A step further from model-fitting for the assessment of the predictability of monthly temperature and precipitation

Georgia Papacharalampous, Hristos Tyralis, and Demetris Koutsoyiannis

European Geosciences Union
General Assembly 2018
Vienna | Austria | 8-13 April 2018

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

Presentation code: EGU2018-864
Available online at: itia.ntua.gr/1808

1. Abstract

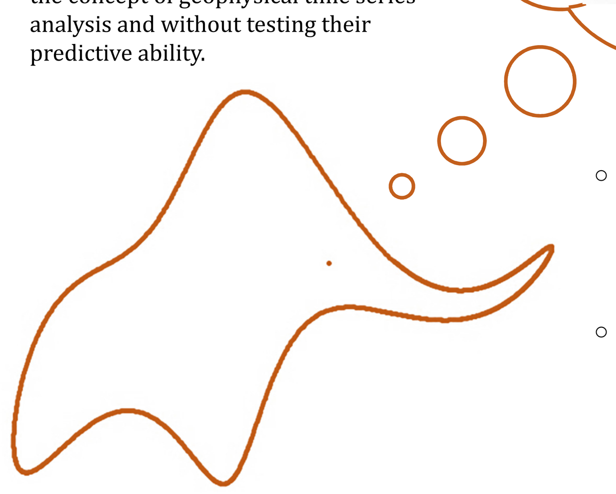
"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk", ~ John von Neumann. This famous quote, literally possible as proved by Mayer et al. (2010), has been widely used to question the parsimony of a model providing a good description of the available data. Still, a significant part of the hydrological literature insists in adding parameters, trend or of other type, to models to increase their descriptive power within the concept of geophysical time series analysis and without testing their predictive ability. Herein, we move a step further from model-fitting and actually run in forecast mode several automatic univariate time series models with the aim to assess the predictability of monthly temperature and precipitation. We examine a sample of 985 monthly temperature and 1552 monthly precipitation time series, observed at stations covering a significant part of the Earth's surface and, therefore, including various real-world process behaviours. All the time series are 40-years long with no missing values. We compare the naïve based on the monthly values of the last year, ARFIMA, exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (BATS), simple exponential smoothing (SES), Theta and Prophet forecasting methods. Prophet is a recently introduced model inspired by the nature of time series forecasted at Facebook and has not been applied to hydrometeorological time series in the past, while the use of BATS, SES and Theta is rare in hydrology. The methods are tested in performing multi-step ahead forecasts for the last 48 months of the data. The results are summarized in global scores, while their examination by group of stations leads to 5 individual scores for temperature and 6 for precipitation. The groups are formed according to the geographical vicinity of the stations.

The findings suggest that all the examined models are accurate enough to be used in long-term forecasting applications. For the total of the temperature time series the use of an ARFIMA, BATS, SES, Theta or Prophet model, instead of the naïve method, leads in about 19-29% more accurate forecasts in terms of root mean square error, or even in about 30-32% more accurate forecasts specifically for the temperature time series observed in North Europe. For the total of the precipitation time series the use of all these automatic methods leads in about 21-22% better forecasts than the use of the naïve method, while for the geographical regions of North America, North Europe and East Asia these percentages are 26-29%, 22-24% and 32-38% respectively. We think that the level of the forecasting accuracy can barely be improved using other methods, as indicated by the experiments of Papacharalampous et al. (2017a).

2. Introduction

"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk", ~ John von Neumann.

- Meaning of this quote: **We should not be surprised by the descriptive ability of a model comprising a large number of parameters.**
- Still, a significant part of the hydrological literature insists in adding parameters, trend or of other type, to models to increase their descriptive power within the concept of geophysical time series analysis and without testing their predictive ability.



Is this quote literally possible?

- An implementation is provided by Mayer et al. (2010). The authors use *"five complex parameters"*, each holding a real and an imaginary part , i.e. 10 parameters in total. The real part of the fifth parameter is the wiggle parameter.
- A relevant discussion and a Python code for the implementation of the method of Mayer et al. (2010), i.e. for drawing the orange elephant on the left, can be found at: https://www.johndcook.com/blog/2011/06/21/how-to-fit-an-elephant/

3. Forecasting: A step further from model-fitting

- We investigate the predictability of monthly temperature and precipitation, and simultaneously assess the multi-step ahead performance of the automatic univariate time series forecasting methods (presented on the right) by applying the latter to the **largest sample of hydrometeorological time series ever used for such purposes**.
- This sample is composed by 985 monthly temperature and 1552 monthly precipitation time series. All the time series are 40-years long, spanning from 1950 to 1989, with no missing values.
- We forecast the monthly values of each time series for the period 1986-1989 using the time series values from the period 1950-1985. The forecast horizon is 48 months ahead.
- For the assessment of the methods we compute the error and absolute error at each time step of the forecast horizon for each forecasting attempt. We further compute the Root Mean Squared Error (RMSE) and the Nash-Sutcliffe Efficiency (NSE) of each multi-step ahead forecast.
- The analyses and visualizations are performed in R Programming Language (R Core Team 2017) using the contributed R packages devtools (Wickham and Chang 2017), forecast (Hyndman and Khandakar 2008, Hyndman et al. 2017), fracdiff (Fraley et al. 2012), gdata (Warnes et al. 2017), ggplot2 (Wickham 2016), HKprocess (Tyralis 2016, Tyralis and Koutsoyiannis 2011), knitr (Xie 2014; 2015; 2017), maps (Brownrigg et al. 2017), prophet (Taylor and Letham 2017a, b), readr (Wickham et al. 2017) and zoo (Zeileis and Grothendieck 2005).

