Predictability of monthly temperature and precipitation using automatic time series forecasting methods

Georgia Papacharalampous*, Hristos Tyralis, and Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Polytechniou 5, 157 80 Zografou, Greece

* Corresponding author, papacharalampous.georgia@gmail.com

Abstract: We investigate the predictability of monthly temperature and precipitation by applying automatic univariate time series forecasting methods to a sample of 985 40-year long monthly temperature and 1552 40-year long monthly precipitation time series. The methods include a naïve one based on the monthly values of the last year, as well as the random walk (with drift), AutoRegressive Fractionally Integrated Moving Average (ARFIMA), exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (BATS), simple exponential smoothing, Theta and Prophet 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 before, while the use of random walk, BATS, simple exponential smoothing and Theta is rare in hydrology. The methods are tested in performing multi-step ahead forecasts for the last 48 months of the data. We further investigate how different choices of handling the seasonality and non-normality affect the performance of the models. The results indicate that (a) all the examined methods apart from the naïve and random walk ones are accurate enough to be used in long-term applications, (b) monthly temperature and precipitation can be forecasted to a level of accuracy which can barely be improved using other methods, (c) the externally applied classical seasonal decomposition results mostly in better forecasts compared to the automatic seasonal decomposition used by the BATS and Prophet methods and (d) Prophet is competitive, especially when it is combined with externally applied classical seasonal decomposition.

Key words: ARFIMA; multi-step ahead forecasting; precipitation forecasting; Prophet; temperature forecasting; time series forecasting

1. Introduction

The role of univariate time series forecasting methods for hydrometeorological and climate forecasting has been emphasized by forecasting experts (Armstrong and Fildes 2006; Green and Armstrong 2007; Green et al. 2009). Relevant reviews linking geosciences with the forecasting scientific field are available in the literature (e.g. Bărbulescu 2016; Sivakumar 2017), while a review of studies applying time series point forecasting methods in hydrology is presented in Papacharalampous et al. (2017a). Moreover, time series analysis is an essential tool for better forecasts; consequently, analysis and forecasting are usually presented simultaneously in specialized textbooks (e.g. Hyndman and Athanasopoulos 2017; Wei 2006).

While time series forecasting is of interest in hydrology, the principles of forecasting are not always conscientiously applied by hydrological scientists, as revealed by the experiments of Papacharalampous et al. (2017a). Furthermore and despite the growing literature specialized in time series prediction and examining subtleties related to the development, application and assessment of methods (see, for example, the review of De Gooijer and Hyndman (2006)), the gained technical know-how has not been fully exploited by geoscientists. This is also suggested by the low number of geoscientific papers citing literature from relevant leading journals (e.g. the International Journal of Forecasting). Armstrong and Fildes (2006) have recognized the need for promoting forecasting in geosciences and proposed the publication of a special issue on applications of traditional forecasting methodologies to climate sciences.

Admittedly, the recent trend in geosciences is focusing on the development of soft computing methods for time series forecasting. These methods can be equally accurate, yet more computationally intensive, compared to the classical forecasting approaches, as presented in Papacharalampous (2017a). In addition to the right above distinction, the various forecasting alternatives can be classified as automatic and non-automatic. The non-automatic or subjective approach to the problem of time series forecasting requires the prior conduct of an exploratory data analysis for each specific individual case to be predicted, as well as human intervention during the forecasting process (Chatfield 1988). Therefore, its implementation can be significantly limited by scale-dependent factors. Taylor and Letham (2017a) identify three types of scale in forecasting related to the number of people making forecasts (and their varying backgrounds), the diversity of the characteristics of each forecasting problem and the number of forecasts needed. Automatic time series forecasting is essential, for example, when a large number of time series is required to be forecasted (Hyndman and Khandakar 2008). Consequently, specialized software for automatic time series forecasting has been developed (Hyndman and Khandakar 2008; Hyndman et al. 2017; Taylor and Letham 2017a,b).

