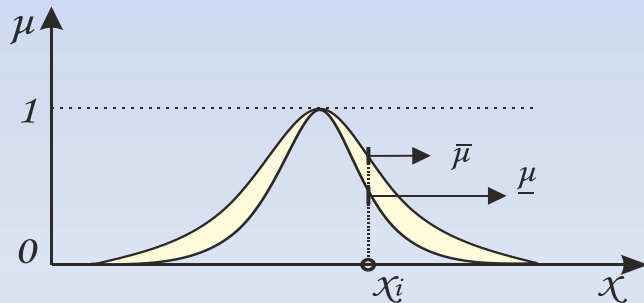
where **x(k)** is an input vector of the states in terms of different time instants.

- Each Neo-Fuzzy Neuron comprises a simple fuzzy inference which produces reasoning to singleton weighting consequents:

$$R^{(i)} : if \ x_i \ is \ \tilde{A}_i^{(i)} \quad then \quad f_i(x_i)$$

- Each element of the input vector is being fuzzified using Type-2 Interval Fuzzy set:

$$\mu_{ij}(x_i) = -\exp\left(\frac{x_i - c_{ij}}{2\sigma_{ij}}\right)^2 = \begin{cases} \overline{\mu}_{ij} \ as \ \sigma_{ij} = \overline{\sigma}_{ij} \\ \underline{\mu}_{ij} \ as \ \sigma_{ij} = \underline{\sigma}_{ij} \end{cases}$$

- The fuzzy inference should match the output of the fuzzifier with fuzzy logic rules performing fuzzy implication and approximation reasoning in the following way:

$$\mu_{ij}* = \begin{cases} \overline{\mu}_{ij}* = \prod_{i=1}^{n} \overline{\mu}_{ij} \\ \underline{\mu}_{ij}* = \prod_{i=1}^{n} \underline{\mu}_{ij} \end{cases}$$

- The output of the network is produced by implementing consequence matching, type reduction and linear combination as follows:

$$\hat{y}(k) = \frac{1}{2} \sum_{j=1}^{l} (\overline{\mu}_{ij}* + \underline{\mu}_{ij}*) f_i(x_i) = \frac{1}{2} \sum_{i=1}^{l} (\overline{\mu}_{ij}* + \underline{\mu}_{ij}*) w_{ij}$$

which in fact represents a weighted product composition of the i^{-th} input to j^{-th} synaptic weight.

- To train the proposed modeling structure an unsupervised learning scheme has been used. Therefore, a defined error cost term is being minimized at each sampling period in order to update the weights:

$$E = \varepsilon^2 \big/ 2 \text{ and } \varepsilon(k) = y_d(k) - \hat{y}(k)$$

- As learning approach of the proposed modeling structure a simple *heuristic backpropagation* approach , where the scheduled parameters depend on the *signum* of the gradient and defined learning rate, is adopted:

$$w_{ij}(k+1) = w_{ij}(k) + \Delta w_{ij}(k) = w_{ij}(k) + \eta_{ij}(k)\,\text{s}\,ign\left(\frac{\partial E(k)}{\partial w_{ij}(k)}\right)$$

$$\Delta w_{ij}(k) = -\eta_{ij}(k)sign\left(\frac{\partial E(k)}{\partial w_{ij}(k)}\right) = -\eta_{ij}(k)sign\left(\frac{\partial E(k)}{\partial \hat{y}(k)}\frac{\partial \hat{y}(k)}{\partial w_{ij}(k)}\right) = -\eta_{ij}(k)sign\left(\varepsilon(k)\frac{\partial \hat{y}(k)}{\partial w_{ij}(k)}\right)$$

- The learning rate is local to each synaptic weight and it is adjusted by taking into account the extent of the gradient in the current and the past sample period as:

$$\eta_{ij}(k) \begin{cases} \min\left(a\eta_{ij}(k-1), \eta_{\max}\right) & if \quad \Delta E_{ij}(k)\Delta E_{ij}(k-1) > 0 \\ \max\left(b\eta_{ij}(k-1), \eta_{\min}\right) & if \quad \Delta E_{ij}(k)\Delta E_{ij}(k-1) < 0 \\ \eta_{ij}(k-1) & if \quad \Delta E_{ij}(k)\Delta E_{ij}(k-1) = 0 \end{cases}$$

where the constants are: a=1.2, b=0.5 and $\eta_{\min}$=10^{-3}, $\eta_{\max}$=5.