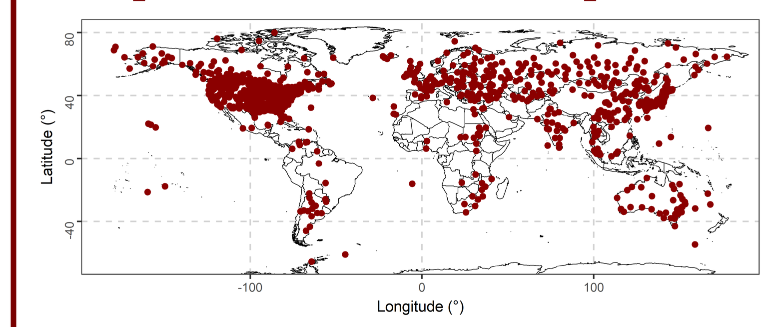
Automatic univariate time series forecasting methods

s/n	Abbreviated name	Type	Handling of seasonality (see bellow)	Handling of non-normality (see bellow)
1	naïve	naïve	1	1
2	arfima_1	AutoRegressive Integrated Moving Average (ARFIMA)	2	1
3	arfima_2		2	2
4	arfima_3		3	1
5	arfima_4		3	2
6	bats_1	exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (BATS)	2	1
7	bats_2		2	2
8	bats_3		3	1
9	bats_4		3	2
10	bats_5		4	1
11	bats_6		4	2
12	ses_1	Simple Exponential Smoothing (SES)	2	1
13	ses_2		2	2
14	ses_3		3	1
15	ses_4		3	2
16	theta_1	Theta	2	3
17	theta_2		3	3
18	prophet_1	Prophet	2	3
19	prophet_2		3	3
20	prophet_3		4	3

s/n	Handling of seasonality
1	Time series offset
2	Classical seasonal decomposition using the additive model of the decompose built-in R algorithm and subsequent addition of the seasonal component to the forecasts
3	Classical seasonal decomposition using the multiplicative model of the decompose built-in R algorithm and subsequent multiplication of the forecasts by the seasonal component
4	Through the forecasting algorithm

s/n	Handling of non-normality
1	-
2	Box-Cox transformation through the forecasting algorithm
3	Default

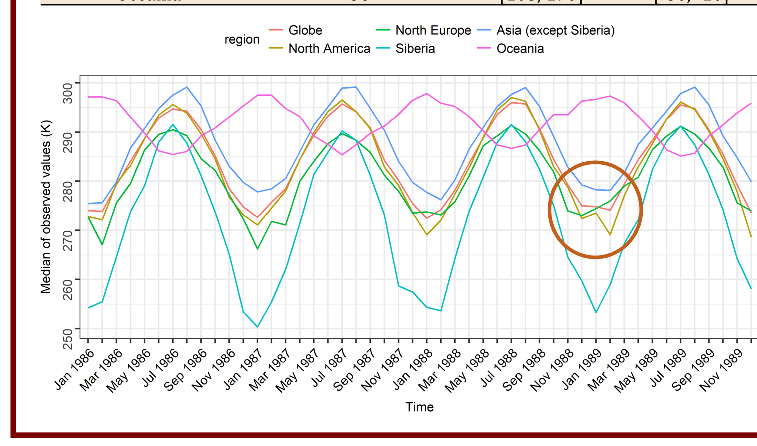
4. Exploration of the temperature dataset

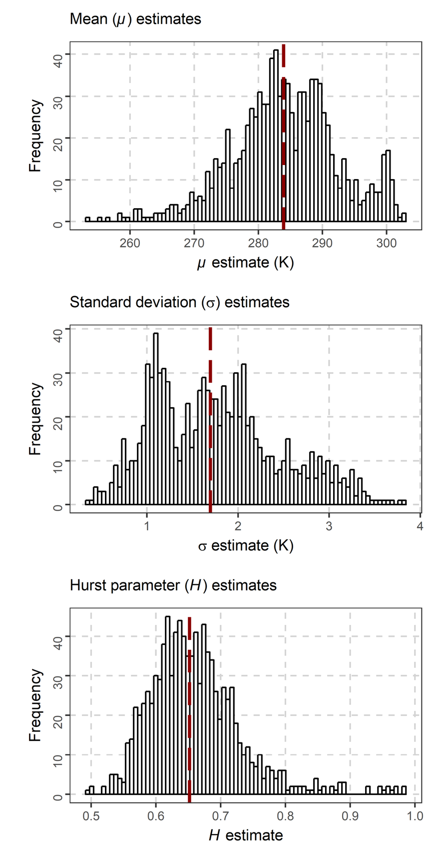


Data source: Lawrimore et al. (2011)