The 'forecast' R package includes mostly methods based on exponential smoothing (Hyndman et al. 2008) or AutoRegressive Integrated Moving Average (ARIMA) (and related) stochastic processes. The ARIMA processes and relevant applications introduced by Box and Jenkins (1968) have been consistently used in hydrology from earlier years (e.g. Carlson et al. 1970) until more recent times (e.g. Montanari et al. 1997, 2000), while exponential smoothing has been used less frequently (e.g. Papacharalampous et al. 2017a,b). Recent methods for automatic time series forecasting that have been used rarely in hydrology include the Theta method (Assimakopoulos and Nikolopoulos 2000, Hyndman and Billah 2003) and the BATS (acronym for Box–Cox transformation, ARMA errors, Trend, and Seasonal components) method (De Livera et al. 2011). A more recently introduced forecasting model is Prophet (Taylor and Letham 2017a), inspired by the nature of time series forecasted at Facebook and available in the 'prophet' R package.

Time series forecasting methodologies can be also classified into two groups according to the forecast horizon, i.e. one- and multi- step ahead forecasting. The latter is more difficult compared to the former. Still, it has been frequently used in hydrology (e.g. Papacharalampous et al. 2017a,b,c; Singh et al. 2011; Tyralis and Koutsoyiannis 2014; Valipour et al. 2013). Relevant reviews on multi-step ahead forecasting methodologies can be found in Chevillon (2007) and Taieb et al. (2012). Multi-step ahead forecasting has been examined theoretically (e.g. De Gooijer and Klein 1992; De Gooijer and Kumar 1992; Franses and Legerstee 2010; Pemberton 1987; Stoica and Nehorai 1989; Taieb and Atiya 2016; Wei 2006, pp. 88-107) and empirically (Papacharalampous et al. 2017a,c). The theoretical examination is not possible for all methods, while the performance of the latter can vary considerably in real case studies. An alternative framework for assessing the performance of forecasting methods is through their application to large datasets. Thus, competitions are organized, in which methods are improved and compared (Makridakis et al. 1987; Makridakis and Hibon 2000) and later published to advance science (e.g. Andrawis et al. 2011).

The culture of evaluating the forecasting performance of methods on large datasets is not a usual practice in hydrology, as noted in Papacharalampous et al. (2017a). The latter study compares stochastic and machine learning methods in multi-step ahead forecasting using a large dataset of simulated and streamflow data. A similar approach has been used in a study of Papacharalampous, Tyralis and Koutsoyiannis (One-step ahead forecasting of geophysical processes within a purely statistical framework: manuscript in review, 2018) in which annual values of mean temperature and precipitation, as well as simulated processes, are used for performing one-step ahead forecasting experiments. These studies are among the first, in which forecasting methods popular in the time series forecasting field (e.g. BATS, Theta and simple exponential smoothing) are used in hydrology.

Here we apply automatic univariate time series forecasting methods to a large sample of 985 40-year long monthly temperature and 1552 40-year long monthly precipitation time series. This sample is the largest used in hydrology for assessing the performance of forecasting methods. We implement a naïve forecasting method based on the monthly values of the last year, as well as the random walk (with drift), AutoRegressive Fractionally Integrated Moving Average (ARFIMA), BATS, simple exponential smoothing, Theta and Prophet forecasting methods. The methods' multi-step ahead forecasting performance is assessed using the last 48 months of the data. The aims of the present study are to:

- Assess the performance of the BATS, simple exponential smoothing and Theta methods when forecasting monthly geophysical time series.
- Compute the minimum forecasting error, which directly expresses the predictability of monthly temperature and precipitation.
- Assess whether using Box-Cox transformations and/or classical seasonal decomposition (additive or multiplicative) results in a better performance of the forecasting methods. For the latter see the relevant discussion in Mills (2011, pp.375–395).
- Assess the performance of the Prophet method.

