Comparison between stochastic and machine learning methods for hydrological multi-step ahead forecasting: All forecasts are wrong!

Georgia Papacharalampous, Hristos Tyralis, and Demetris Koutsoyiannis

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

European Geosciences Union
General Assembly 2017
Vienna | Austria | 23-28 April 2017

Session NP4.1/AS4.13/CL5.06: Time Series Analysis, Prediction, Verification and Inter-Comparison of Geoscientific Observations and Model Data (co-organized)

Time series forecasting in hydrology and beyond

- Time series forecasting is of great importance in operational hydrology (Wang et al. 2009).
- Machine learning (ML) algorithms are widely used as an alternative to stochastic methods. Popular ML algorithms are the:
  - ✓ **Neural Networks (NN)**
  - ✓ **Random Forests (RF)**
  - ✓ **Support Vector Machines (SVM)**
- Research often focuses on comparing ML forecasting methods to stochastic.
- The comparisons performed are usually based on **case studies.**
- Within the field of **hydrology**:

  Large number of relative studies, e.g. Jain et al. (1999), Kisi (2004), Khan and Coulibaly (2006), Lin et al. (2006), Han et al. (2007), Yu and Liong (2007), Koutsoyiannis et al. (2008), Wang et al. (2009), Chen et al. (2012), Kisi et al. (2012), Valipour et al. (2013), Belayneh et al. (2014), Patel and Ramachandran (2014), Papacharalampous et al. (2017b)

The broader perspective

- There are **theoretical questions** regarding time series forecasting remaining unanswered in the literature.
- **Generalized results** could be derived through the conduct of large-scale computational experiments, which use real-world and/or simulated data in conjunction with statistical methods.
- Beyond the field of hydrology there are some **few examples** of studies pursuing generalized results to some extent, namely:
  - ✓ Makridakis and Hibon (2000): Comparison between 23 stochastic and one ML forecasting methods on 3 003 historical time series.
  - ✓ Ahmed et al. (2010): Comparison between 8 ML forecasting methods on 1 045 historical time series.
  - ✓ Zhang (2001): Comparison between 25 NN and 3 ARIMA non-automatic forecasting models on 8 stochastic processes from the ARMA family and 30 simulated time series for each stochastic process.
  - ✓ Thissen et al. (2003): Comparison between one stochastic and 2 machine learning non-automatic forecasting methods on two long simulated time series and one historical time series.

An extensive comparison within the field of hydrology

- We performed an **extensive comparison** between several stochastic and ML methods for the **multi-step ahead** forecasting of hydrological processes.
- The comparison is conducted in both:
  - ✓ theoretical level → derivation of generalized results

    contribution in hydrology and beyond
  - ✓ empirical level → reinforcement of the findings

    highlight of important facts
- The theoretical comparison is available in Papacharalampous et al. (2017a), while the preliminary form of the latter study can be found in Papacharalampous (2016).
- The **innovation** of the present study lays on the use of:
  - ✓ **simulated time series**, which compose a wide range of different cases
  - ✓ a sufficient number of forecasting methods from both the stochastic and the ML categories
  - ✓ an adequate number of metrics corresponding to several criteria for the comparative assessment of the forecasting methods
- Papacharalampous et al. (2017d) is a companion to the present study.

Research questions regarding time series forecasting

- **Do the ML methods exhibit different forecasting performance from the stochastic?**
- To which extent might the performance of a specific forecasting method differ across the various time series?
- **Do sophisticated forecasting methods necessarily provide better forecasts than simple forecasting methods?**
- Are there forecasting methods standing out because of their good or bad performance?
- Is it possible to name several advantages/disadvantages of the forecasting methods?
- Is the classification of the forecasting methods possible?
- Moreover, is a general ranking of the forecasting methods possible?
- Finally, **can we decide on a universally best forecasting method?**

A methodological framework aiming at generalized results

- We conduct **12 large-scale simulation experiments**.
- Additionally, we conduct **92 case studies**.
- We compare **20 forecasting methods**.
- We quantify the forecast quality using **18 metrics**.

<u>Time series</u>

- ✓ 12 x **2 000 time series** of **310 values**, resulted from the simulation of ARMA and ARFIMA processes
- ✓ **92** mean montly time series of streamflow, which contain at least 10 years of continuous observations

<u>Forecasting methods</u>

- ✓ 11 stochastic methods originating from the families: ARMA, ARIMA, ARFIMA, Exponential Smoothing, State Space
- ✓ 9 ML methods: 3 NN methods, 3 RF methods, 3 SVM methods

<u>Metrics</u>

18 metrics providing assessment regarding the following criteria:

- ✓ accuracy
- ✓ capture of the variance
- ✓ correlation

- The metrics are computed on test set, which is the last 10 values for the simulated time series and the last 12 values for the real-world time series.