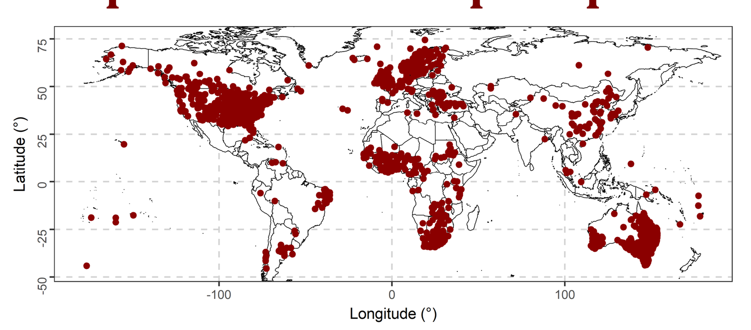
Groups of the temperature stations

Geographical region	Number of stations	Longitude (°)	Latitude (°)
North America	410	[-140, -50]	[20, 65]
North Europe	80	[-15, 40]	[45, 75]
Siberia	70	[40, 175]	[50, 75]
Asia (except Siberia)	259	[40, 150]	[5, 50]
Oceania	36	[105, 170]	[-50, -10]





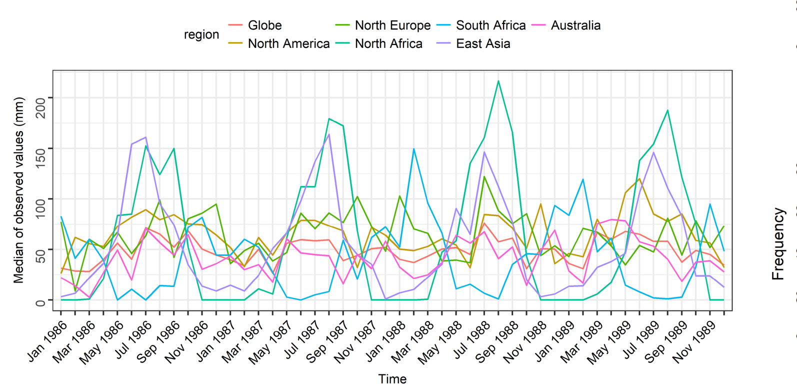
5. Exploration of the precipitation dataset

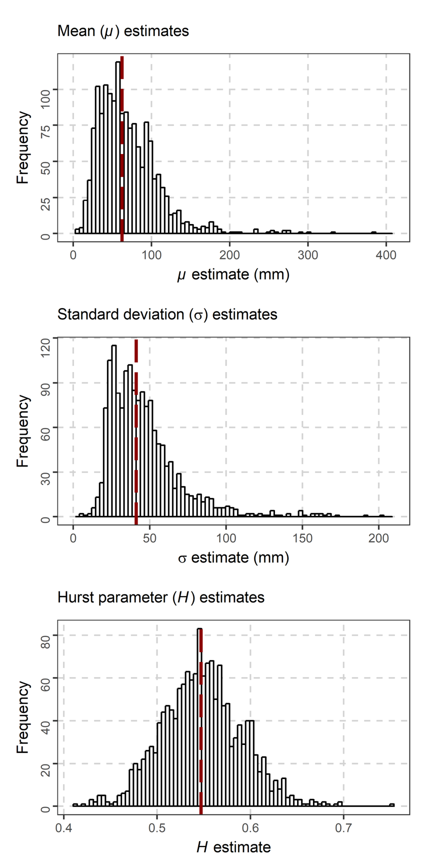


Data source: Peterson and Vose (1997)