2. Methodological framework

2.1 Data

We use monthly temperature and precipitation instrumental data from the stations presented in Fig. 1, while the primary datasets are documented in Lawrimore et al. (2011) and Peterson and Vose (1997) respectively. The data span from 1950 to 1989 (i.e. 480 months). Since we need a large dataset without missing values, we do not use more recent data. Indeed the number of stations without missing data decreases rapidly after 1990. Furthermore, 480 months are sufficient for a reliable inference regarding the performance of the forecasting methods. The stations are located in regions with different climates; therefore, our testing could be affected by the dispersion of the stations. To mitigate this effect, we group the stations according to their locations, as presented in Table 1 for the temperature stations and Table 2 for the precipitation ones. We note that 130 temperature and 149 precipitation stations are left out of the formed groups, due to their remote locations compared to the rest of the stations.

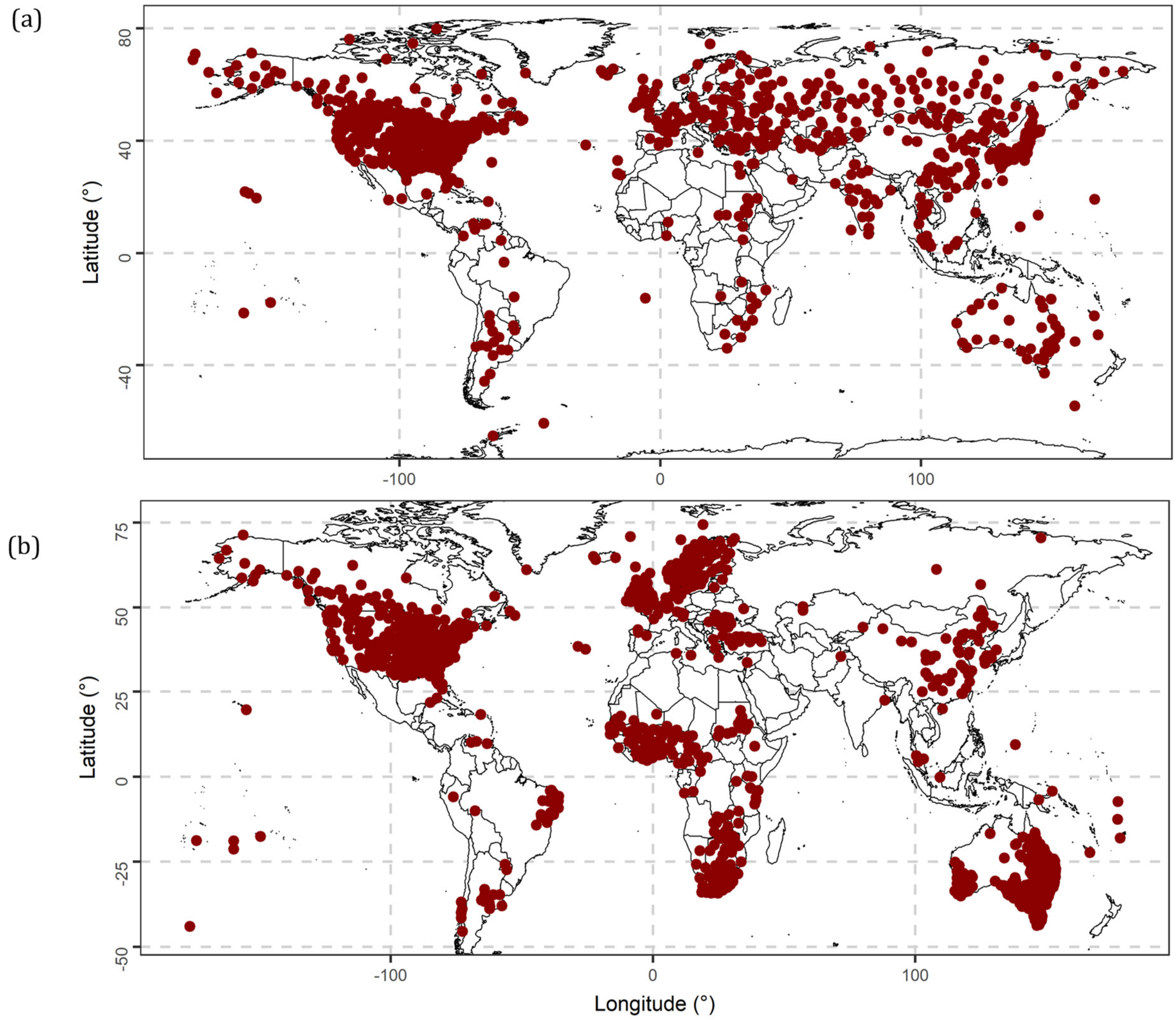

Fig. 1. Maps of the (a) temperature and (b) precipitation stations; their sources are Lawrimore et al. (2011) and Peterson and Vose (1997) respectively.