The main advantage of the proposed approach is that the information about the gradient is neglected, which accelerates significantly the learning process!

- To test the modeling capabilities of the proposed NEO-fuzzy neural network, a numerical experiments in prediction of two common chaotic time series (Mackey-Glass a and Rossler)are investigated.

- The Rossler chaotic time series are described by three coupled first-order differential equations:
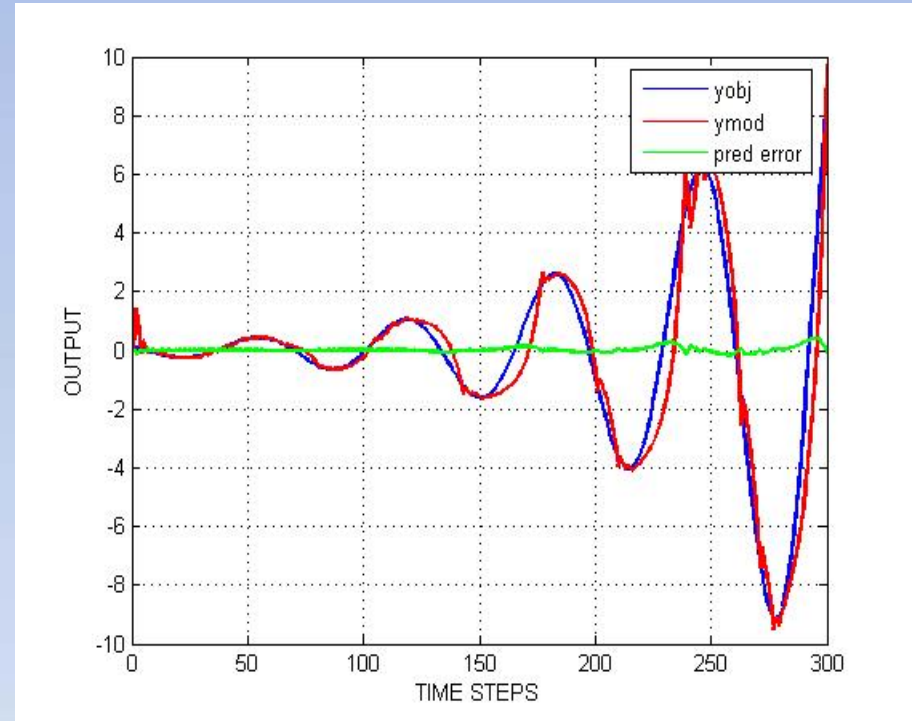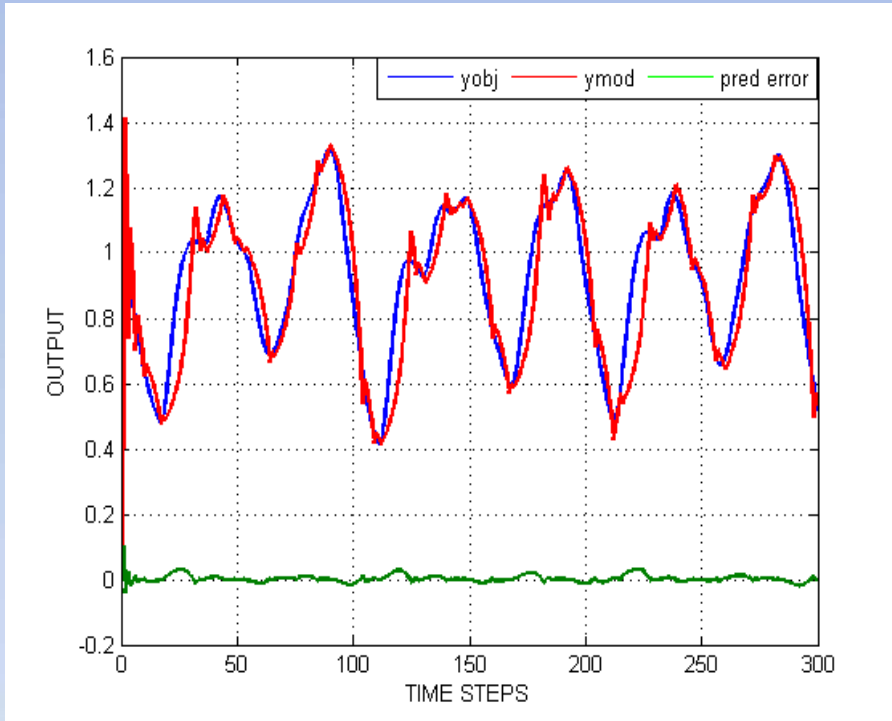
$$\frac{dx}{dt} = -y - z \qquad \frac{dy}{dt} = x + ay \qquad \frac{dz}{dt} = b + z(x - c)$$

