An extensive review and comparison of R packages on the long-range dependence estimators

Presentation · August 2017

CITATIONS
0

READS
161

3 authors, including:

Hristos Tyralis
National Technical University of Athens
73 PUBLICATIONS **108** CITATIONS

SEE PROFILE

Demetris Koutsoyiannis
National Technical University of Athens
710 PUBLICATIONS **6,652** CITATIONS

SEE PROFILE

An extensive review and comparison of R packages on the long-range dependence estimators

Hristos Tyralis, Panayiotis Dimitriadis, and Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering
School of Civil Engineering
National Technical University of Athens
(montchrister@gmail.com)

Session HS06: Hydroinformatics

Presentation available online: **itia.ntua.gr/1721**

1. Abstract

The long-range dependence (LRD) is a well-established property of climatic variables such as temperature and precipitation. A long list of estimators of the LRD parameters exist while a few comparison studies of their properties have been published. The emergence of R as one of the favourite programming languages among the hydrological community and its increasing number of packages enable the fast implementation of statistical methods in hydrological studies. Many R packages include functions for the estimation of the parameter, which characterizes the LRD, e.g. the Hurst parameter of the Hurst-Kolmogorov behaviour or the *d* parameter of the ARFIMA model. Here we present an extensive review of all R packages containing functions used to estimate the LRD parameter. Furthermore, we examine the properties of the implemented estimators and we perform an extended simulation experiment to compare them.

2. Introduction

- The Hurst-Kolmogorov process (HKp, also known as fractional Gaussian noise, fGn) and the Autoregressive Fractional Integrated Moving Average models (ARFIMA) are two processes suitable for modelling the long-range dependence.
- Hydrological processes are usually modelled by long-range dependent processes.
- The magnitude of the long-range dependence is characterized by the *H* (Hurst) parameter of the HKp and the *d* parameter of the ARFIMA model.
- Numerous methods for the simulation of the HKp and the ARFIMA model and numerous estimators of the *H* and *d* parameters exist.
- Comparison of the estimators of *H* and *d* has been performed in many studies. A literature review is presented in Witt and Malamud (2013), while two recent studies are Tyralis and Koutsoyiannis (2011) and Rea et al. (2013).
- Many simulators and estimators have been implemented in the R programming language.
- The R programming language has become particularly popular in the hydrological science.
- Here we present R functions which simulate the HKp and the ARFIMA.
- Furthermore we present R functions which estimate the *H* and *d* parameters.
- Lastly we compare a set of the functions in the estimation of *H* and *d*, using different estimators.
- Reference of each R package which includes the respective R functions is given in the References slide.

3. HKp and ARFIMA simulators in R

- Five functions for simulating the HKp.
- One function uses three algorithms.
- Five functions for simulating the ARFIMA.
- One functions uses two algorithms.
- The arfima, fArma and longmemo packages include functions for the simulation of both the HKp and the ARFIMA.
- In the present and the following slides we present in bold the functions which will be used in the study.

HKp simulators	
function	R package
arfima.sim	arfima
fgnSim (3)	**fArma**
SimulateFGN	FGN
lmSimulate	fractal
simFGN0	**longmemo**

ARFIMA simulators	
function	R package
arfima.sim	**arfima**
farimaSim (2)	**fArma**
fracdiff.sim	fracdiff
simARMA0	longmemo
arfimapath	rugarch

4. Hurst parameter estimators in R

- 27 functions in 10 packages.
- A wide list of estimators including methods based on wavelets, maximum likelihood estimators, Whittle estimators, least squares based on variance and least squares based on standard deviation, DFA, R/S, ….

function	R package	function	R package
arfima	arfima	**DFA**	**fractal**
absvalFit	fArma	FDWhittle	fractal
aggvarFit	fArma	**hurstACVF**	**fractal**
boxperFit	**fArma**	hurstBlock	fractal
diffvarFit	fArma	hurstSpec	fractal
higuchiFit	fArma	**mleHK**	**HKprocess**
pengFit	fArma	lssd	HKprocess
perFit	fArma	**lsv**	**HKprocess**
rsFit	**fArma**	liftHurst	liftLRD
waveletFit	**fArma**	WhittleEst	longmemo
earfima	FGN	HurstIndex	PerformanceAnalytics
FitFGN	FGN	hurstexp	pracma
HurstK	**FGN**	hurst.est	Rwave
warfima	FGN		

5. ARFIMA parameter *d* estimators in R

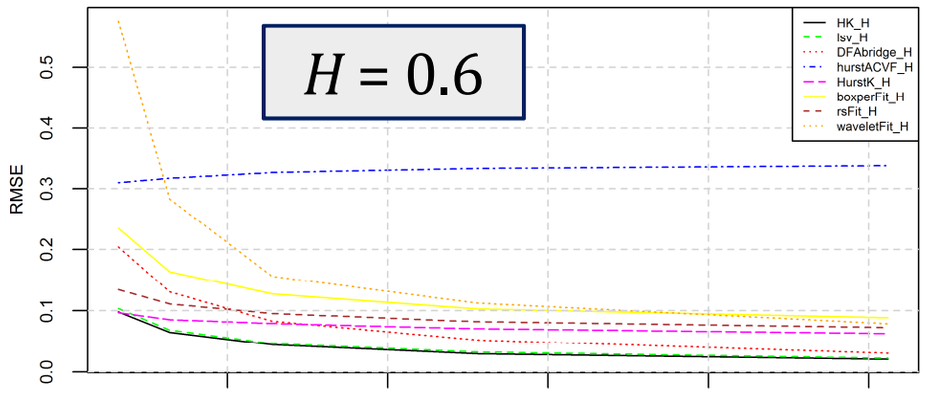
- 11 functions in 8 packages.
- A wide list of estimators including maximum likelihood estimators, Whittle estimators, methods based on periodogram, methods based on wavelets.
- Some functions wrap other R functions for estimating *d*, but also for performing additional tasks.
- The arfima, earfima, warfima, lifthurst, Whittleest functions are also used to estimate the Hurst parameter.