Table 1. Groups of temperature stations with the respective number of stations per group and regions' geographical boundaries.

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]

Table 2. Groups of precipitation stations with the respective number of stations per group and regions' geographical boundaries.

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]

The Hurst-Kolmogorov behaviour (or long-range dependence) is an inherent property of geophysical processes (see, for example, Tyralis and Koutsoyiannis 2011 and the references therein). Its magnitude is measured by the value of the Hurst parameter $H \in (0, 1)$. The long-range dependence is strong when H is high, while $H = 0.5$ corresponds to uncorrelated random variables. In Fig. 2 we present the H parameter estimates of the time series along with their estimated means (denoted as μ) and standard deviations (denoted as σ). To estimate the parameters, we first decompose the time series using a classical additive model, as described in Hyndman and Athanasopoulos (2017, Chapter 6.3), and subsequently we use the maximum likelihood estimator presented in Tyralis and Koutsoyiannis (2011). The magnitude of the estimated long-range dependence is significant in the temperature time series, while long-range dependence is also observed in the precipitation time series. This is important in the forecasting procedure, since we implement a method which can model the Hurst-Kolmogorov behaviour and take advantage of this prior knowledge (see Section 2.3).

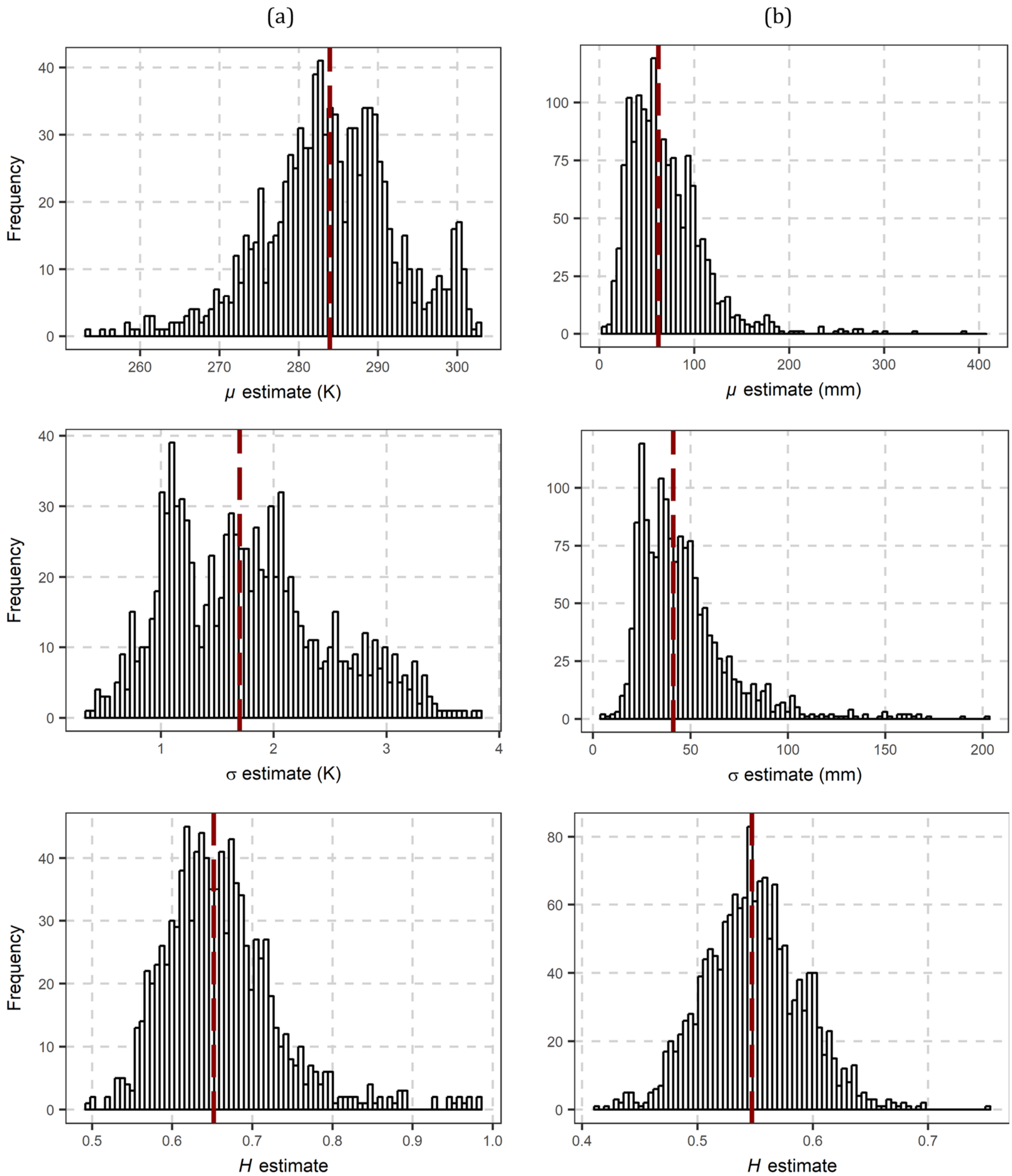

Fig. 2. Mean (μ), standard deviation (σ) and Hurst parameter (H) estimates for the total of the deseasonalized (a) temperature and (b) precipitation time series. The vertical red dashed line denotes the median value of the estimates.

2.2 Definition of the forecasting problem

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 time series forecasting is univariate, i.e. each time series is forecasted based on its past values. The observed values of the period 1986-1989 are used for testing the performance of the forecasting methods (and are not used for the fitting of the models). Let also $x_1, x_2, \ldots, x_n$ represent the observations (in the period 1986-1989) and $f_1, f_2, \ldots, f_n$ represent their forecasts.

In Fig. 3 we present the medians of the observed temperature values to be forecasted for all groups of Table 1. The seasonal patterns are obvious in each region, while the minima and maxima clearly depend on the hemisphere. We present a similar illustration for the precipitation time series in Fig. 4. The seasonal patterns are again apparent, while zero precipitation appears in Africa regions.