R code

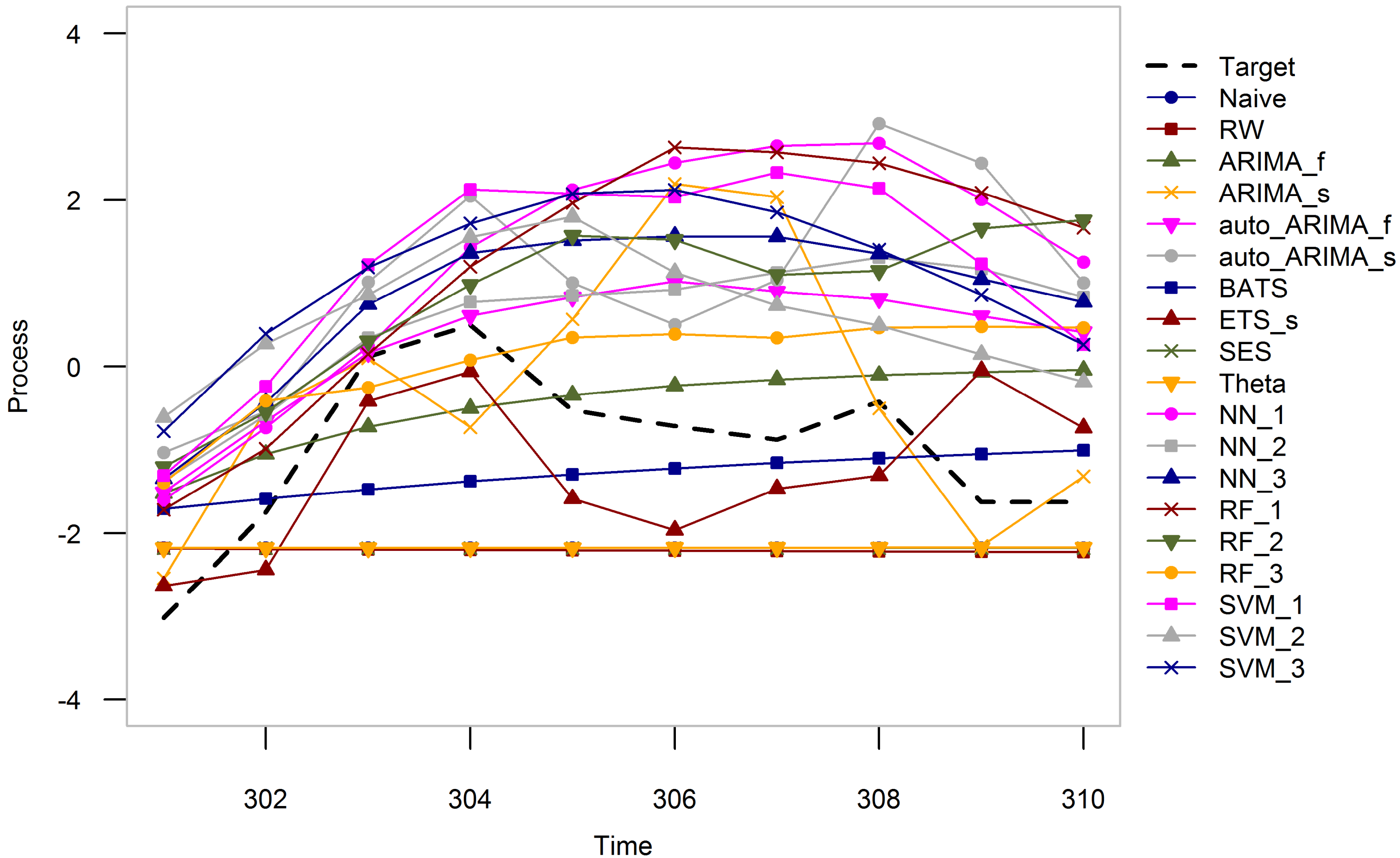
- We combined functions mainly originating from the following packages.
  - ✓ **stats** (R Core Team 2017)
  - ✓ **CombMSC** (Smith 2012)
  - ✓ **EnvStats** (Millard 2013)
  - ✓ **forecast** (Hyndman and Khandakar 2008, Hyndman et al. 2017)
  - ✓ **fracdiff** (Fraley et al. 2012)
  - ✓ **ggExtra** (Attali 2016)
  - ✓ **ggplot2** (Wickham 2009)
  - ✓ **HKprocess** (Tyralis 2016, Tyralis and Koutsoyiannis 2011)
  - ✓ **kernlab** (Karatzoglou et al. 2004)
  - ✓ **nnet** (Venables and Ripley 2002)
  - ✓ **randomForest** (Liaw and Wiener 2002)
  - ✓ **rminer** (Cortez 2010, 2016)