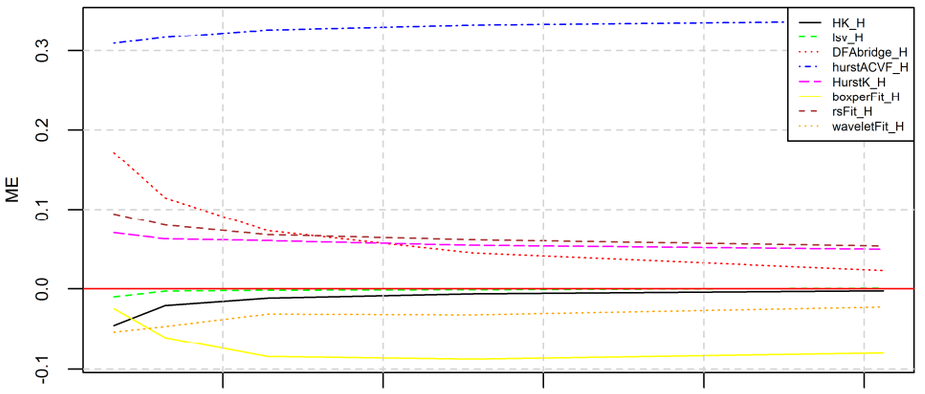
function	R package
arfima	**arfima**
armaFit	fArma
earfima	FGN
warfima	**FGN**
arfima	forecast
fdGPH	**fracdiff**
fdSperio	**fracdiff**
fracdiff	fracdiff
liftHurst	liftLRD
WhittleEst	longmemo
arfimafit	rugarch

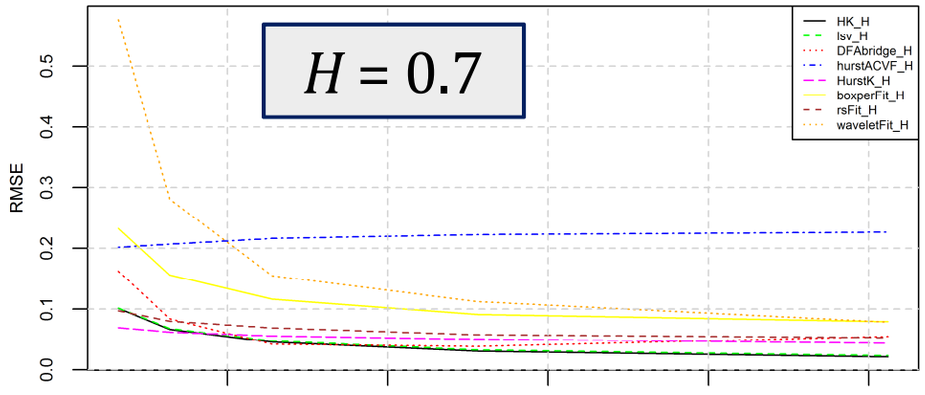
6. Simulation experiment

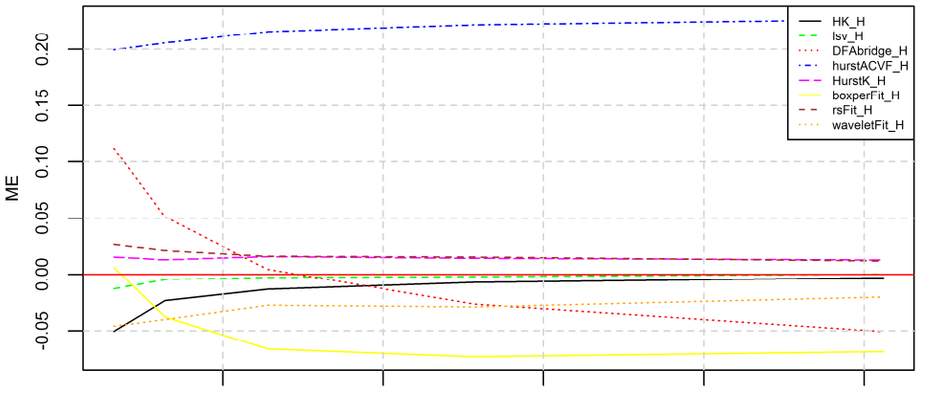
- Two simulators of the Hurst-Kolmogorov process (fgnSim, simFGN0) and another two ARFIMA simulators (arfima.sim, farimaSim) were applied.
- 8 functions for estimating H and 4 functions for estimating d were applied.
- The fuctions used for estimating H were boxperFit, rsFit, waveletFit, HurstK, DFA, hurstACVF, mleHK, lsv.
- The functions used for estimating d were arfima, warfima, fdGPH, fdSperio.
- Simulation lengths were equal to 64, 128, 256, 512, 1024.
- Three values of H (= 0.6, 0.7, 0.8) and three values of d (= 0.1, 0.2, 0.3) were used in the simulation experiment.
- 1000 simulated time series were produced for each simulation length.
- The Mean Error (ME) and the Root Mean Squared Error (RMSE) were calculated.
- The mean error is defined by ME = $(1/1000)\ \Sigma(H_{i,\text{est}} - H_{\text{sim}})$, H_{sim} = 0.6, 0.7, 0.8.
- The Root Mean Squared Error is defined by RMSE = $((1/1000)\ \Sigma(H_{i,\text{est}} - H_{\text{sim}})^2)^{1/2}$, d_{sim} = 0.1, 0.2, 0.3.