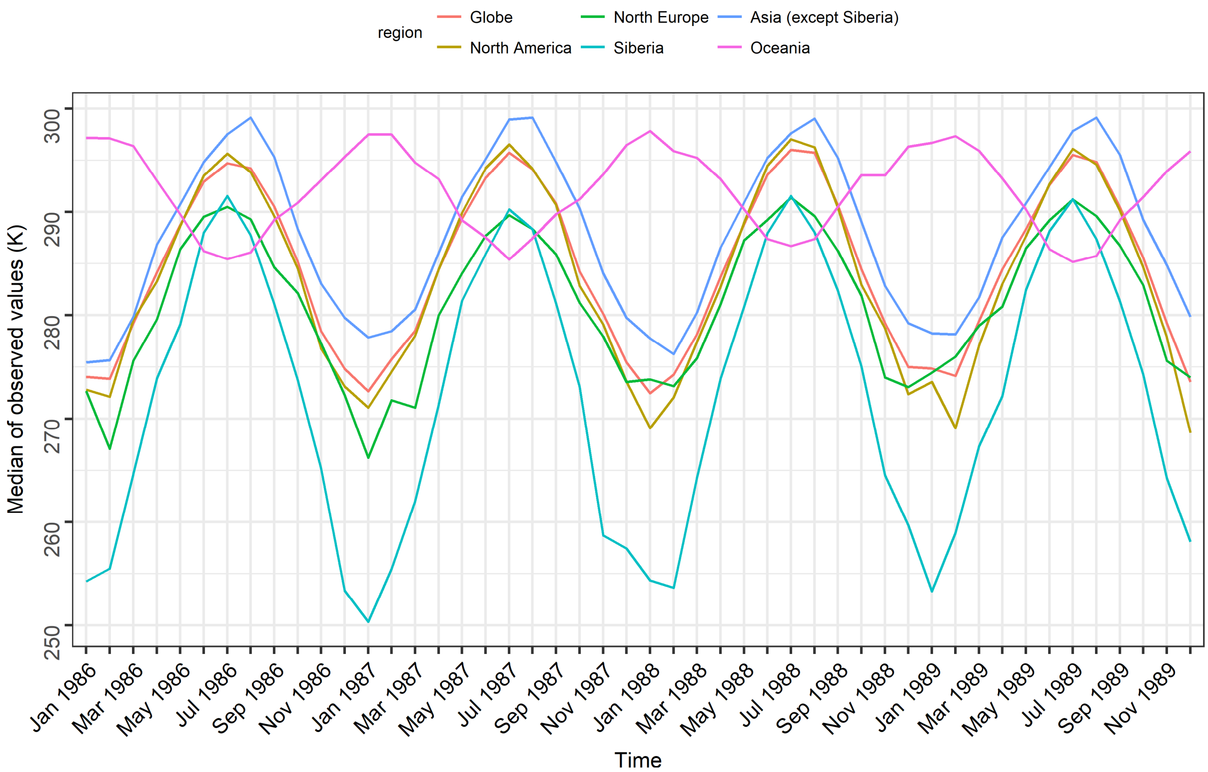

Fig. 3. Medians of the observed temperature values to be forecasted per group presented in Table 1.