built-in R functions

R packages

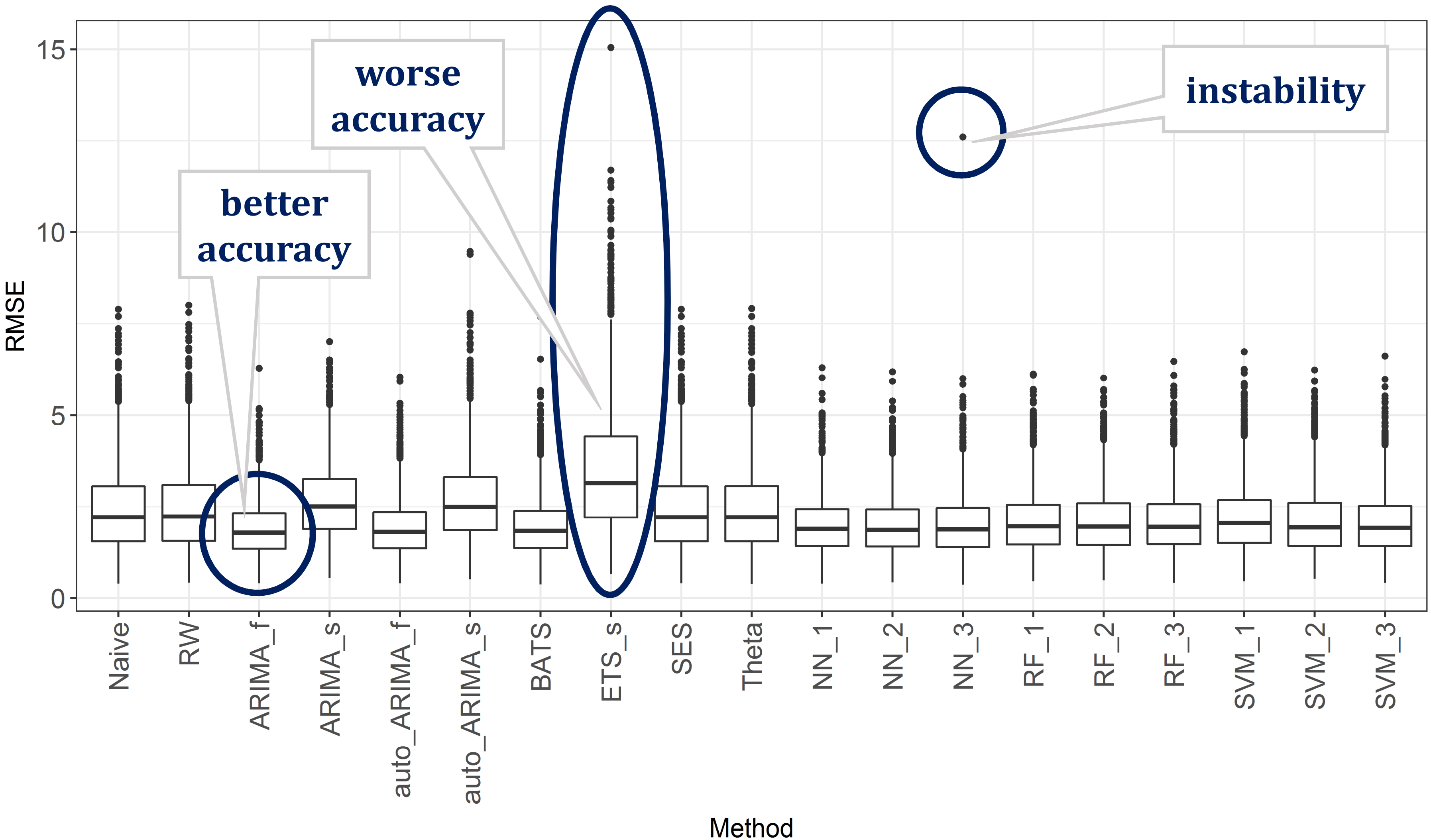
Forecasting examples on simulated time series

time series of 310 values resulting from the simulation of ARMA(1,1)