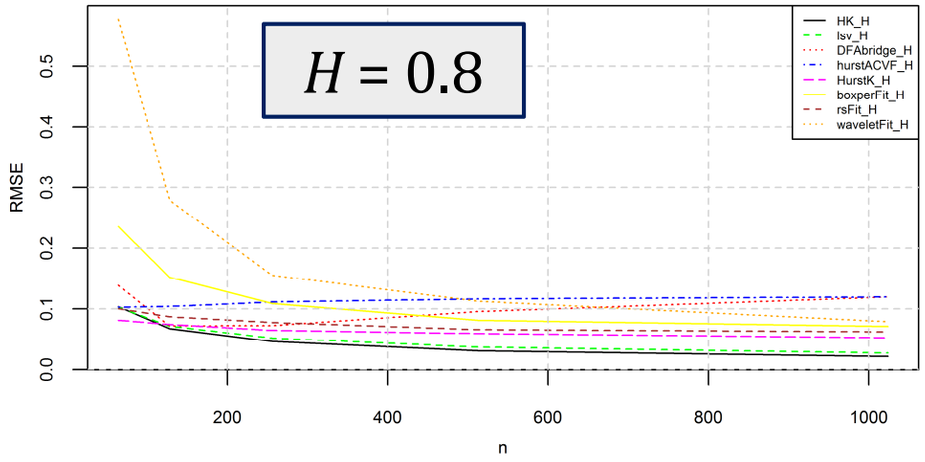
7. First Hurst-Kolmogorov simulator

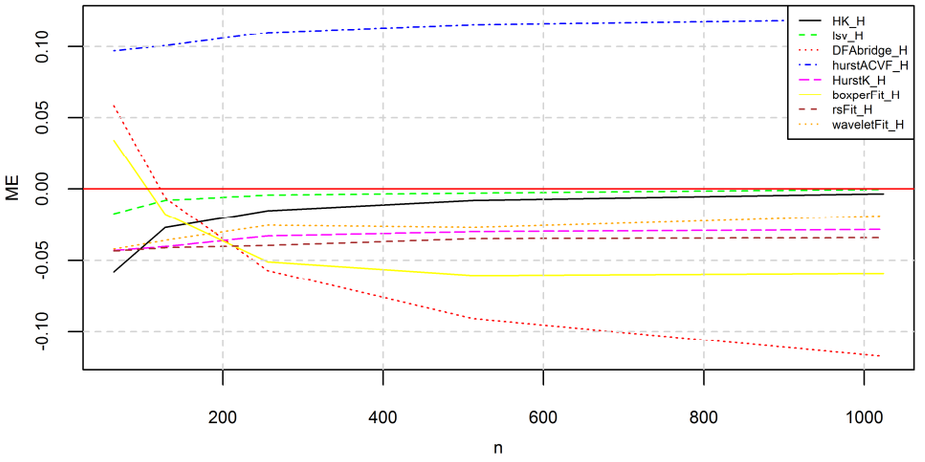




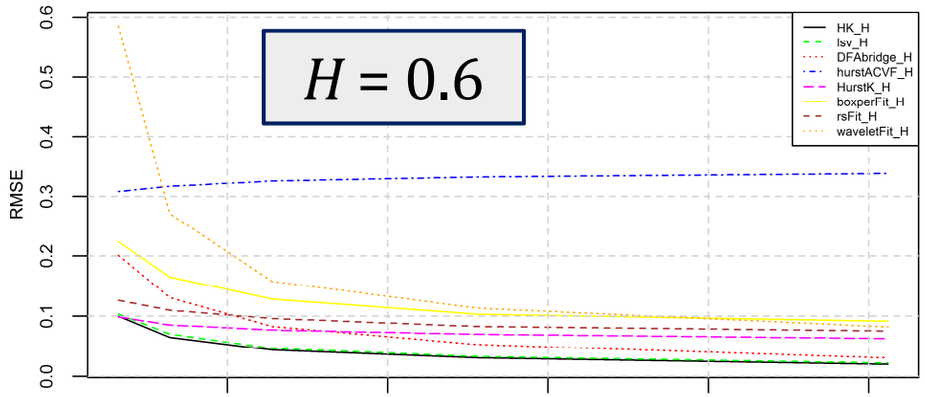


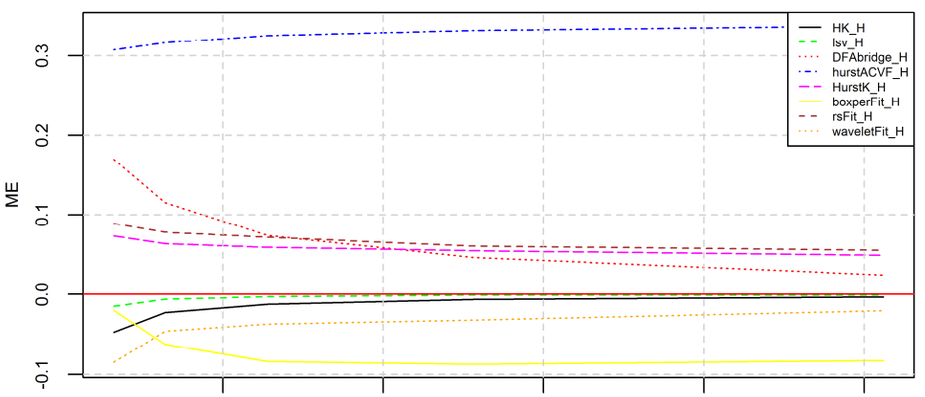


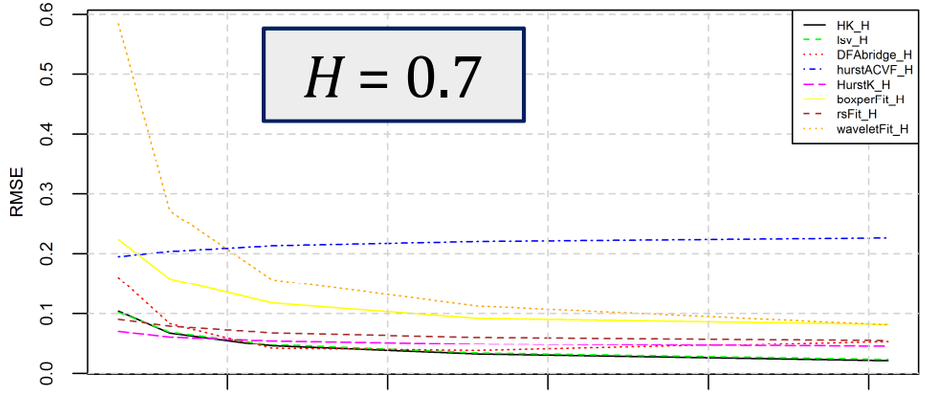


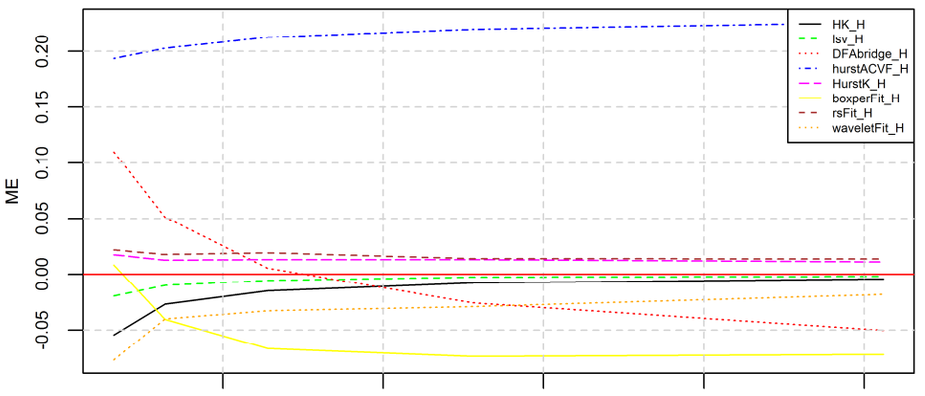


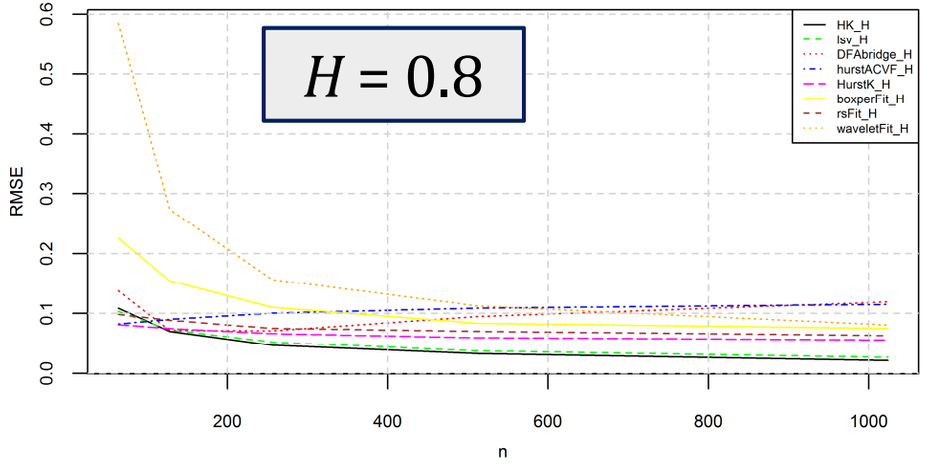
8. Second Hurst-Kolmogorov simulator

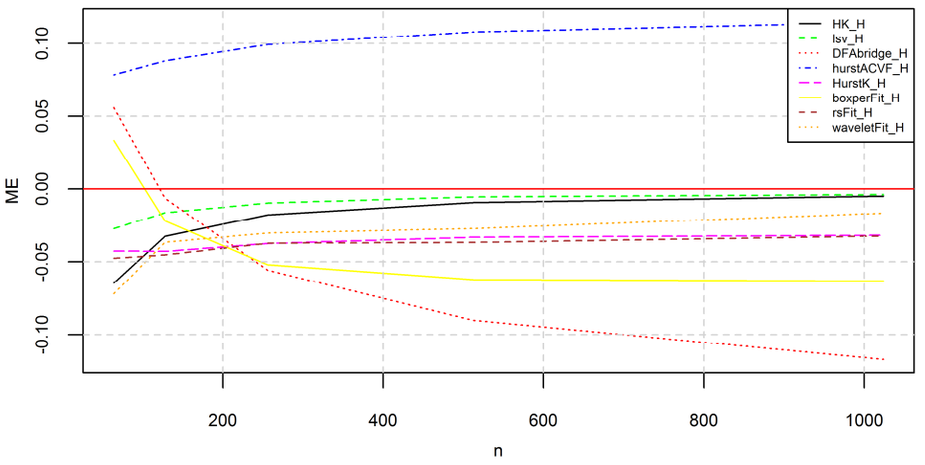




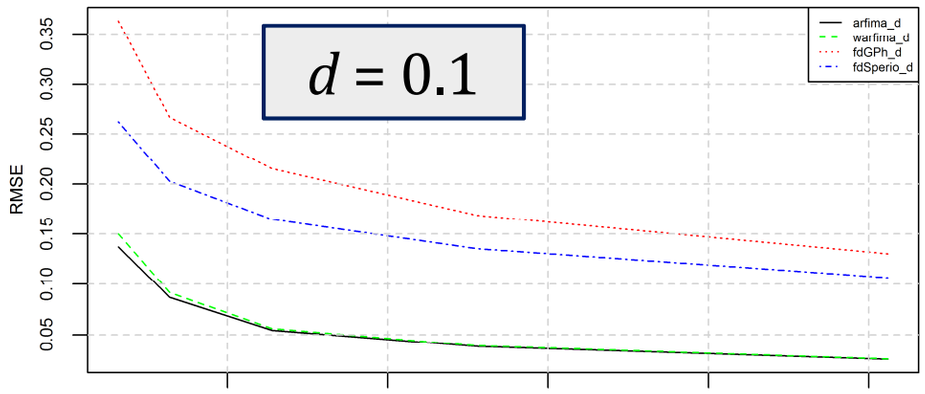


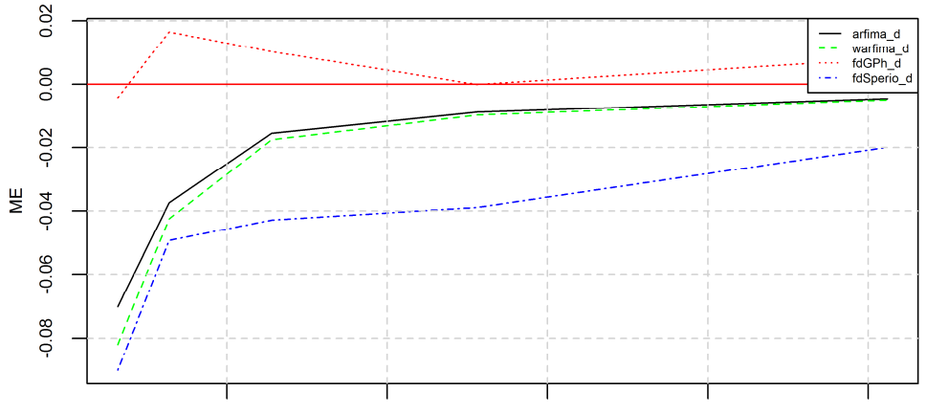


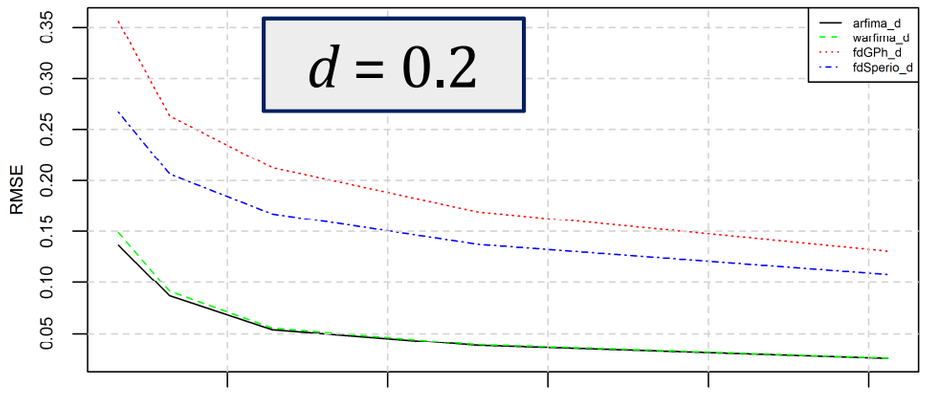


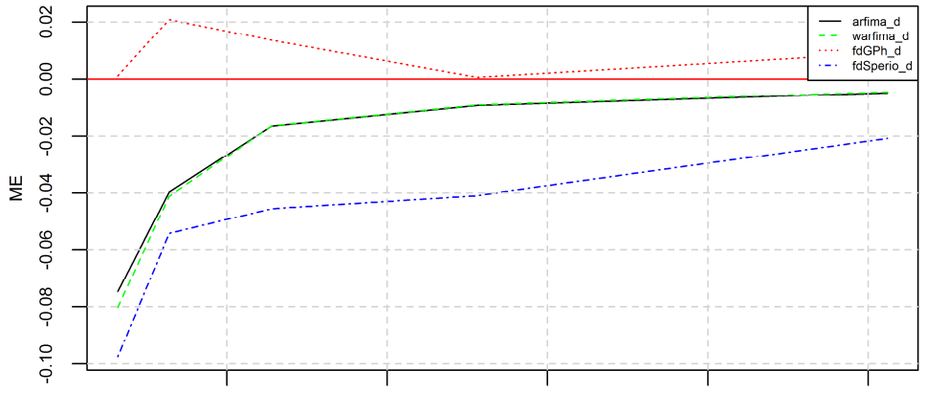


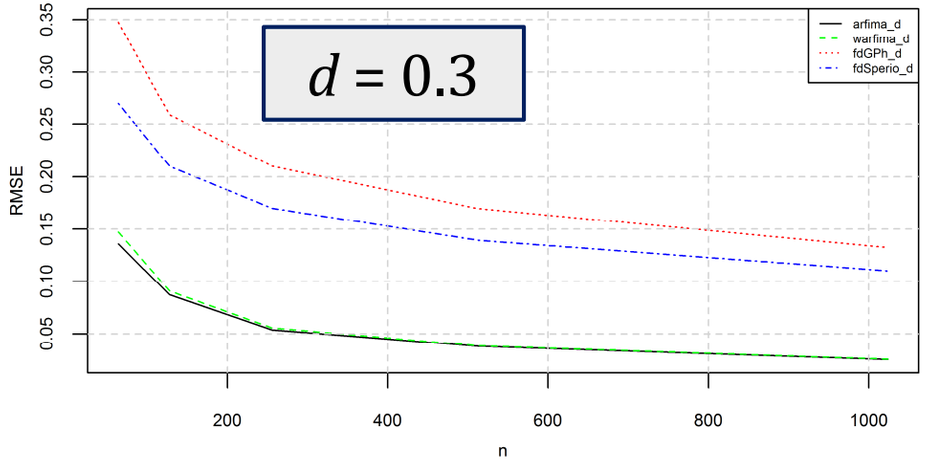
9. First ARFIMA simulator

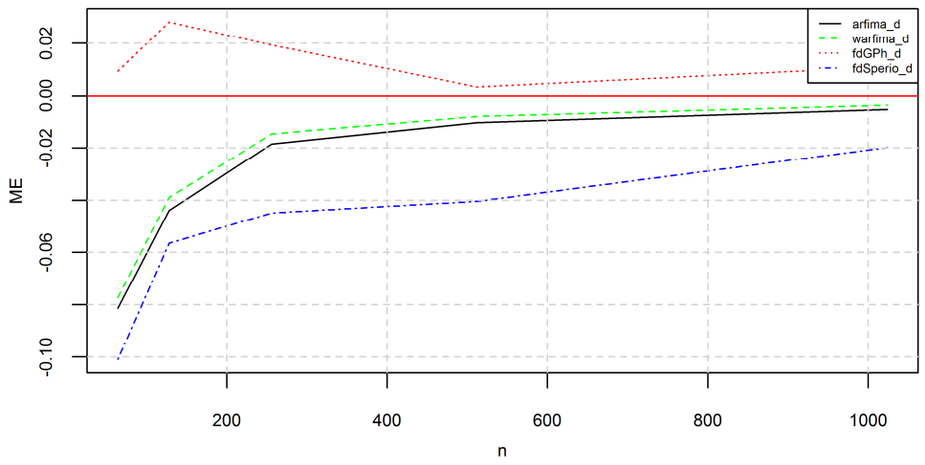




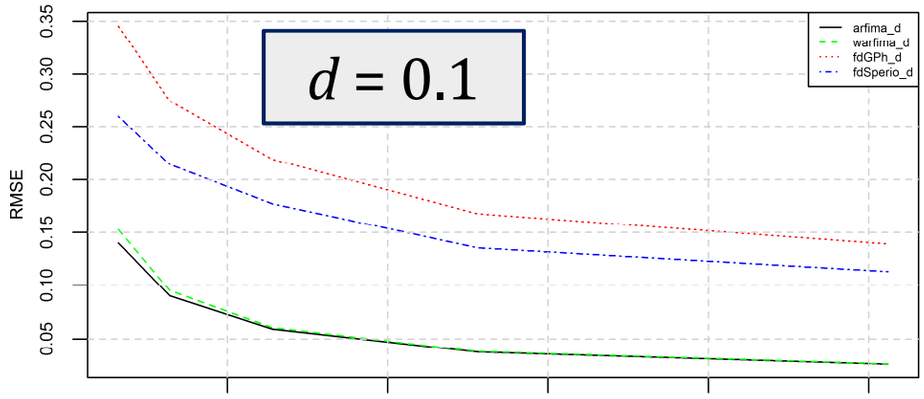


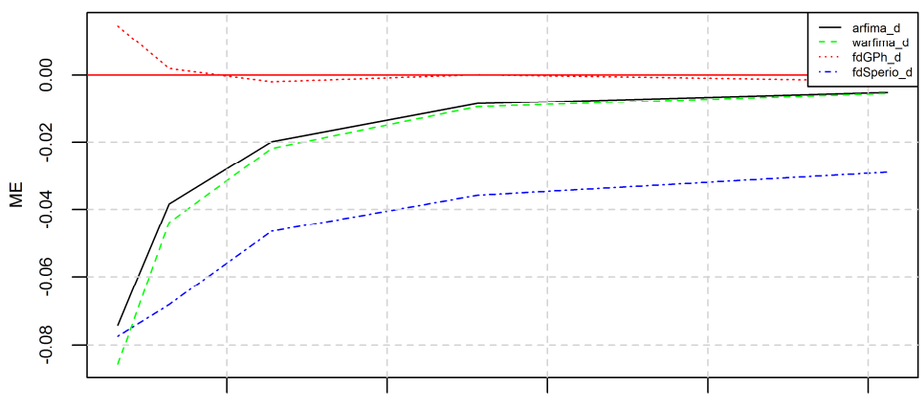


