

Neural Networks Cartridges for Data Mining on Time Series

Eduardo Ogasawara¹, Leonardo Murta²,
Geraldo Zimbrão¹, Marta Mattoso¹

¹ Computer Science Department - COPPE/UFRJ, Rio de Janeiro, Brazil

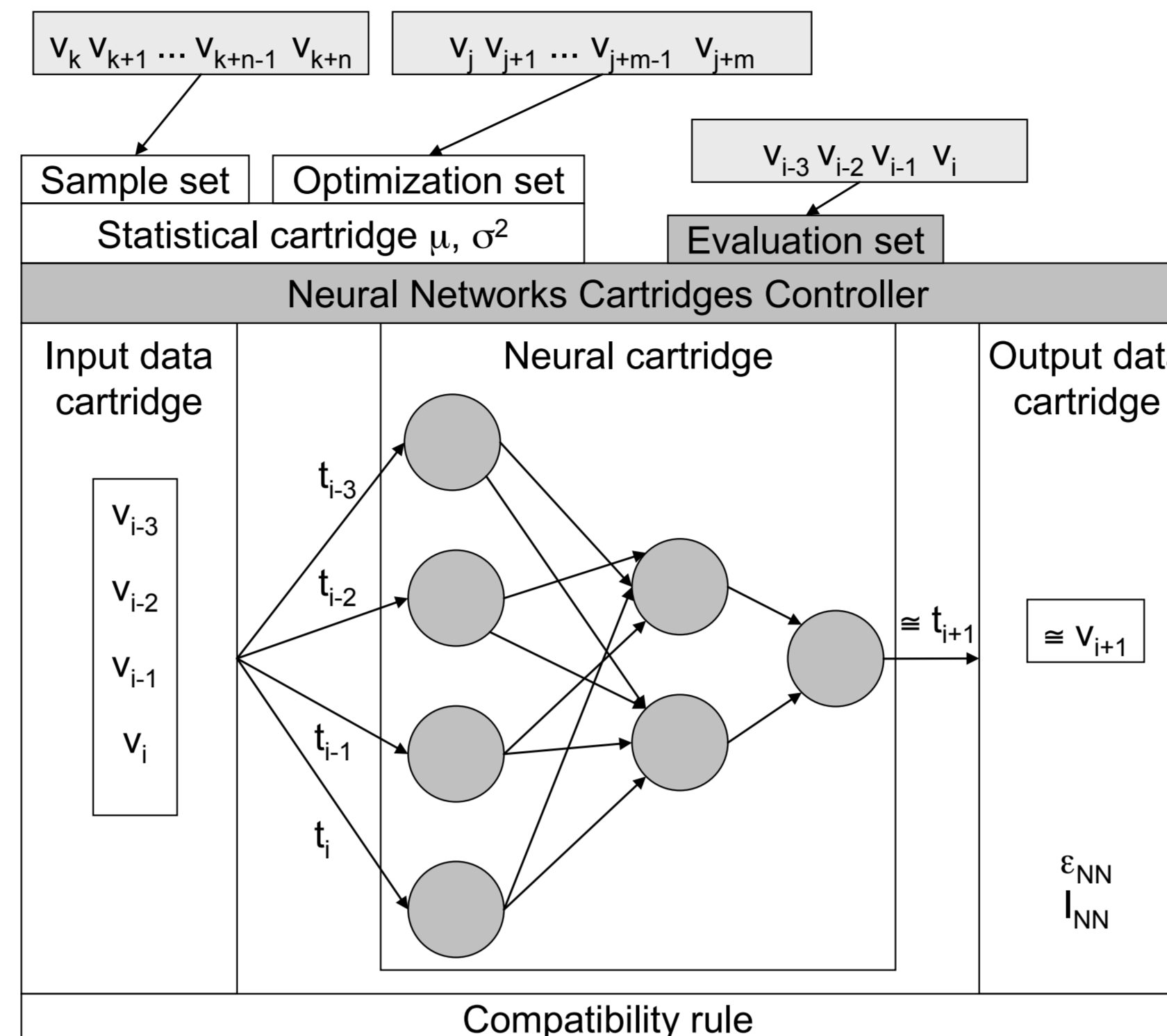
² Computer Science Institute - UFF, Niterói, Brazil