  a=0.2; b=0.4; c=5.7 and initial conditions x_0=0.1; y_0=0.1; z_0=0.1
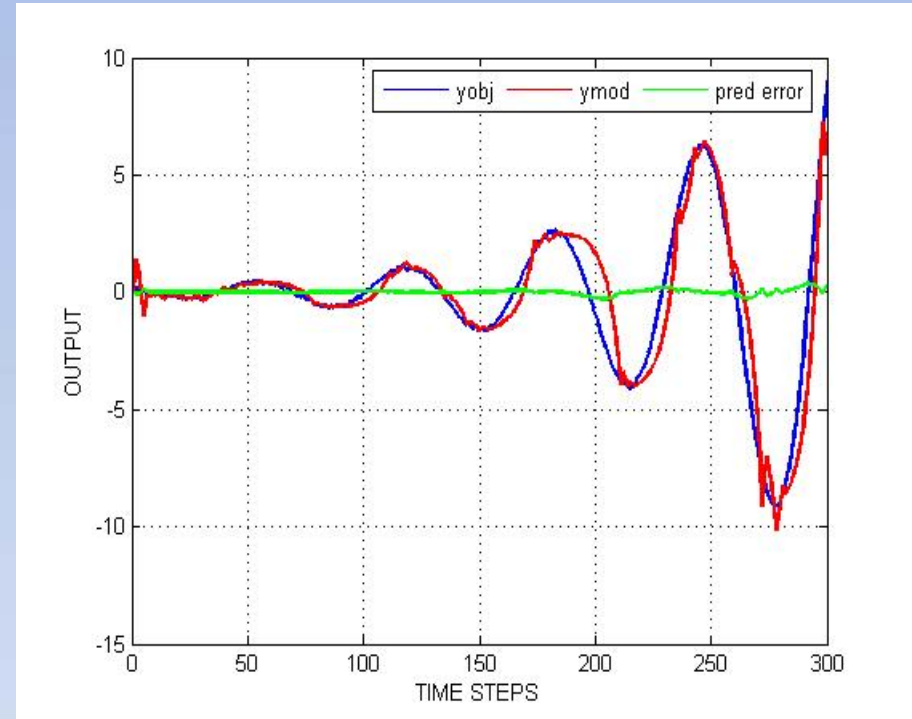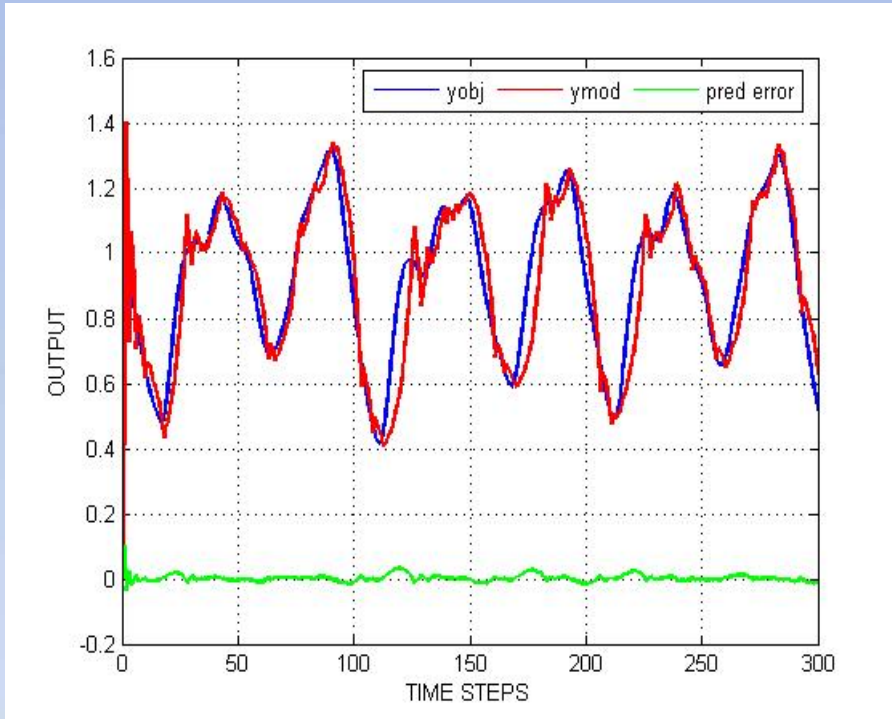
- The Mackey-Glass (MG) chaotic time series is described by the following time-delay differential equation:

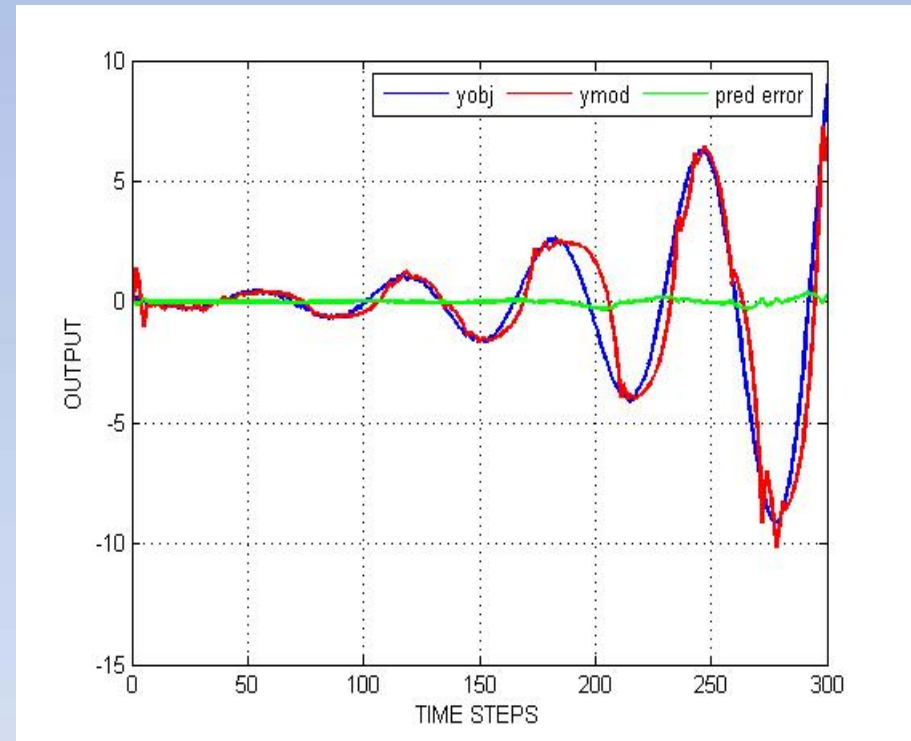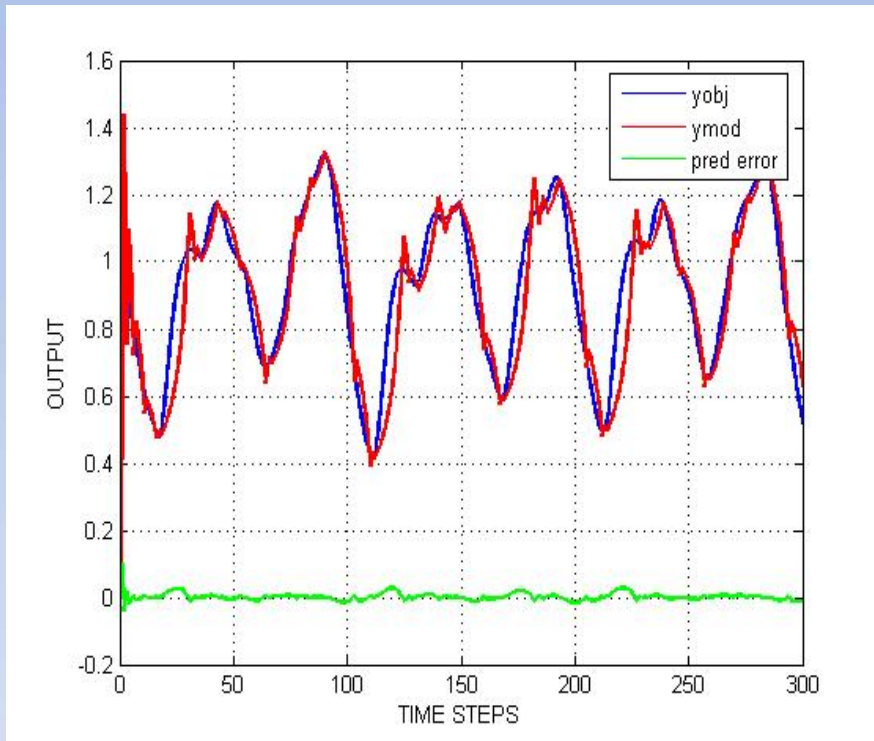$$x(i + 1) = \frac{x(i) + ax(i - s)}{(1 + x^c(i - s)) - bx(i)}$$