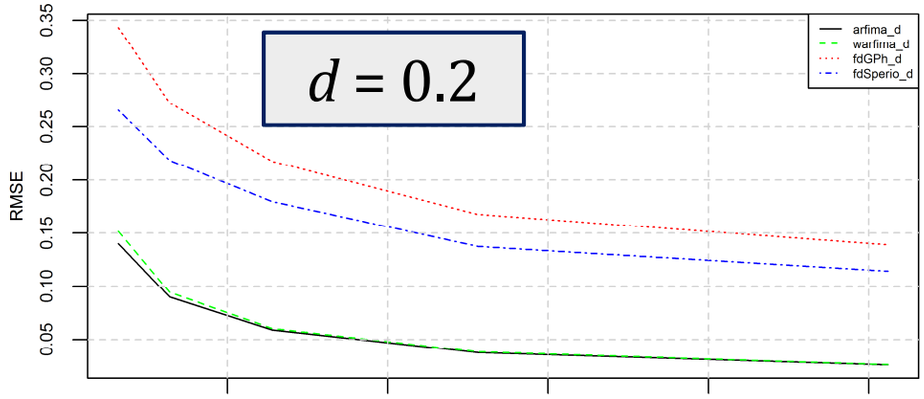


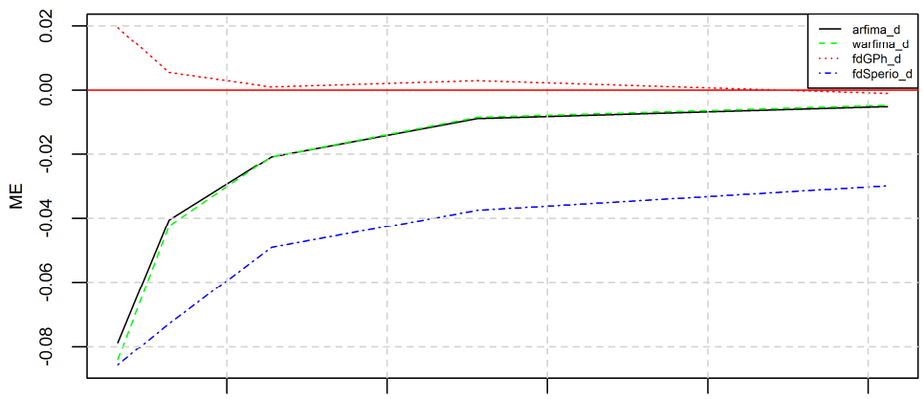


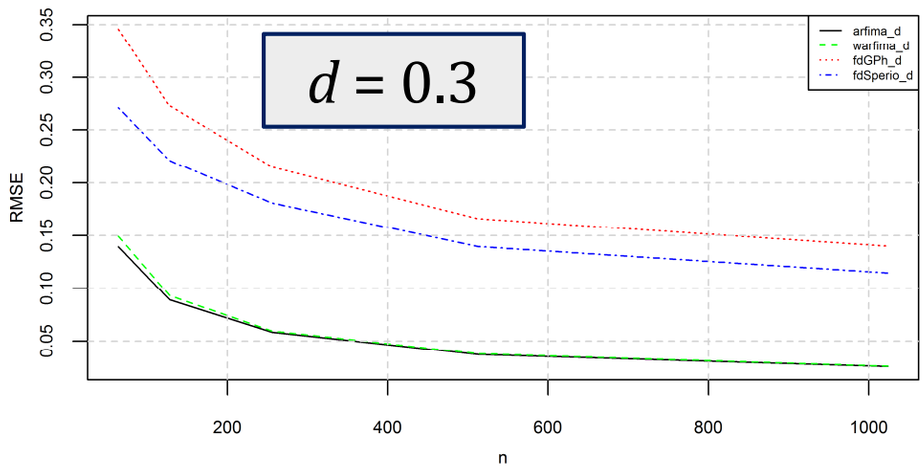
10. Second ARFIMA simulator

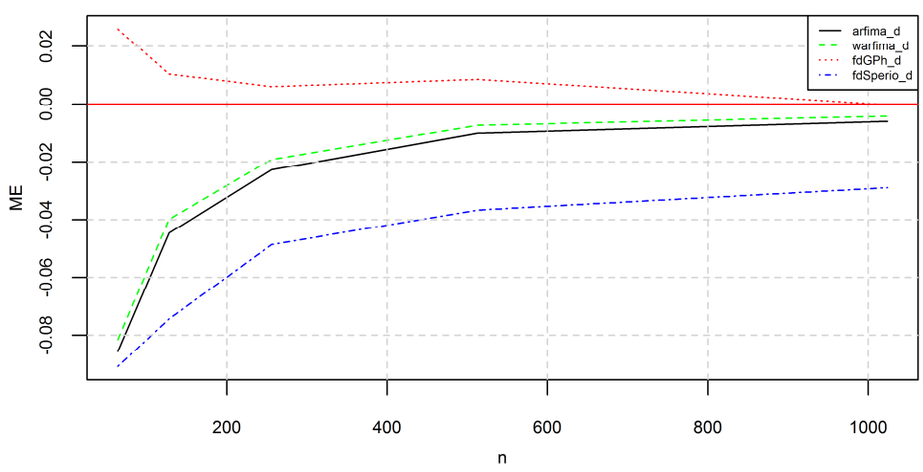












11. Conclusions

- When estimating *H*, the function mleHK had the best RMSE and the lsv the best ME, regardless the estimator used.
- Regarding the other functions estimating *H*, the results were mixed depending on the series length and the value of the parameter used for the simulation.
- The HurstK and the rsFit functions performed well in all simulation experiments, while the performance of the DFA depended on the value of *H* used for the simulation.
- Most estimators were negatively biased, while none of them was unbiased.