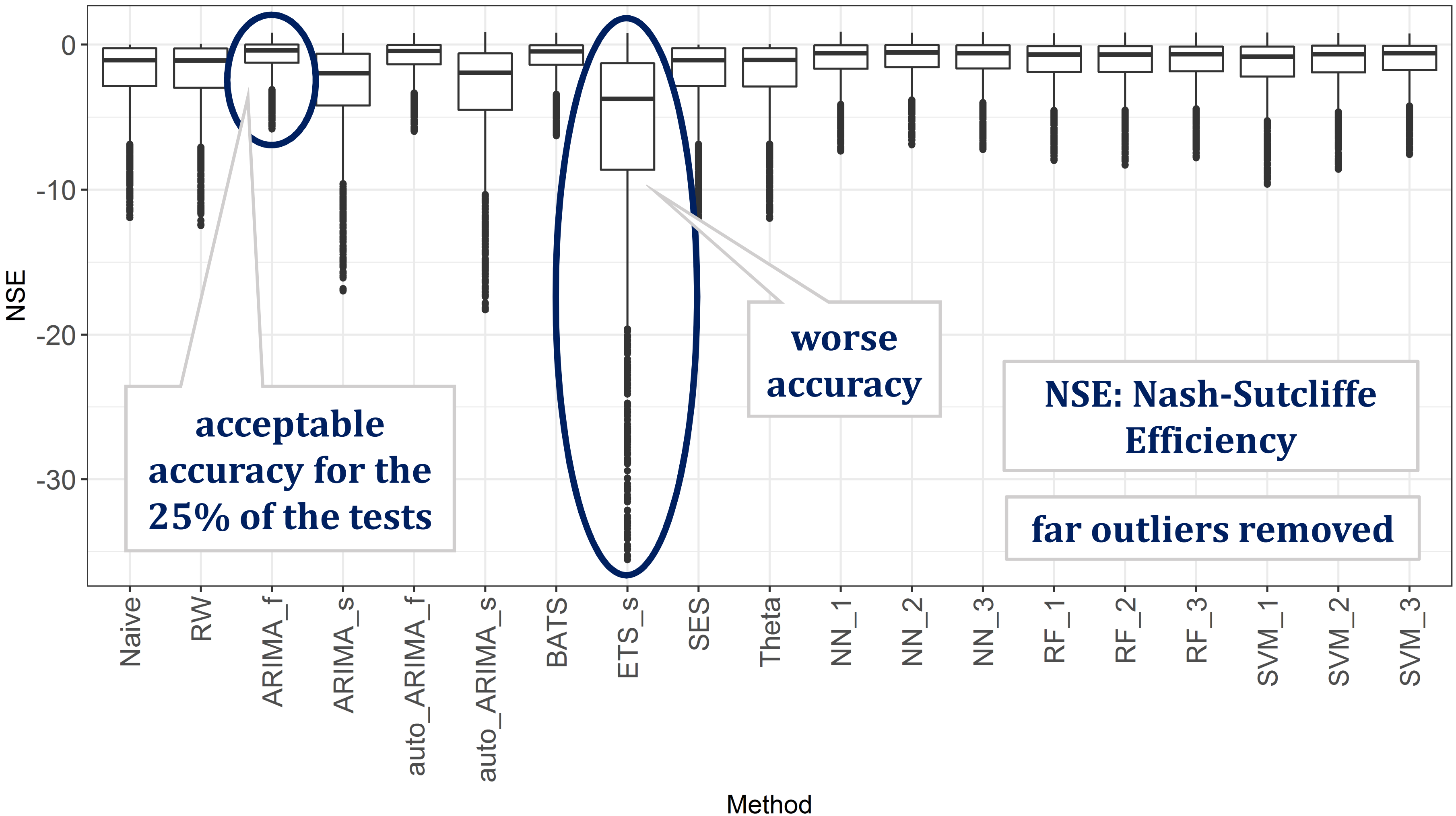
Comparison of the forecasting methods in terms of RMSE

2 000 time series of 310 values resulting from the simulation of ARMA(1,1)