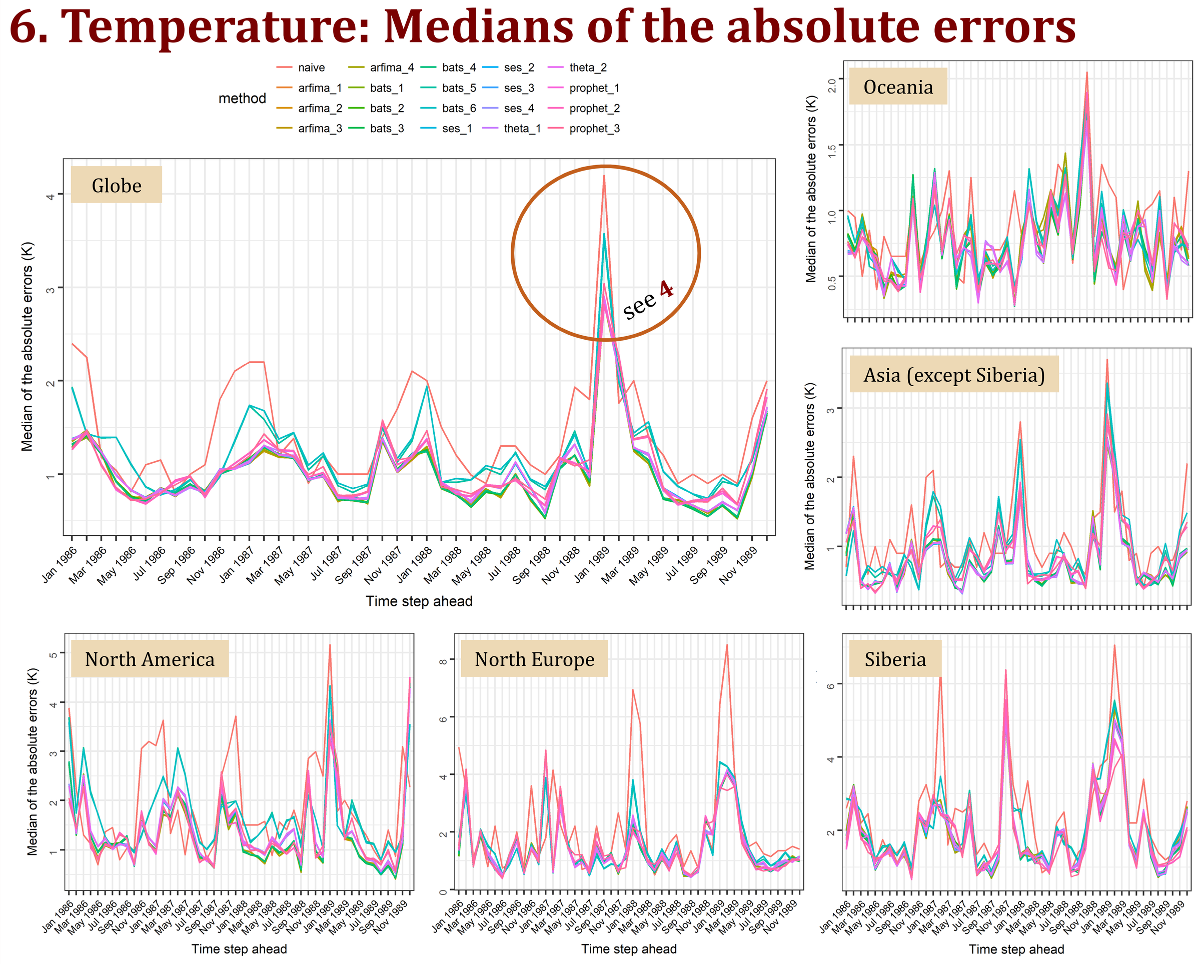
Groups of the precipitation stations

Geographical region	Number of stations	Longitude (°)	Latitude (°)
North America	388	[-135, -60]	[20, 55]
North Europe	182	[-15, 35]	[50, 75]
North Africa	100	[-20, 40]	[0, 20]
South Africa	120	[-20, 45]	[-35, 0]
East Asia	50	[95, 135]	[15, 50]
Australia	563	[110, 155]	[-45, - 15]





6. Temperature: Medians of the absolute errors



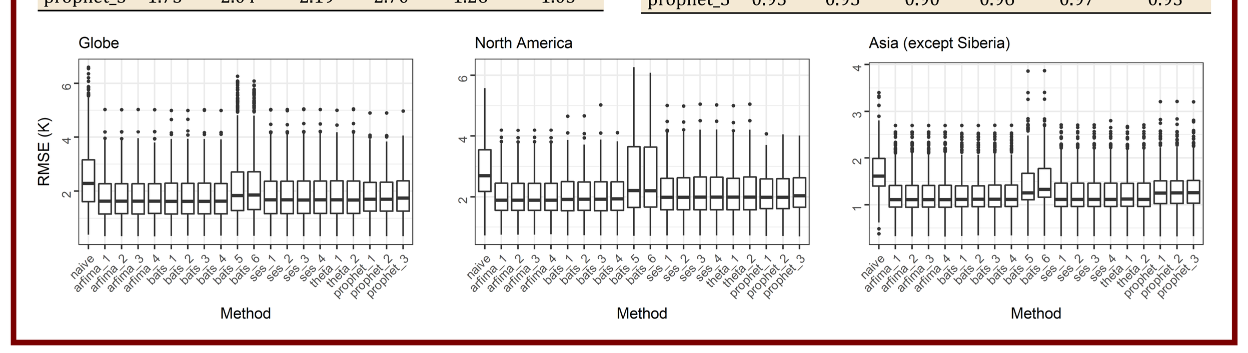
7. Temperature: RMSE and Nash-Sutcliffe values

Medians of the RMSE values (K)