- When estimating *d*, the function arfima had the best RMSE followed by the warfima. On the other hand the warfima had the best ME when $d \leq 0.2$, while the earfima had the best RMSE when $d > 0.2$. The results were similar for both simulators.
- The fdSperio had lower RMSE compared to the fdGPH. However the fdGPH had better ME. The results were similar for both simulators.
- The results of the present study refer to the performance of the functions which implement the estimators. Some of the functions use some tuning parameters, which were set constant, regardless of the simulation experiment.
- Regarding the implementation of the estimators, most of them are implemented in more than one R functions.

References

[1] Beran, J., Whitcher, B. and Maechler, M., 2011. **longmemo**: Statistics for Long-Memory Processes (Jan Beran) – Data and Functions. R package version 1.0-0.

[2] Borchers, H. W., 2017. **pracma**: Practical Numerical Math Functions. R package version 2.0.4.

[3] Carmona, R., Torresani, B., Whitcher, B. Lees, J. M., 2017. **Rwave**: Time-Frequency Analysis of 1-D Signals. R package version 2.4-5.

[4] Constantine, W. and Percival, D., 2016. **fractal**: Fractal Time Series Modeling and Analysis. R package version 2.0-1.

[5] Fraley, C., Leisch, F., Maechler, M., Reisen, V. and Lemonte, A., 2012. **fracdiff**: Fractionally differenced ARIMA aka ARFIMA(p,d,q) models. R package version 1.4-2.

[6] Ghalanos, A., 2015. **rugarch**: Univariate GARCH Models. R package version 1.3-6.

[7] Hyndman, R., O'Hara-Wild, M., Bergmeir, C., Razbash, S. and Wang E., 2017. **forecast**: Forecasting Functions for Time Series and Linear Models. R package version 8.0.