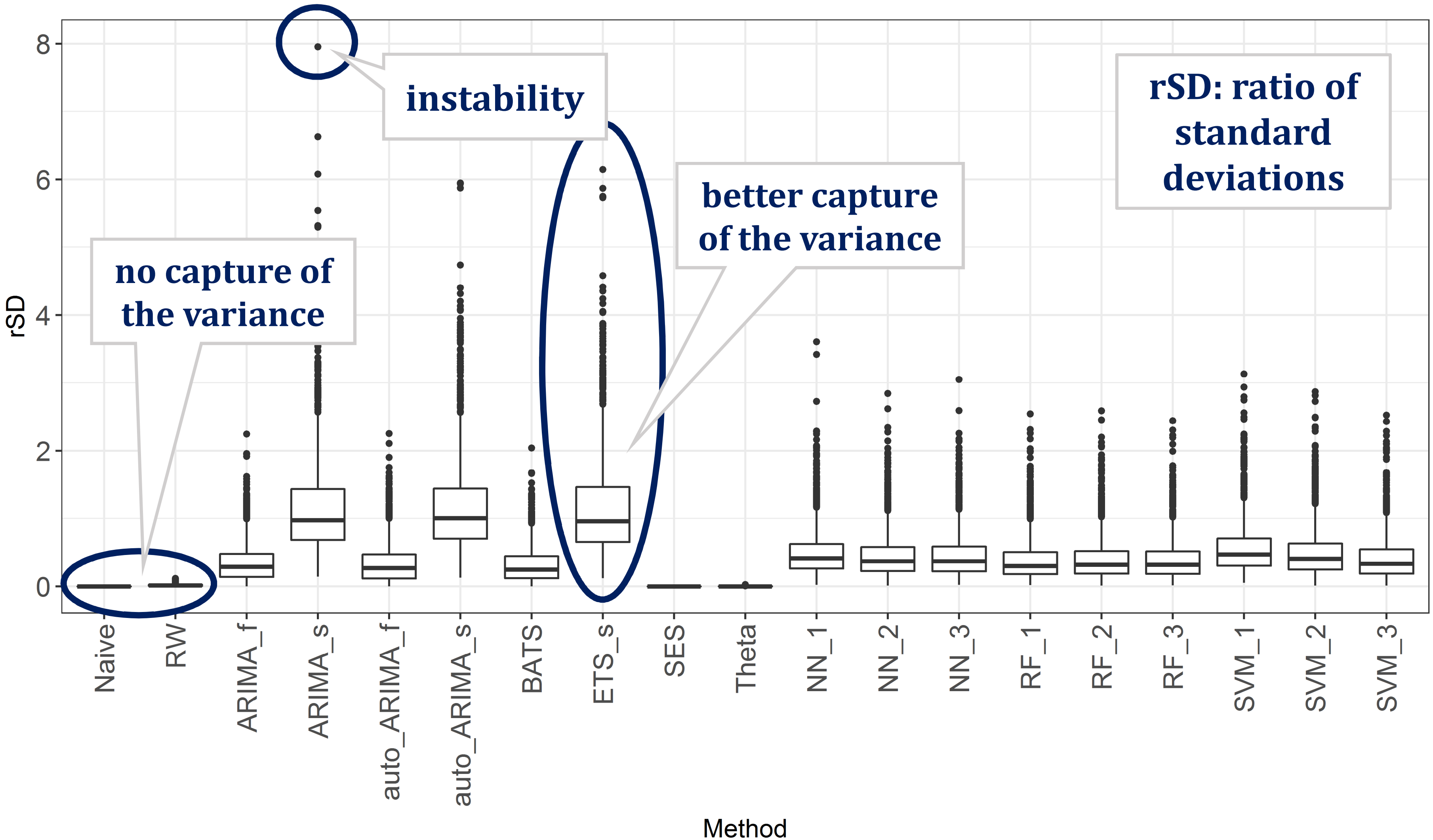
Comparison of the forecasting methods in terms of NSE

2 000 time series of 310 values resulting from the simulation of ARMA(1,1)