  a=0.2; b=0.1; C=10; initial conditions x_0=0.1 and τ= 17s.

Modeling of Mackey-Glass and Rossler chaotic time series and
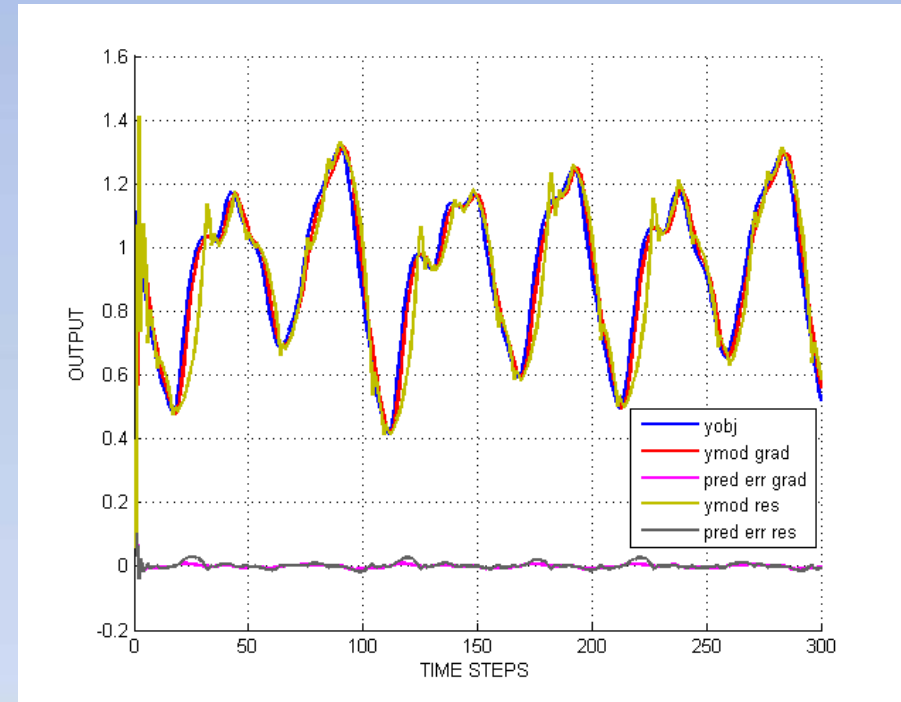the estimated error in the noiseless case.

