Large scale simulation experiments for the assessment of one-step ahead forecasting properties of stochastic and machine learning point estimation methods

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)
Background information

- The scientific literature includes a large number of studies assessing the one-step ahead forecasting performance of stochastic and/or machine learning methods when applied to geophysical processes within case studies, e.g.:

  Lambrakis et al. (2000); Ballini et al. (2001); Yu et al. (2004); Yu and Liong (2007); Hong (2008); Koutsoyiannis et al. (2008); Papacharalampous et al. (2017b)

- However, generalized information about the forecasting methods cannot be extracted from case studies.

- Makridakis and Hibon (2000) presented the results of the M3-Competition. In the latter the one- and multi-step ahead forecasting performance of several methods were assessed on 3 003 real-world time series.

- Recently, Papacharalampous et al. (2017a) compared several stochastic and machine learning methods regarding their multi-step ahead forecasting properties when applied to stationary stochastic processes. The methods were tested on 48 000 simulated time series.

- In a similar vein, Tyralis and Papacharalampous (2017) compared several random forests methods regarding their one-step ahead forecasting performance on 16 000 simulated time series. The aim was to suggest an optimal set of time lags to be used in the fitting phase.
The present study

- We have focused on one-step ahead forecasting in geoscience.
- We have conducted 12 large scale simulation experiments.
- Additionally, we have conducted a real-world multiple-case study.
- We have compared 20 forecasting methods.

<table>
<thead>
<tr>
<th>Time series</th>
<th>Forecasting methods (see 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ 12 x 2 000 time series of 100 values, resulted from the simulation of ARMA and ARFIMA processes (see 4), which are widely used for the modelling of geophysical processes</td>
<td>✓ 11 stochastic methods originating from the families: simple, ARMA, ARIMA, ARFIMA, exponential smoothing, state space</td>
</tr>
<tr>
<td>✓ 135 mean annual time series of temperature, which contain 100 continuous observations (see 5)</td>
<td>✓ 9 machine learning methods originating from the families: neural networks, random forests, support vector machines</td>
</tr>
</tbody>
</table>

- The comparative assessment of the methods has been based on the error and the absolute error of the forecast of the last value.
The simulations were performed with zero mean and standard deviation of 1.

The definitions of the ARMA and ARFIMA stochastic processes can be found in Wei (2006).
Real-world time series

135 mean annual time series of temperature

Hurst parameter estimation

- R package HKprocess (Tyralis 2016, see also Tyralis and Koutsoyiannis 2011)
### Forecasting methods

<table>
<thead>
<tr>
<th></th>
<th>Naive simple</th>
<th>RW simple</th>
<th>ARIMA_f ARMA</th>
<th>ARIMA_s ARMA</th>
<th>auto_ARIMA_f ARIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>auto_ARIMA_s ARIMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>auto_ARFIMA ARFIMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BATS state space</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETS_s state space</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES exponential smoothing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theta exponential smoothing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN_1 neural networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN_2 neural networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN_3 neural networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF_1 random forests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF_2 random forests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF_3 random forests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM_1 support vector machines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM_2 support vector machines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM_3 support vector machines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Implementation of the forecasting methods
- R package kernlab (Karatzoglou et al. 2004)
- R package nnet (Venables and Ripley 2002)
- R package randomForest (Liaw and Wiener 2002)
- R package rminer (Cortez 2010, 2016)
Simulation experiments: Errors

![Box plots showing errors for different forecasting methods.](image)

- SE_1
- SE_2
- SE_9
- SE_12

### Forecasting Method

- Naïve
- RW
- ARIMA
- ARIMA	
- auto_ARIMA
- auto_ARIMA	
- BATS
- ETS
- SES
- Theta
- NN_1
- NN_2
- NN_3
- RF_1
- RF_2
- RF_3
- SVM_1
- SVM_2
- SVM_3
Simulation experiments: Absolute errors

Absolute errors

[Box plots showing absolute errors for different forecasting methods]

Forecasting Method

SE_6

SE_8

SE_7

SE_10
Simulation experiments: Average-case performance

The darker the colour the better the forecasts.
135 temperature time series: Forecasted vs observed
The darker the colour the better the forecasts.
Summary and conclusions

- We have conducted **large scale simulation experiments** for the assessment of the one-step ahead forecasting properties of several stochastic and machine learning point estimation methods.

- Our findings indicate that **the results can vary significantly** across the different simulation experiments and across the different time series.

- ARIMA_f, auto_ARIMA_f and BATS were proven to be the most accurate forecasting methods on the ARMA processes. The same applies to auto_ARFIMA, BATS, SES and Theta on the ARFIMA processes.

- The simple forecasting methods (Naive and RW) are also competent.

- Most of the observed **far outliers** were produced by neural networks.

- We have additionally applied our methodology to 135 mean annual time series of temperature.

- The **Theta** method, presented by Assimakopoulos and Nikolopoulos (2000), exhibited the best performance within this real-world case study being slightly better than BATS and SES.
References