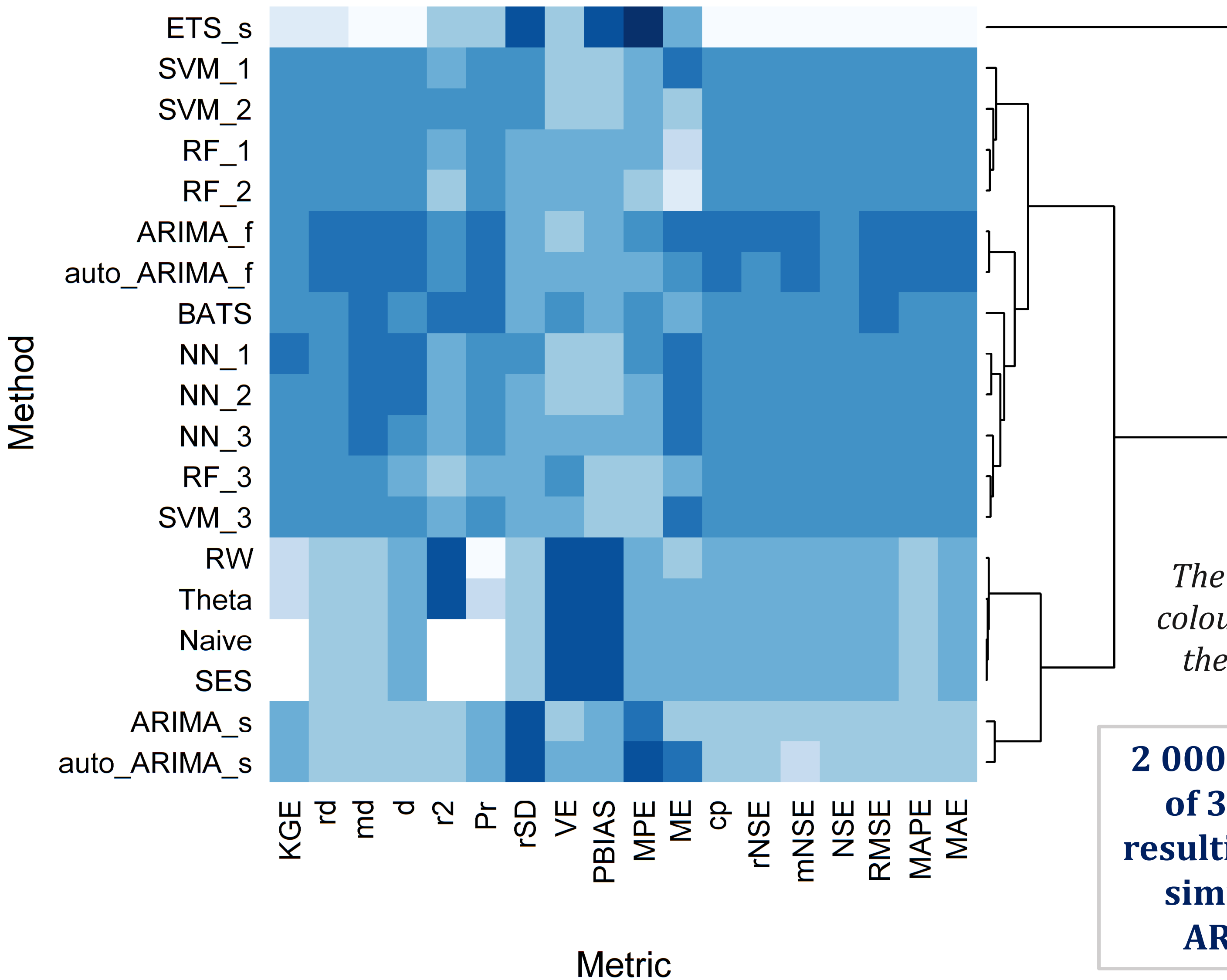


Comparison of the forecasting methods in terms of rSD

2 000 time series of 310 values resulting from the simulation of ARMA(1,1)



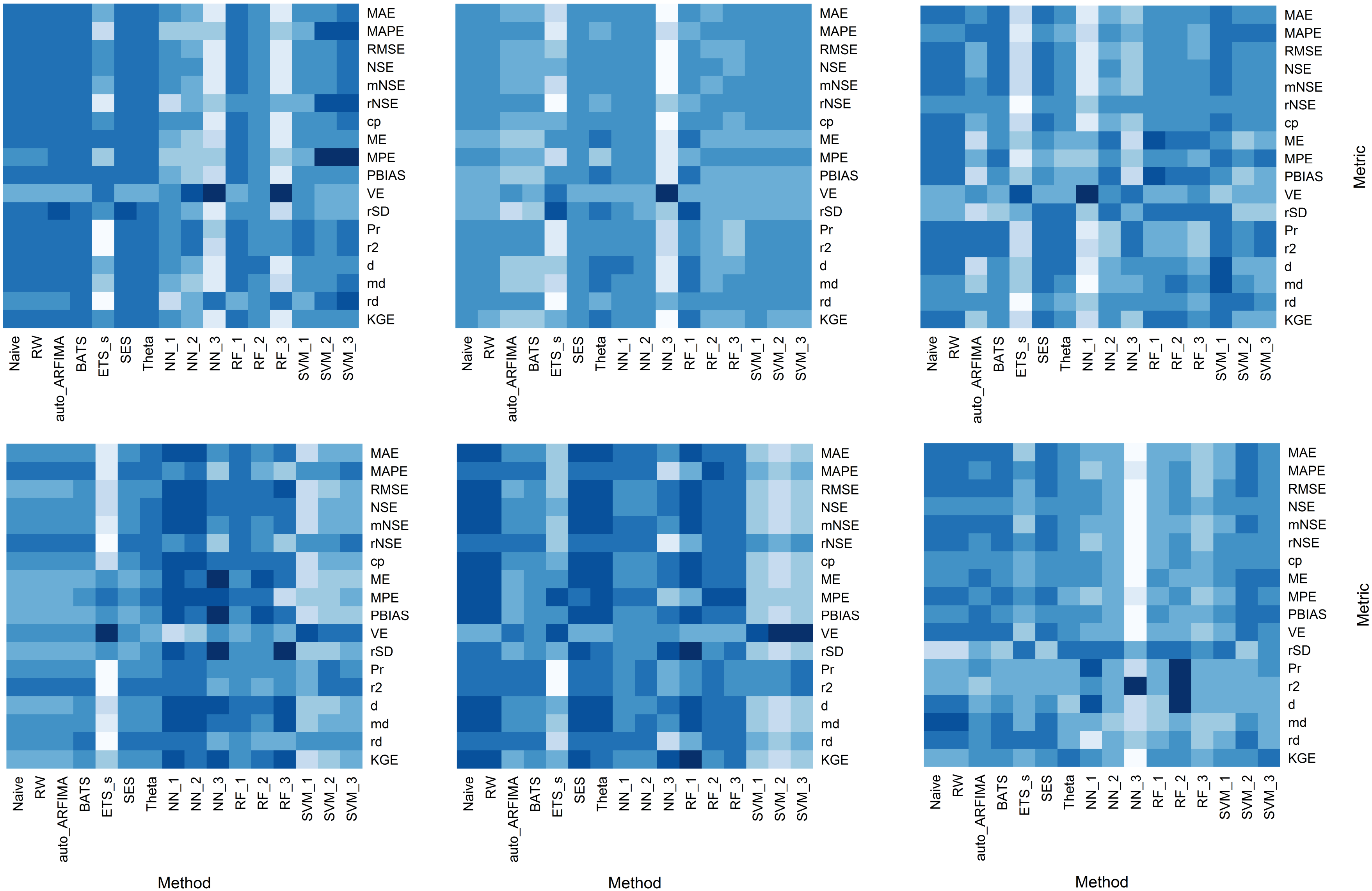
Average-case performance of the forecasting methods



