



## **Univariate time series forecasting properties of random forests**

Hristos Tyralis (1) and Georgia Papacharalampous (2)

(1) National Technical University of Athens, School of Civil Engineering, Department of Water Resources and Environmental Engineering, Athens, Greece (montchrister@gmail.com), (2) National Technical University of Athens, School of Civil Engineering, Department of Water Resources and Environmental Engineering, Athens, Greece (papacharalampous.georgia@gmail.com)

The random forests' univariate time series forecasting properties have remained unexplored. Here we assess the performance of random forests in one-step forecasting using two large datasets of short time series with the aim to suggest an optimal set of predictor variables. Furthermore, we compare their performance to benchmarking methods. The first dataset consists of 16 000 simulated time series from a variety of Autoregressive Fractionally Integrated Moving Average (ARFIMA) models. The second dataset consists of 135 mean annual temperature time series. The random forests performed better mostly when using a few recent lagged predictor variables. A possible explanation of this result is that increasing the number of lagged variables decreases the length of the training set and simultaneously decreases the information exploited from the original time series during the model fitting phase. Furthermore, the random forests were comparable to the benchmarking methods.