# TIME SERIES PREDICTION FOR KERNEL-BASED ADAPTIVE FILTERS USING VARIABLE BANDWIDTH, ADAPTIVE LEARNING-RATE, AND DIMENSIONALITY REDUCTION

S. Garcia-Vega E. A. Leon-Gomez G. Castellanos-Domínguez

Universidad Nacional de Colombia, Signal Processing and Recognition Group,
Campus La Nubia, Km 7 vía al Magdalena, 170003, Manizales, Colombia
segarciave@unal.edu.co, ealeong@unal.edu.co, cgcastellanosd@unal.edu.co

## Motivation

- **Time series prediction** techniques have been used in a wide variety of real-world applications, e.g., financial markets, electric utility load, weather state, and human emotions, among others.
- **In practice**, the underlying system models and data generating processes are usually not known, resulting in a challenge to build accurate and unbiased estimation of timeseries data.

## Problem Statement

- The baseline solutions to perform prediction tasks are the statistical methods, mostly employing some improved versions of regressive models. **However**, their imposed analytic models frequently face numerous restrictions when dealing with non-stationarities and nonlinearities of data
- To overcome nonlinearities, data-driven approaches are widely used like Neural Networks (NN), employing one or more layers of non-linear units to predict outputs. **However,** NN algorithms tend to demand long training time and may get stuck in local minima.
- In contrast to NNs, kernel-based adaptive filters have convex optimization and moderate computational complexity. **However** the kernel methods pose three main open issues: *i)* selection of an appropriate kernel bandwidth; *ii)* learning-rate parameter; *iii)* selection of samples to train the model.

## Main Contributions

- The proposed framework sequentially optimize the **bandwidth** and **learning-rate** parameters using stochastic gradient algorithms that maximize the correntropy function.
- A **sparsification** approach based on dimensionality reduction is proposed to remove redundant samples.

## Materials and Methods

- **The goal** is to learn a continuous input-output mapping $f\colon\mathcal{U}\rightarrow\mathbb{R}$ based on a paired sequence of input-output examples $\{\boldsymbol{u}_1, y_1\}, \ldots, \{\boldsymbol{u}_t, y_t\}$.
- The input-output mapping function $f$ can be learned using a **kernel-based adaptive filter**, yielding the following sequential rule through the time domain:

$$f_t = \begin{cases} f_{t-1} + \eta\epsilon_t\kappa_\sigma(\boldsymbol{u}_t, \cdot), & \forall t \neq 0 \\ 0, & t = 0 \end{cases} \tag{1a}$$

$$\epsilon_t = y_t - f_{t-1}(\boldsymbol{u}_t) \tag{1b}$$

- We propose to optimize the adaptive filter parameters using the **correntropy cost function** expressed over time as follows:

$$J_t = \arg\max_{\forall\sigma,\eta}\{\exp\left(-\epsilon_t^2(\sigma_t, \eta_t)/2\lambda^2\right)\} \tag{2}$$