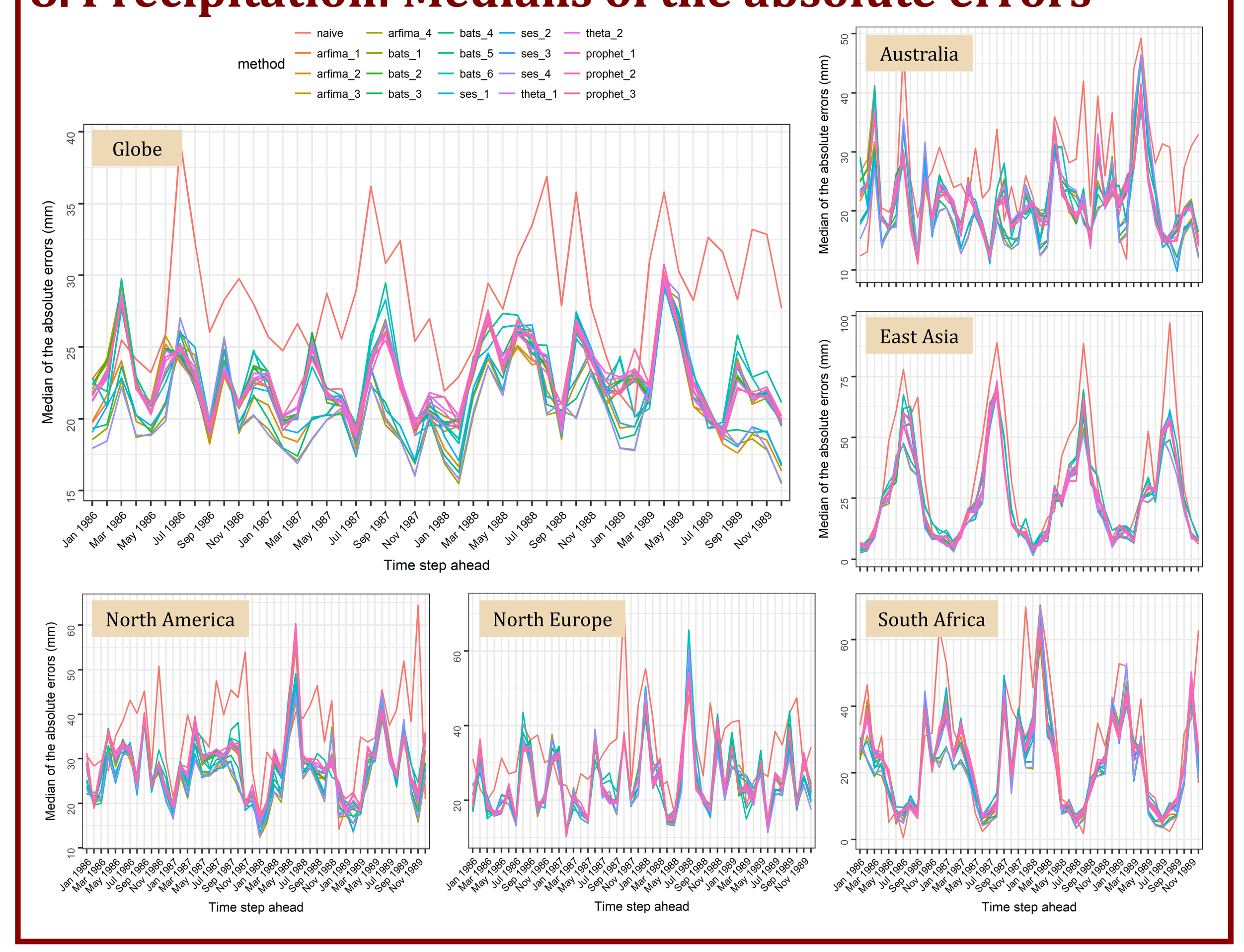
Method	Globe	North America	North Europe	Siberia	Asia (except Siberia)	Oceania
naïve	2.29	2.70	3.14	3.49	1.61	1.19
arfima_1	1.63	1.90	2.13	2.66	1.11	1.03
arfima_2	1.63	1.90	2.13	2.66	1.11	1.04
arfima_3	1.63	1.90	2.13	2.66	1.11	1.03
arfima_4	1.63	1.90	2.13	2.66	1.11	1.06
bats_1	1.62	1.92	2.13	2.66	1.11	1.01
bats_2	1.62	1.93	2.13	2.67	1.12	1.01
bats_3	1.62	1.92	2.13	2.68	1.12	1.01
bats_4	1.64	1.94	2.13	2.67	1.12	1.01
bats_5	1.84	2.21	2.16	2.84	1.26	1.07
bats_6	1.86	2.20	2.18	2.84	1.33	1.04
ses_1	1.68	1.99	2.15	2.66	1.12	1.02
ses_2	1.68	1.99	2.15	2.67	1.12	1.02
ses_3	1.68	2.00	2.15	2.67	1.12	1.02
ses_4	1.68	2.00	2.15	2.68	1.12	1.02
theta_1	1.68	1.99	2.17	2.64	1.12	1.01
theta_2	1.68	2.00	2.15	2.67	1.12	1.02
prophet_1	1.70	1.99	2.12	2.62	1.25	1.03
prophet_2	1.71	1.99	2.13	2.63	1.25	1.03
prophet_3	1.75	2.04	2.19	2.70	1.26	1.03