Summary

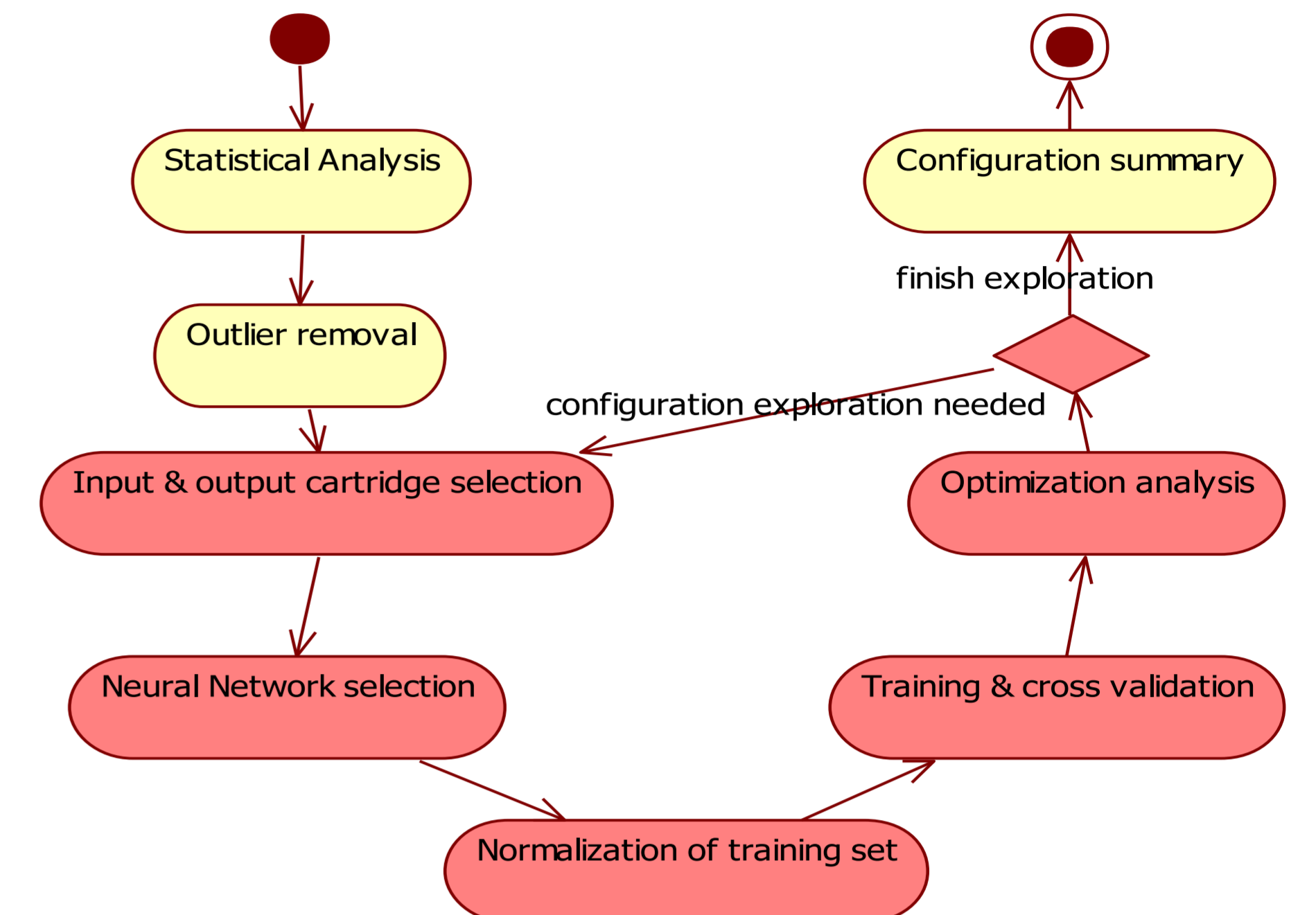
The performance of neural networks for time series forecast is affected by several parameters such as the type of neural network used and its configuration, but also by the quality of data preprocessing. All these parameters need to be explored in order to obtain an optimal neural network adequate for the data being analyzed. However, the manual establishment of different neural networks configurations for selecting the best ones may be error-prone and time-consuming. This paper proposes the development of neural networks cartridges to systematically empower neural network performance by means of data mining activities that explore different neural network structures and different data preprocessing techniques.

Architecture



Each cartridge may be exchanged during the exploratory phase using the training set of the time series and validated using the optimization set.