The darker the colour the better the forecasts.

2 000 time series of 310 values resulting from the simulation of ARMA(1,1)

Case studies using mean monthly time series of streamflow



Contribution in hydrology and beyond

- There is not a universally best or worst forecasting method.

All forecasts are wrong!

- There are forecasting methods regularly better or worse than others with respect to specific metrics, while there are also forecasting methods sharing a quite similar performance.

Some forecasts are more useful than others!

- More sophisticated methods do not necessarily provide better forecasts compared to simpler methods.
- Although a general ranking of the forecasting methods is not possible, their classification based on their similar or contrasting performance in the various metrics is possible to some extent.
- The ML methods **do not differ dramatically** from the stochastic, except for the fact that the former are computationally intensive.
- The forecasting methods resulting from the implementation of the same algorithm can exhibit a far distant performance.

Contribution in hydrology and beyond

- The values of the metrics can **vary significantly** across the different simulation experiments and across the different time series within a specific simulation experiment.
- The above worded findings highlight the efficiency of our methodological approach in producing **generalized results**.
- The empirical comparison emphasizes on important issues, which exhibit greater interest when presented on real-world data, while it also reinforces the findings of the theoretical comparison.
- Someone who examines both the results of the simulation experiments and the case studies has a more complete picture of the underlying phenomena than whom considering only the results of the simulation experiments.
- The **use of simulated processes** has proved pivotal in delivering the pursued generalization.
- Using fewer forecasting methods and fewer metrics would have led to a very different overall picture, particularly if those fewer metrics corresponded to fewer criteria.

Recommendations for further research

- We propose the conduct of several large-scale simulation experiments for the comparison of stochastic and ML forecasting methods regarding their **one-step ahead forecasting** properties, which are also of practical interest.
- We recommend the adoption of our methodological framework for the assessment of any forecasting method exhibiting theoretical and/or practical interest.
- The understanding of the **theoretical properties** of the forecasting methods, presupposes systematic and focused on each of them research.
- The investigation of the capabilities that each metric provides regarding the quantification of the forecasting methods' performance is also required.
- Regarding time series forecasting using ML algorithms, we recommend the conduct of an extensive study aiming at the investigation of the effect of the **hyperparameter optimization** and the **time lag selection**. The latter requires an expansion of the existing relative studies:

  Zhang (2001), Papacharalampous et al. (2017b), Papacharalampous et al. (2017c)
- Above all, the intensification of the research on **probabilistic forecasting** (e.g. Tyralis and Koutsoyiannis 2014) and its effective exploitation by the users (e.g. Ramos et al. 2013) should be thoroughly considered by the scientific community.

References

Ahmed, N.K., Atiya, A.F., Gayaran, N.E., and El-Shishiny, H., 2010. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29 (5-6), 594-621. doi:10.1080/07474938.2010.481556

Attali, D., 2016. ggExtra: Add Marginal Histograms to 'ggplot2', and More 'ggplot2' Enhancements. R package version 0.5.

Belayneh, A., Adamowski, J., Khalil, B., and Ozga-Zielinski, B., 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508, 418-429. doi:10.1016/j.jhydrol.2013.10.052

Chen, J., Li, M., and Wang, W., 2012. Statistical uncertainty estimation using random forests and its application to drought forecast. *Mathematical Problems in Engineering*, 2012 (915053). doi:10.1155/2012/915053

Cortez, P., 2010. Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool. *In*: P. Perner, eds. *Advances in Data Mining. Applications and Theoretical Aspects*. Springer Berlin Heidelberg, pp 572-583. doi:10.1007/978-3-642-14400-4_44

Cortez, P., 2016. rminer: Data Mining Classification and Regression Methods. R package version 1.4.2.

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.

Han, D., Chan, L., and Zhu, N., 2007. Flood forecasting using support vector machines. *Journal of Hydroinformatics*, 9 (4), 267-276. doi:10.2166/hydro.2007.027

Hyndman, R.J., and Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27 (3), 1-22. doi:10.18637/jss.v027.i03

Hyndman, R.J., 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.