Medians of the NSE values

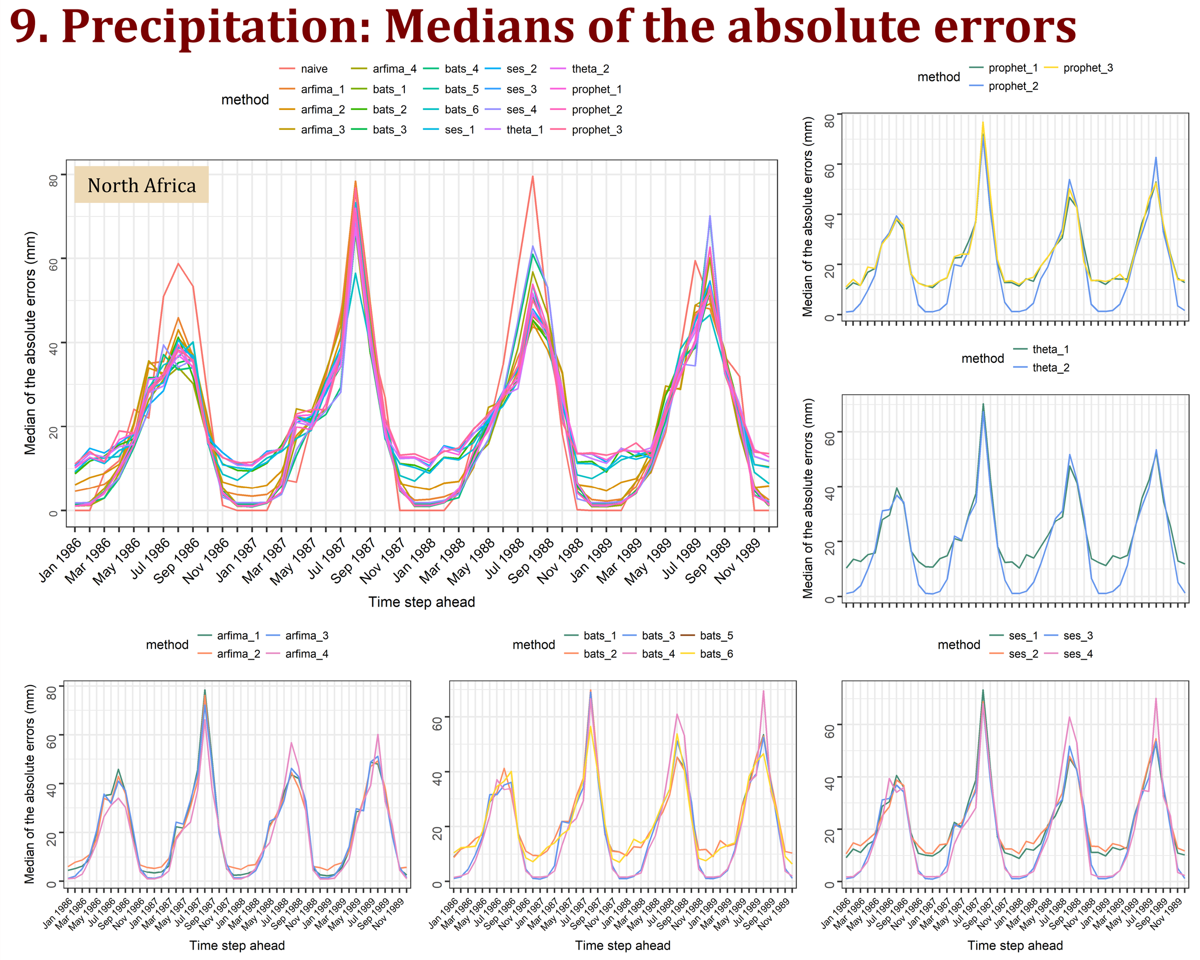
Method	Globe	North America	North Europe	Siberia	Asia (except Siberia)	Oceania
naïve	0.91	0.90	0.79	0.93	0.95	0.90
arfima_1	0.95	0.95	0.91	0.96	0.98	0.93
arfima_2	0.95	0.95	0.91	0.96	0.98	0.93
arfima_3	0.95	0.95	0.91	0.96	0.98	0.93
arfima_4	0.95	0.95	0.91	0.96	0.98	0.92
bats_1	0.95	0.95	0.91	0.96	0.98	0.93
bats_2	0.95	0.95	0.91	0.96	0.98	0.93
bats_3	0.95	0.95	0.91	0.96	0.98	0.93
bats_4	0.95	0.95	0.91	0.96	0.98	0.93
bats_5	0.94	0.93	0.89	0.95	0.97	0.93
bats_6	0.94	0.93	0.89	0.95	0.97	0.93
ses_1	0.95	0.95	0.90	0.96	0.98	0.93
ses_2	0.95	0.95	0.90	0.96	0.98	0.93
ses_3	0.95	0.95	0.90	0.96	0.98	0.93
ses_4	0.95	0.95	0.90	0.96	0.98	0.93
theta_1	0.95	0.95	0.90	0.96	0.98	0.93
theta_2	0.95	0.95	0.90	0.96	0.98	0.93
prophet_1	0.95	0.95	0.91	0.96	0.97	0.93
prophet_2	0.95	0.95	0.91	0.96	0.97	0.93
prophet_3	0.95	0.95	0.90	0.96	0.97	0.93