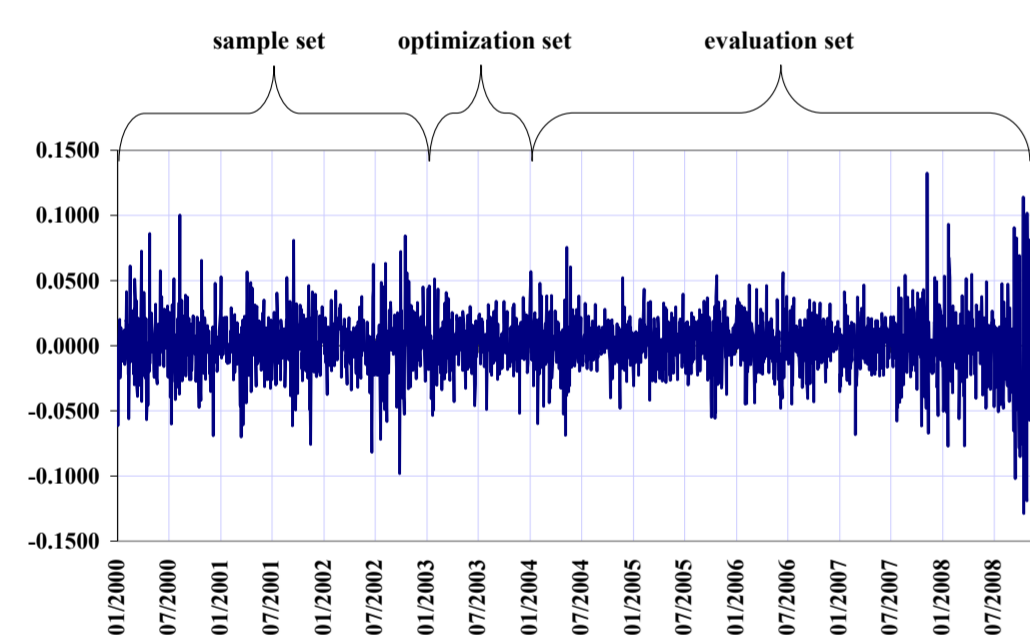
Optimization process



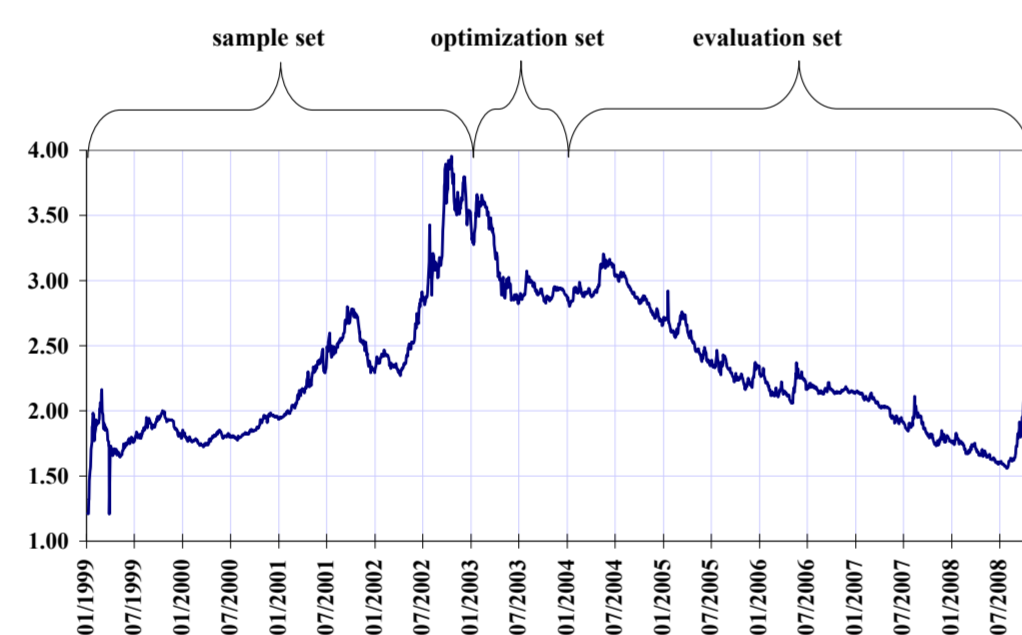
The statistical analysis and outlier removal are applied into the training and optimization set. The optimization process explores different data normalization techniques, different neural networks types, and different neural networks configuration.