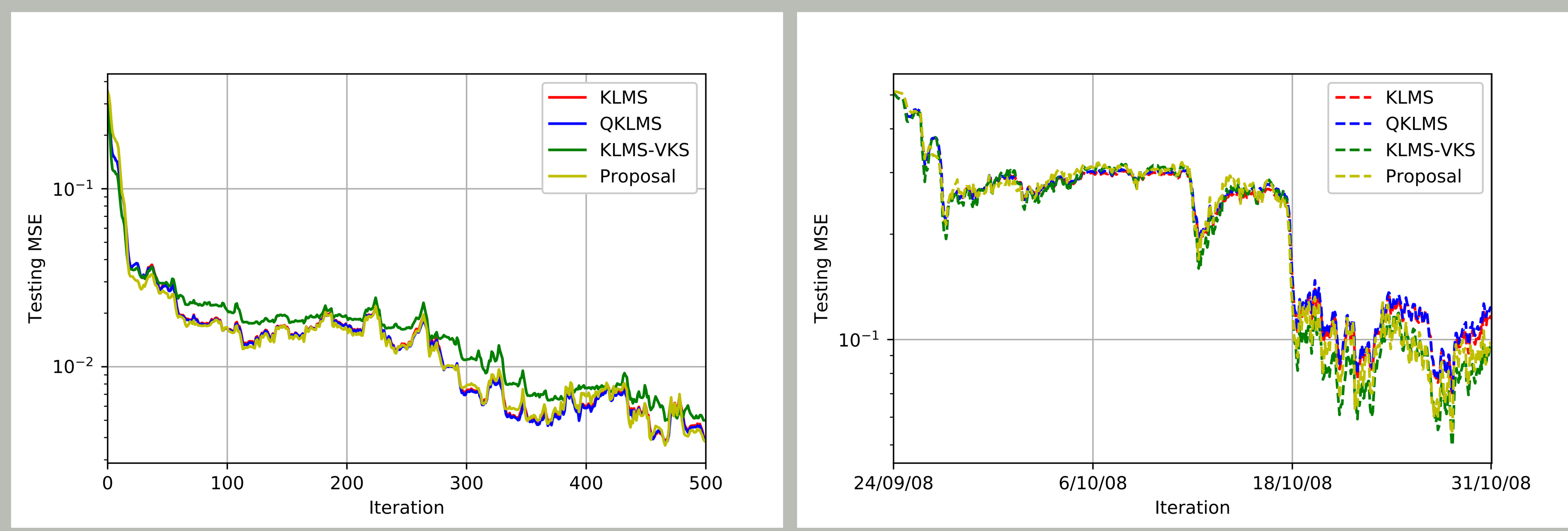
- In addition, the proposed kernel-based **dimensionality reduction** method aims to select only the input data that encodes the **global structures** extracted from training samples.

## Datasets

- **Mackey-Glass**: This time-series is a chaotic system whose states are governed by a set of time-delayed differential equations
- **Wind Speed**: This collection holds hourly wind speed records from the northern region of Colombia.

## Results

The task is to predict the current value using the previous ten consecutive samples.




(a) Mackey-Glass (b) Wind Speed

Figure: Learning curves of each compared method on tested datasets.

| Dataset | Method | Measure | Iteration | | | | |
|---|---|---|---|---|---|---|---|
| | | | *100* | *200* | *300* | *400* | *500* |
| Mackey-Glass | KLMS | MSE | 0.016 | 0.017 | 0.007 | 0.006 | 0.004 |
| | | DS | 100 | 200 | 300 | 400 | 500 |
| | QKLMS | MSE | 0.017 | 0.017 | 0.007 | 0.006 | 0.004 |
| | | DS | 80 | 103 | 126 | 136 | 150 |
| | KLMS-VKS | MSE | 0.021 | 0.019 | 0.011 | 0.008 | 0.005 |
| | | DS | 100 | 200 | 300 | 400 | 500 |
| | Proposal | MSE | 0.016 | 0.007 | 0.007 | 0.006 | 0.004 |
| | | DS | **57** | **71** | **86** | **100** | **104** |
| | | | Iteration | | | | |
| | | | *28/09/08* | *06/10/08* | *14/10/08* | *23/10/08* | *31/10/08* |
| Wind Speed | KLMS | MSE | 0.253 | 0.299 | 0.249 | 0.084 | 0.115 |
| | | DS | 100 | 300 | 500 | 700 | 900 |
| | QKLMS | MSE | 0.252 | 0.302 | 0.255 | 0.087 | 0.122 |
| | | DS | 81 | 193 | 280 | 357 | 371 |
| | KLMS-VKS | MSE | 0.241 | 0.311 | 0.253 | 0.066 | 0.094 |
| | | DS | 100 | 300 | 500 | 700 | 900 |
| | Proposal | MSE | 0.262 | 0.311 | 0.272 | 0.074 | 0.095 |
| | | DS | **62** | **88** | **96** | **108** | **121** |

Table: Performed results on tested datasets at different iterations. The best overall method of each column are marked with bold notation. *MSE*-mean square error. *DS*-Dictionary Size.

## Conclusions

- In this study, a **framework for kernel-based adaptive filters is introduced** that addresses three main challenges of their online implementation: selection of appropriate bandwidth, learning-rate, and training samples.
- Validation on both datasets, synthetic and real-world, proves that **the proposed framework converges to relatively low values of mean-square-error**, avoiding overfitting while providing stable solutions in real-world applications.
- We are in the process of expanding our research to **other information theoretic measures**. In the future, we plan to extend the results to the case where a more elaborate hyper-parameter tuning procedure is introduced into the compared kernel-based adaptive filters.