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A., 2004. kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11 (9), 1-20.

Khan, M.S., and Coulibaly, P., 2006. Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering*, 11 (3), 199-205. doi:10.1061/(ASCE)1084-0699(2006)11:3(199)

Kisi, O., 2004. River flow modeling using artificial neural networks. *Journal of Hydrologic Engineering*, 9 (1), 60-63. doi:10.1061/(ASCE)1084-0699(2004)9:1(60)

Kisi, O., Shiri, J., and Nikoofar, B., 2012. Forecasting daily lake levels using artificial intelligence approaches. *Computers & Geosciences*, 41, 169-180. doi:10.1016/j.cageo.2011.08.027

Koutsoyiannis, D., Yao, H., and Georgakakos, A., 2008. Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. *Hydrological Sciences Journal*, 53 (1), 142-164. doi:10.1623/hysj.53.1.142

Liaw, A., and Wiener, M., 2002. Classification and regression by randomForest. *R News*, 2 (3), 18-22.

Lin, J.Y., Cheng, C.T., Chau, K.W., 2006. Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51 (4), 599-612. doi:10.1623/hysj.51.4.599

Makridakis, S., and Hibon, M., 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16 (4), 451-476. doi:10.1016/S0169-2070(00)00057-1

Millard, S.P., 2013. EnvStats: An R Package for Environmental Statistics. *In*: G. Høst, eds. *Encyclopedia of Environmetrics*. doi:10.1002/9780470057339.vae043.pub2

References

Papacharalampous, G.A., 2016. Theoretical and empirical comparison of stochastic and machine learning methods for hydrological processes forecasting. MSc thesis.

Papacharalampous, G.A., Tyralis, H., and Koutsoyiannis, D., 2017a. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Journal of Hydrology*. In review.

Papacharalampous, G.A., Tyralis, H., and Koutsoyiannis, D., 2017b. Forecasting of geophysical processes using stochastic and machine learning algorithms. Proceedings of the 10th World Congress of EWRA on Water Resources and Environment.

Papacharalampous, G.A., Tyralis, H., and Koutsoyiannis, D., 2017c. Investigation of the effect of the hyperparameter optimization and the time lag selection in time series forecasting using machine learning algorithms. European Geosciences Union General Assembly 2017, Geophysical Research Abstracts, 19, Vienna.

Papacharalampous, G.A., Tyralis, H., and Koutsoyiannis, D., 2017d. Multi-step ahead streamflow forecasting for the operation of hydropower reservoirs. European Geosciences Union General Assembly 2017, Geophysical Research Abstracts, 19, Vienna.

Patel, S.S., and Ramachandran, P., 2015. A comparison of machine learning techniques for modeling river flow time series: the case of upper Cauvery river basin. *Water Resources Management*, 29 (2), 589-602. doi:10.1007/s11269-014-0705-0

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramos, M.H., Van Andel, S.J., and Pappenberger, F., 2013. Do probabilistic forecasts lead to better decisions?. *Hydrology and Earth System Sciences*, 17, 2219-2232. doi:10.5194/hess-17-2219-2013

Smith, A.K., 2012. CombMSC: Combined Model Selection Criteria. R package version 1.4.2.

Thissen, U., Van Brakel, R., De Weijer, A.P., Melssena, W.J., and Buydens, L.M.C., 2003. Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems*, 69 (1-2), 35-49. doi:10.1016/S0169-7439(03)00111-4

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

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

Tyralis, H., and Koutsoyiannis, D., 2014. A Bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. *Climate Dynamics*, 42 (11), 2867-2883. doi:10.1007/s00382-013-1804-y

Valipour, M., Banihabib, M.E., and Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology*, 476 (7), 433-441. doi:10.1016/j.jhydrol.2012.11.017

Venables, W.N., and Ripley, B.D., 2002. *Modern Applied Statistics with S, fourth edition*. New-York: Springer-Verlag. doi:10.1007/978-0-387-21706-2

Wang, W.C., Chau, K.W., Cheng, C.T., and Qiu, L., 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374 (3-4), 294-306. doi:10.1016/j.jhydrol.2009.06.019

Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Yu, X., and Liong, S.Y., 2007. Forecasting of hydrologic time series with ridge regression in feature space. *Journal of Hydrology*, 332 (3-4), 290-302. doi:10.1016/j.jhydrol.2006.07.003

Jain, S.K., Das, A., and Srivastava, D.K., 1999. Application of ANN for reservoir inflow prediction and operation. *Journal of Water Resources*, 125 (5), 263-271. doi:10.1061/(ASCE)0733-9496(1999)125:5(263)

Zhang, G.P., 2001. An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research*, 28 (12), 1183-1202. doi:10.1016/S0305-0548(00)00033-2