Stationary/Non Stationary

A non stationary affects the network training



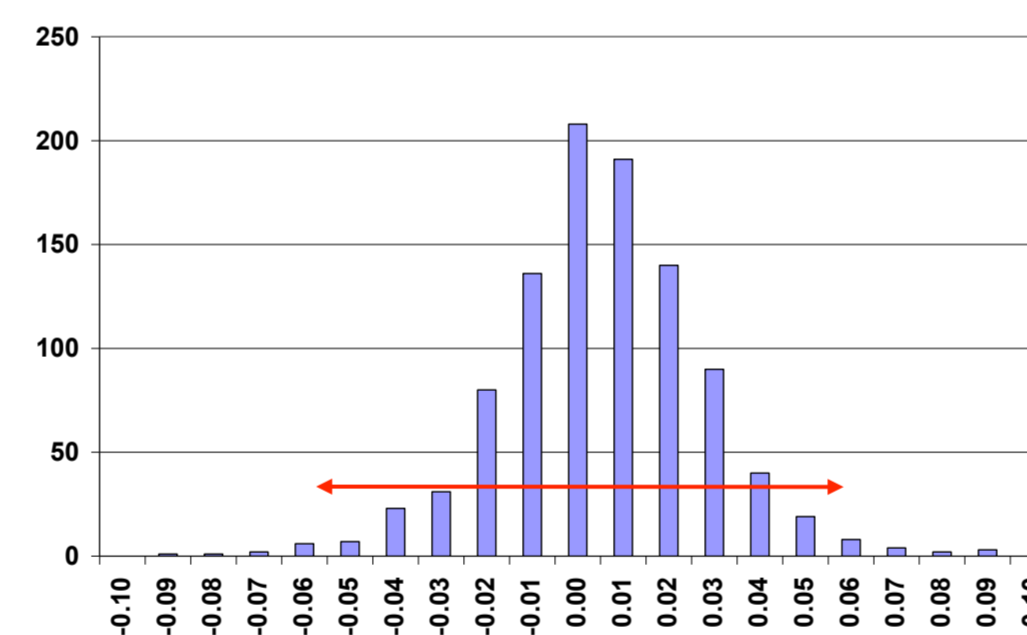
Stationary log return time series for the Brazilian Oil Company Stock (PETR4)



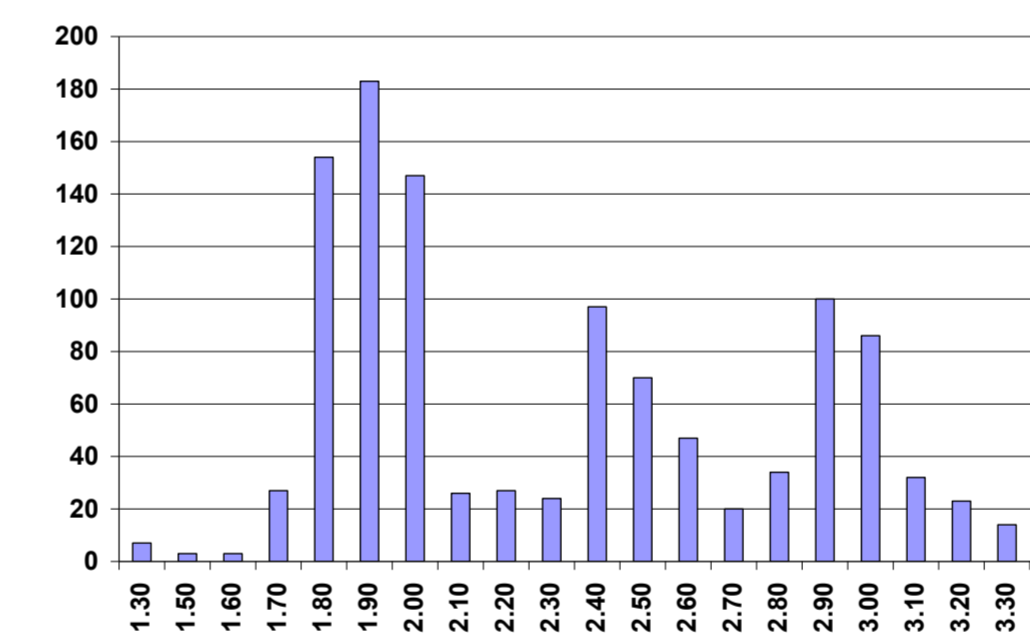
Pseudo-stationary time series for the U.S. Dollar to Brazilian Real Exchange Rate