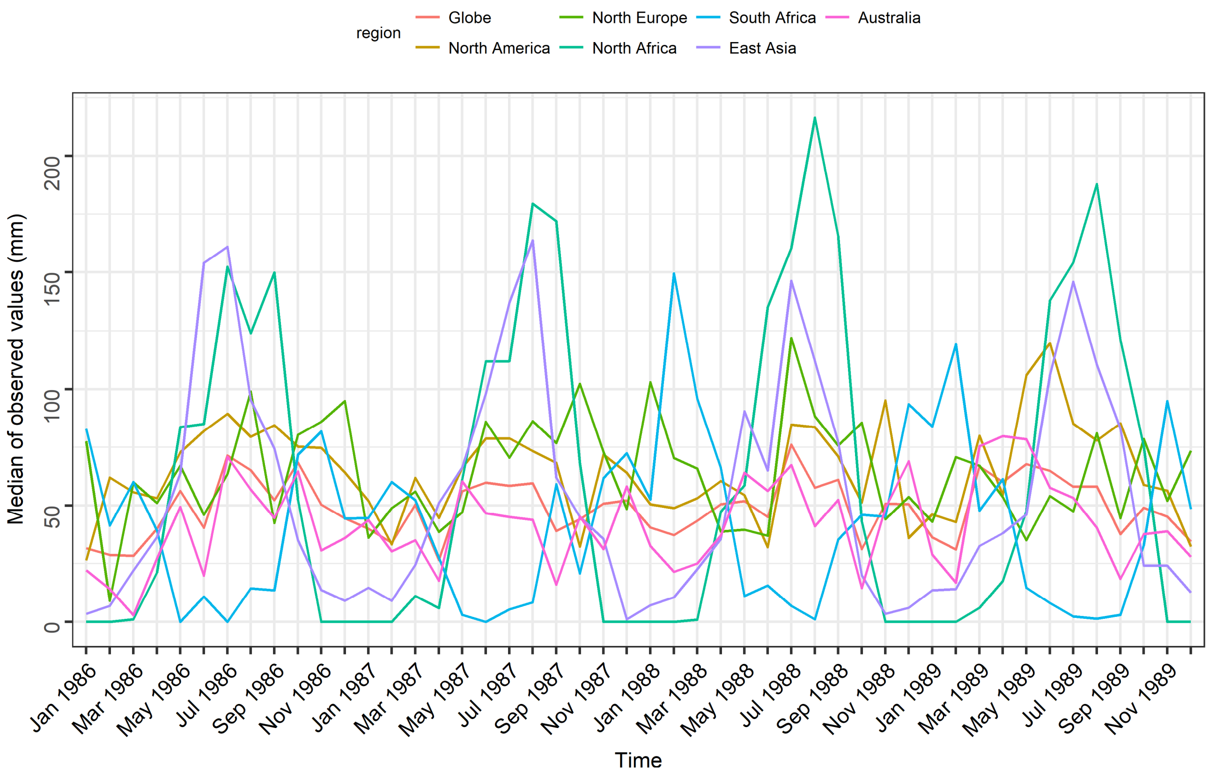

Fig. 4. Medians of the observed precipitation values to be forecasted per group presented in Table 2.

2.3 Forecasting methods

We use the seven forecasting methods described in Table 3 with their documentation and implementation in R programming language (R Core Team 2017). In particular, most of the used algorithms are implemented through the 'forecast' R package (Hyndman et al. 2017), while the Prophet method is implemented through the 'prophet' R package (Taylor and Letham 2017b). Subsequently, we present these seven methods, omitting the mathematical parts, since they are well documented in the literature, while highlighting important aspects from a practical point of view.

Table 3. Forecasting methods implemented in the present study.

s/n	Name	R Function	Category	Reference	R package
1	naïve		simple		
2	random walk	rwf	simple	Hyndman and Athanasopoulos (2017)	Hyndman et al. (2017)
3	ARFIMA	arfima	ARFIMA		Hyndman et al. (2017)
4	BATS	bats	state space	De Livera et al. (2011)	Hyndman et al. (2017)
5	simple exponential smoothing	ses	exponential smoothing	Hyndman et al. (2008)	Hyndman et al. (2017)
6	Theta	thetaf	exponential smoothing	Assimakopoulos and Nikolopoulos (2000)	Hyndman et al. (2017)
7	Prophet	predict.prophet	curve-fitting	Taylor and Letham (2017a)	Taylor and Letham (2017b)

The naïve method is the simplest available forecasting method. In our case, the forecasted value of the naïve method for each month of the testing period is equal to the observed value of the respective month of the year 1985, i.e. the last year of