Modeling of Mackey-Glass and Rossler chaotic time series and
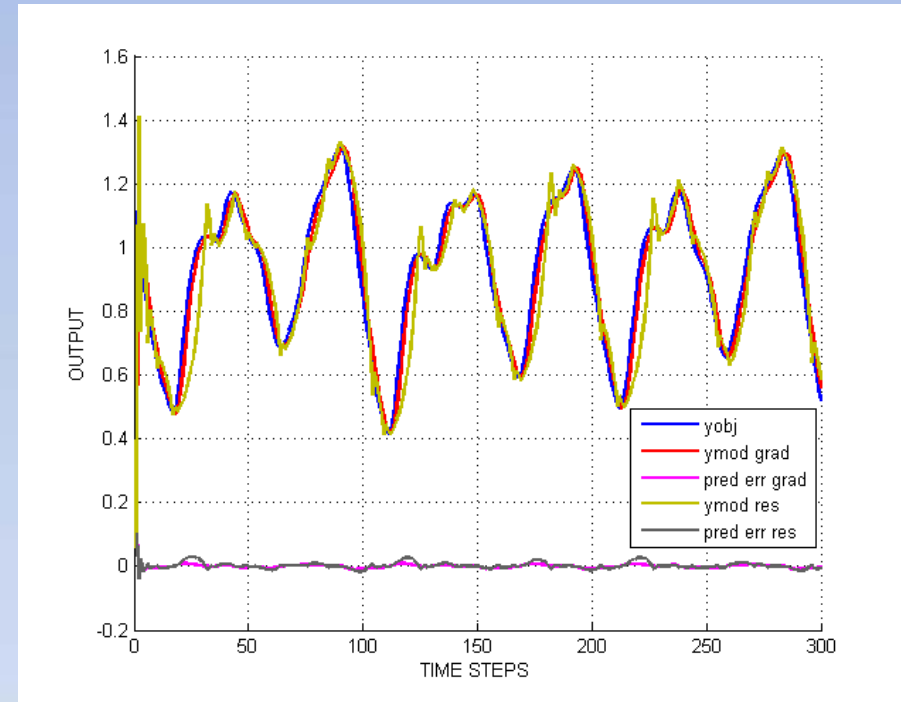the estimated error in the case of 5% additive noise and 5% FOU.

Modeling of Mackey-Glass chaotic and Rossler time series and the estimated error in the case of 5% additive noise and 10% FOU.

ACoMIn: Advanced Computing for Innovation

Mean Squared Errors			
Time step	*Without noise* 10^{-4}	*With noise and 5% FOU,* 10^{-4}	*With noise and 10% FOU,* 10^{-4}
50	4.70	4.66	4.62
100	2.86	2.70	2.64
150	3.37	3.90	2.95
200	8.07	7.47	6.97
250	39.88	22.33	21.82
300	81.71	72.81	70.13

Comparison of the proposed heuristic algorithm to the classical Gradient Descent.

Mean Squared Errors			
Time step	*Without noise 10^{-4}*	*With noise and 5% FOU, 10^{-4}*	*With noise and 10% FOU, 10^{-4}*
50	4.70	4.66	4.62
100	2.86	2.70	2.64
150	3.37	3.90	2.95
200	8.07	7.47	6.97
250	39.88	22.33	21.82
300	81.71	72.81	70.13

Comparison of the proposed heuristic algorithm to the classical Gradient Descent.

AComIn: Advanced Computing for Innovation

http://www.iict.bas.bg

- The achieved results in modeling chaotic time series show a good model performance in the cases without/with 5% additive noise on the input signal, which proves the ability of the network the handle uncommon and uncertain additive noises, with relative unchanged error variance.

- The comparison with the classical Gradient Descent shows a similar model operation and error variance.

- Since, most of the real time signals concerning modeling in purpose of process control have a slower dynamics (frequency and amplitude changes), the proposed approach may be a promising solution with application in modern control systems.

- A future work and extension of the proposed network is its inclusion in Model Predictive Control scheme.