8. Precipitation: Medians of the absolute errors



9. Precipitation: Medians of the absolute errors



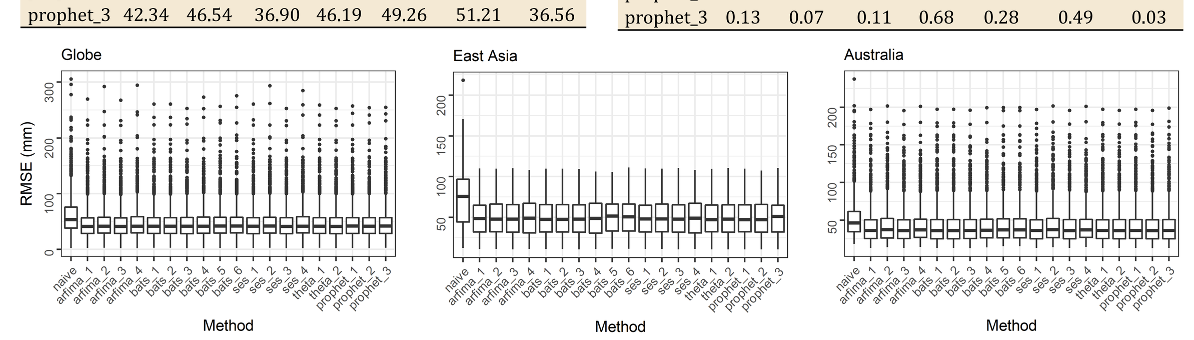
10. Precipitation: RMSE and Nash-Sutcliffe values

Medians of the RMSE values (mm)

Method	Globe	North America	North Europe	North Africa	South Africa	East Asia	Australia
naïve	53.74	63.20	47.89	59.91	58.84	75.61	46.38
arfima_1	41.75	45.16	36.65	46.18	48.34	48.57	36.51
arfima_2	42.07	45.29	37.26	45.27	48.81	47.98	37.49
arfima_3	41.67	45.61	36.65	45.36	48.20	47.79	36.25
arfima_4	42.01	45.34	37.09	46.19	49.60	49.07	37.26
bats_1	41.88	45.78	36.59	46.02	48.85	47.56	36.21
bats_2	41.90	45.62	36.59	45.75	48.85	47.56	36.21
bats_3	41.98	46.15	36.54	45.15	48.56	48.01	36.51
bats_4	42.06	45.28	36.69	47.69	50.04	48.85	36.75
bats_5	42.39	45.80	36.78	47.50	49.99	51.71	37.28
bats_6	42.34	45.56	37.52	47.50	49.99	50.77	37.13
ses_1	41.88	45.54	36.60	45.77	48.83	48.07	36.35
ses_2	42.23	45.49	36.78	46.16	49.17	48.03	37.45
ses_3	41.79	45.90	36.30	45.17	48.40	47.87	36.16
ses_4	42.13	45.39	36.95	48.93	50.29	49.12	37.26
theta_1	42.08	46.24	36.87	46.09	48.95	47.22	36.50
theta_2	41.79	45.90	36.30	45.17	48.40	47.87	36.16
prophet_1	42.16	46.22	37.06	46.03	48.70	47.18	36.56
prophet_2	41.85	46.72	36.84	46.26	48.89	47.08	36.31
prophet_3	42.34	46.54	36.90	46.19	49.26	51.21	36.56

Medians of the NSE values

Method	Globe	North America	North Europe	North Africa	South Africa	East Asia	Australia
naïve	−0.38	−0.55	−0.45	0.54	−0.04	0.05	−0.44
arfima_1	0.15	0.10	0.12	0.69	0.29	0.49	0.04
arfima_2	0.11	0.09	0.08	0.69	0.25	0.50	−0.04
arfima_3	0.14	0.09	0.12	0.70	0.30	0.49	0.05
arfima_4	0.08	0.08	0.05	0.71	0.18	0.51	−0.04
bats_1	0.14	0.08	0.13	0.70	0.29	0.49	0.04
bats_2	0.14	0.08	0.13	0.70	0.29	0.49	0.04
bats_3	0.14	0.08	0.13	0.71	0.29	0.50	0.04
bats_4	0.11	0.10	0.11	0.69	0.18	0.52	−0.01
bats_5	0.11	0.08	0.10	0.70	0.25	0.45	0.01
bats_6	0.10	0.07	0.05	0.70	0.25	0.44	0.01
ses_1	0.14	0.09	0.12	0.70	0.30	0.49	0.04
ses_2	0.10	0.09	0.08	0.69	0.22	0.50	−0.04
ses_3	0.14	0.08	0.12	0.71	0.29	0.49	0.05
ses_4	0.09	0.09	0.05	0.68	0.16	0.51	−0.04
theta_1	0.14	0.08	0.13	0.69	0.29	0.49	0.04
theta_2	0.14	0.08	0.12	0.71	0.29	0.49	0.05
prophet_1	0.14	0.07	0.13	0.69	0.28	0.50	0.04
prophet_2	0.14	0.07	0.13	0.70	0.29	0.50	0.04
prophet_3	0.13	0.07	0.11	0.68	0.28	0.49	0.03