Outlier removal

Outlier removal using the central limit theorem



Outlier removed for the log return of the Brazilian Oil Company Stock (PETR4)



No outlier removed for the U.S. Dollar to Brazilian Real Exchange Rate

Normalization

$$\text{Min/Max normalization: } t = (H - L) \cdot \frac{v - \min_v}{\max_v - \min_v} + L$$

$$\text{Z-score normalization: } t = \frac{v - \bar{V}}{\sigma V}$$

$$\text{Decimal normalization: } t = \frac{v}{10^j}$$

Error Measurement

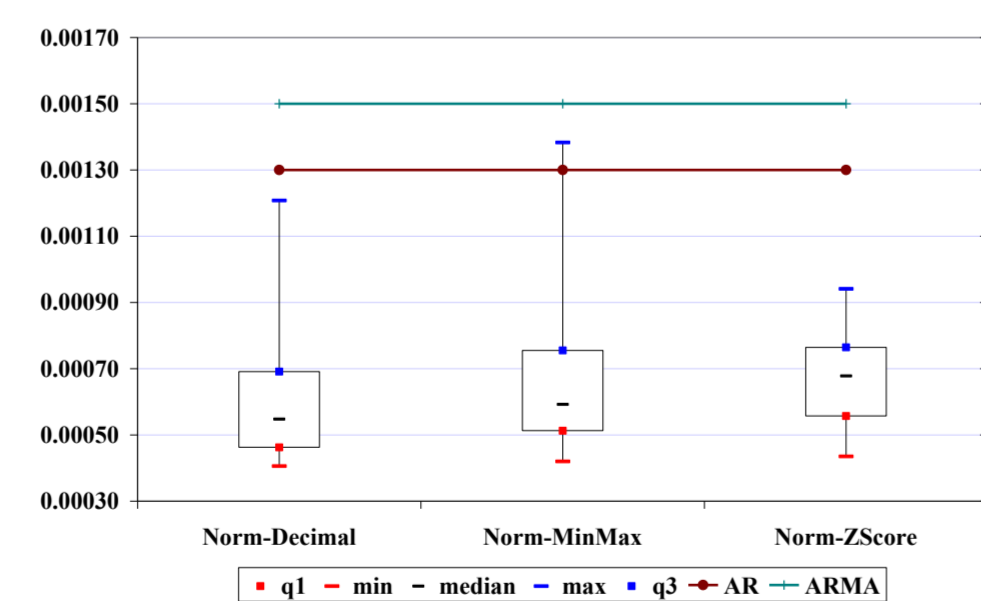
White noise: $\omega_i = f_i - v_i$

$$\text{Accuracy: } \delta = \frac{\sum_{i=1}^N \omega_i^2}{N}$$

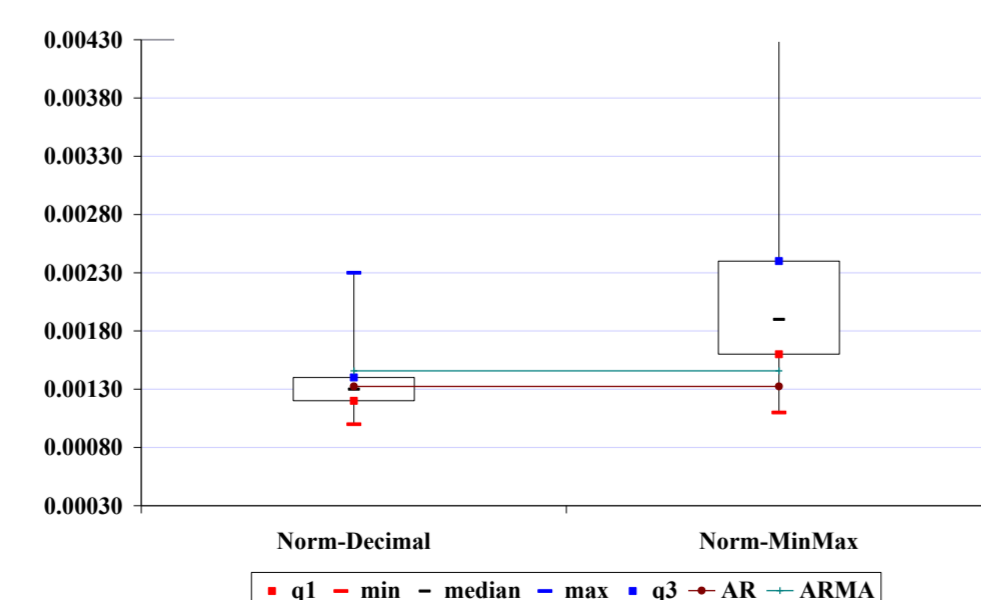
$$\text{Precision: } \rho = t \frac{\sigma(\omega)}{\sqrt{N}}$$