[8] Knight, M., Nason, G. and Nunes, M., 2016. **liftLRD**: Wavelet Lifting Estimators of the Hurst Exponent for Regularly and Irregularly Sampled Time Series. R package version 1.0-5.

[9] McLeod, A. I. and Veenstra, J. Q., 2014. **FGN**: Fractional Gaussian Noise and power law decay time series model fitting. R package version 2.0-12.

[10] Peterson, B. G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., Lestel, M., Balkissoon, K. and Wuertz, D., 2014. **PerformanceAnalytics**: Econometric tools for performance and risk analysis. R package version 1.4.3541.

[11] Rea, W., Oxley, L., Reale, M. and Brown, J., 2013. Not all estimators are born equal: The empirical properties of some estimators of long memory. *Mathematics and Computers in Simulation*, 93, 29–42. doi: 10.1016/j.matcom.2012.08.005.

[12] Tyralis, H. and Koutsoyiannis, D., 2011. Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. *Stochastic Environmental Research and Risk Assessment*, 25 (1), 21–33. doi: 10.1007/s00477-010-0408-x.

[13] Tyralis, H., 2016. **HKprocess**: Hurst-Kolmogorov Process. R package version 0.0-2.

[14] Veenstra, J. Q. and McLeod, A. I., 2015. **arfima**: Fractional ARIMA (and Other Long Memory) Time Series Modeling. R package version 1.3-4.

[15] Witt, A. and Malamud, B.D., 2013. Quantification of Long-Range Persistence in Geophysical Time Series: Conventional and Benchmark-Based Improvement Techniques. Surveys in Geophysics, 34 (5), 541-651. doi: 10.1007/s10712-012-9217-8.

[16] Wuertz, D., 2013. **fArma**: ARMA Time Series Modelling. R package version 3010.79.