11. Discussion and conclusions

- The present study is available in Papacharalampous et al. (2018b).
- We suggest its reading alongside with its companion studies, i.e. Tyralis and Koutsoyiannis (2014), Papacharalampous et al. (2017a, b, c; 2018a, c) and Tyralis and Papacharalampous (2017).
- The results indicate that all the examined models (apart from the naïve one) are accurate enough to be used in long-term forecasting applications.
- Even the SES and Theta models, which exhibit a rather moderate performance in terms of RMSE and NSE in the experiments of Papacharalampous et al. (2017a), here are found to be equally competitive with the ARFIMA and BATS models, which are the most accurate in terms of RMSE and NSE in the above-mentioned study.
- This may be explained by the fact that the experiments of Papacharalampous et al. (2017a) use non-seasonal simulated processes, with different predictability than the monthly temperature and precipitation processes. Seasonality can be assumed to be the deterministic term of a process and its proper handling leads to a significant improvement of the forecasts.
- Regarding the investigation of the present study on how different choices of handling seasonality and non-normality affect the performance of the models, the results do not suggest any specific combination of choices for the external handling of seasonality and non-normality as best.
- Nevertheless, the handling of seasonality through the BATS and Prophet models (the only models that offer this possibility amongst the used ones) mostly leads to less accurate forecasts than the external handling, especially for the former model.
- Admittedly, the quantitative information provided by the present study is also important, since it directly expresses the predictability of monthly temperature and precipitation. Excluding the naïve method, the respective RMSE values range between 1.01 K and 2.84 K for temperature, and 36.16 mm and 51.71 mm for precipitation.
- In more detail, for the total of the temperature time series the use of an ARFIMA, BATS, SES, Theta or Prophet model, instead of the naïve method, leads to about 19-29% more accurate forecasts in terms of RMSE, or even in about 30-32% more accurate forecasts specifically for the temperature time series observed in North Europe.
- For the total of the precipitation time series the use of all these automatic methods leads to about 21-22% better forecasts than the use of the naïve method, while for the geographical regions of North America, North Europe and East Asia these percentages are 26-29%, 22-24% and 32-38% respectively.
- We think that the level of the forecasting accuracy can barely be improved using other methods as the experiments of Papacharalampous et al. (2017a) suggest.

References

[1] Brownrigg, R., Minka, T.P., and Deckmyn, A., 2017. maps: Draw Geographical Maps. R package version 3.2.0. https://CRAN.R-project.org/package=maps

[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. https://CRAN.R-project.org/package=fracdiff

[3] 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

[4] 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

[5] Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S., and Rennie, J. 2011. An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *Journal of Geophysical Research: Atmospheres*, 116 (D1912). doi: 10.1029/2011JD016187

[6] Mayer, J., Khairy, K., and Howard, J., 2010. Drawing an elephant with four complex parameters. *American Journal of Physics*, 78 (6). doi:10.1119/1.3254017

[7] Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D., 2017a. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Preprints*, 2017100133. doi:10.20944/preprints201710.0133.v1

[8] Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D., 2017b. Error evolution in multi-step ahead streamflow forecasting for the operation of hydropower reservoirs. *Preprints*. doi:10.20944/preprints201710.0129.v1

[9] Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D., 2017c. Forecasting of geophysical processes using stochastic and machine learning algorithms. *European Water*, 59, 161–168

[10] Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D., 2018a. One-step ahead forecasting of geophysical processes within a purely statistical framework. *Geoscience Letters*. doi:10.1186/s40562-018-0111-1

[11] Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D., 2018b. Predictability of monthly temperature and precipitation using automatic time series forecasting methods. *Acta Geophysica*. doi:10.1007/s11600-018-0120-7

[12] Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D., 2018c. Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: A multiple-case study from Greece. In review

[13] Peterson, T.C., and Vose, R.S. 1997. An Overview of the Global Historical Climatology Network Temperature Database. Bulletin of the American Meteorological Society, 78, 2837–2849. doi: 10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2

[14] R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[15] Taylor, S.J., and Letham, B., 2017a. Forecasting at scale. *The American Statistician*. doi:10.1080/00031305.2017.1380080

[16] Taylor, S.J., and Letham, B., 2017b. prophet: Automatic Forecasting Procedure. R package version 0.2. https://CRAN.R-project.org/package=prophet

[17] Tyralis, H., 2016. HKprocess: Hurst-Kolmogorov Process. R package version 0.0-2. https://CRAN.R-project.org/package=HKprocess

[18] 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

[19] 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

[20] Tyralis, H., and Papacharalampous, G., 2017. Variable selection in time series forecasting using random forests. *Algorithms*, 10 (4), 114. doi:10.3390/a10040114

[21] Warnes, G.R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., MacQueen, D., Magnusson, A., Rogers, J., and others, 2017. gdata: Various R Programming Tools for Data Manipulation. R package version 2.18.0. https://CRAN.R-project.org/package=gdata

[22] Wickham, H., 2016. ggplot2. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

[23] Wickham, H., and Chang, W., 2017. devtools: Tools to Make Developing R Packages Easier. R package version 1.13.4. https://CRAN.R-project.org/package=devtools

[24] Wickham, H., Hester, J., Francois, R., Jylänki, J., and Jørgensen, M., 2017. readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr

[25] Zeileis, A., and Grothendieck, G., 2005. zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, 14 (6), 1–27. doi:10.18637/jss.v014.i06

[26] Xie, Y., 2014. knitr: A Comprehensive Tool for Reproducible Research in R. In: V. Stodden, F. Leisch, R.D. Peng, eds. Implementing Reproducible Computational Research. Chapman and Hall/CRC

[27] Xie, Y., 2015. Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC

[28] Xie, Y., 2017. knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.17. Available from: https://CRAN.R-project.org/package=knitr