Global error: $\epsilon = \delta + \rho$

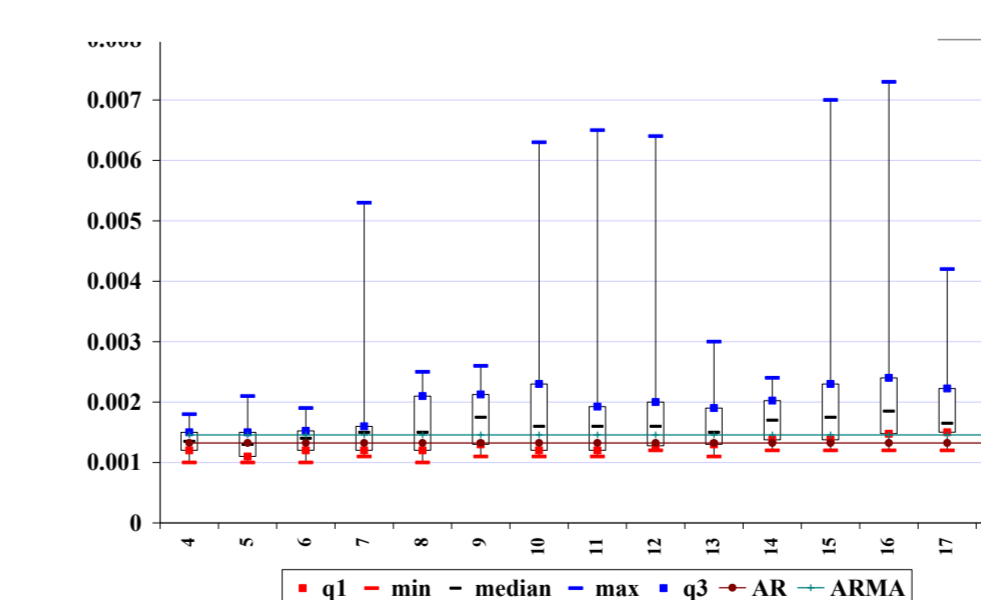
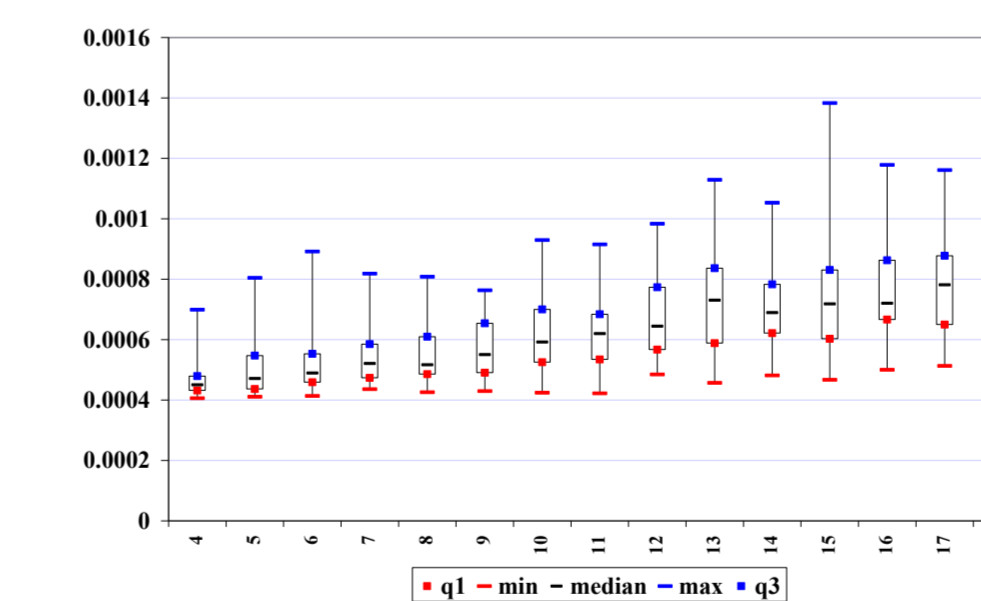
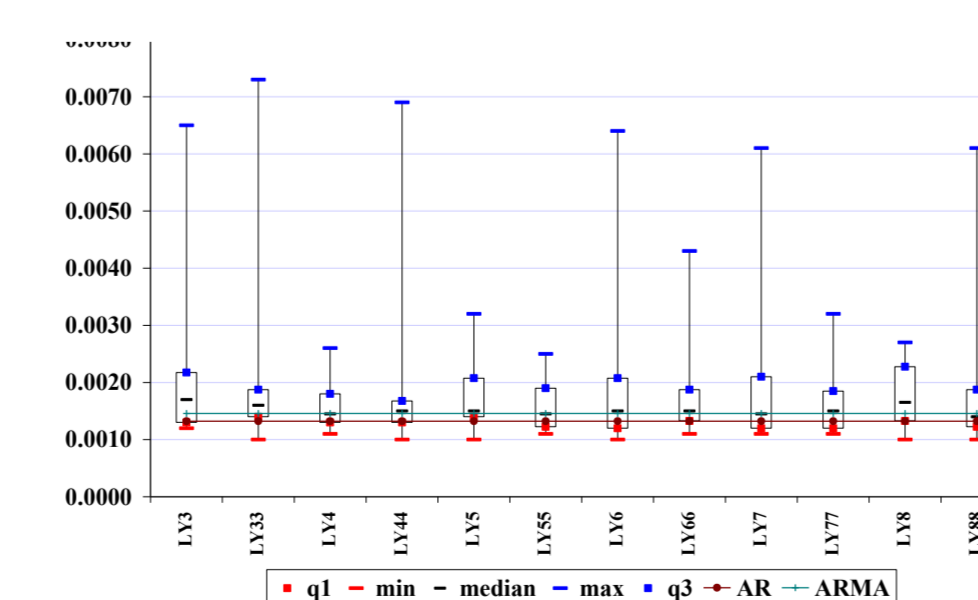
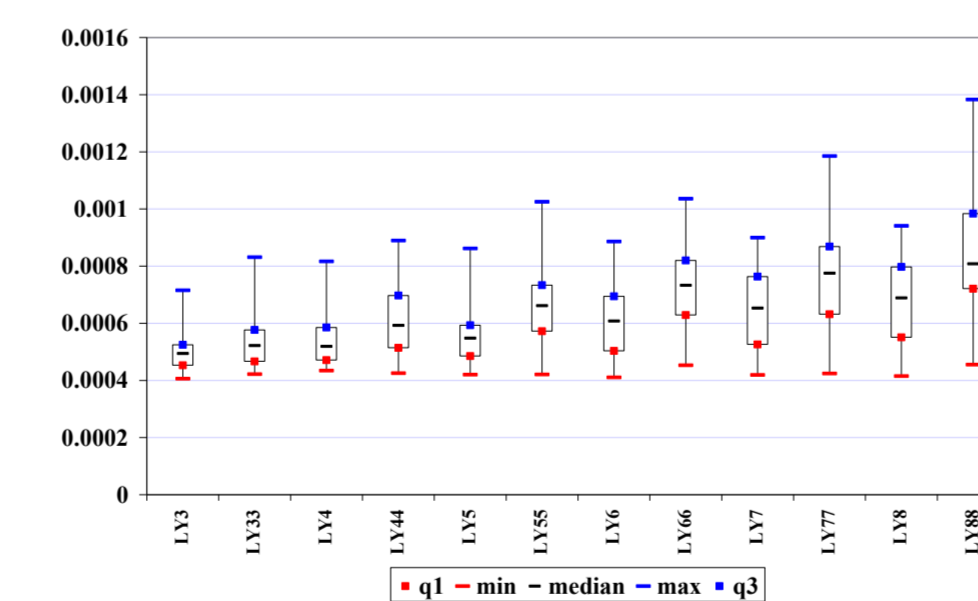
Neural network evaluation using optimization set



Influence of data normalization, number of layers, number of input entries for the Brazilian Oil Company Stock training



Influence of data normalization, number of layers, number of input entries for the U.S. Dollar to Brazilian Real Exchange Rate



Overall performance

Evaluation of the neural network performance