the fitting period. The random walk method, defined in Hyndman et al. (2017), fits a random walk model with drift to the fitting set and then uses the fitted model for forecasting. Both simple methods are based on modelling the data using discrete-time martingales. The difference is that in the latter case a drift is added to the model.

The ARFIMA models are widely used in hydrology. The reader is referred to Wei (2006, pp. 6-65, 489-494) for their detailed definitions. Let $d \in (-0.5, 0.5)$. The stochastic process $\{\underline{x}_t\}$ is an ARFIMA(p, d, q), if the following equation holds:

$$\varphi_p(\mathrm{B})(1 - \mathrm{B})^d \underline{x}_t = \theta_q(\mathrm{B})\underline{a}_t \tag{1}$$

where

$$\varphi_p(\mathrm{B}) := 1 - \varphi_1\mathrm{B} - \ldots - \varphi_p\mathrm{B}^p \tag{2}$$

$$\theta_q(\mathrm{B}) := 1 + \theta_1\mathrm{B} + \ldots + \theta_q\mathrm{B}^q \tag{3}$$

$$\mathrm{B}^j\underline{x}_t = \underline{x}_{t-j} \tag{4}$$

and $\{\underline{a}_t\}$ is a white noise process. The parameter d of the model is a measure of the estimated long-range dependence of the time series (see also Section 2.1). The parameters p and q are the orders of the autoregressive and moving average components of the ARFIMA model. The ARFIMA forecasting method fits an ARFIMA(p, d, q) model to the fitting set. The fitting procedure is explained in Hyndman et al. (2017). The parameters p and q are estimated using the Hyndman and Khandakar (2008) algorithm and the parameter d is estimated using the Haslett and Raftery (1989) algorithm. The method combines functions from the 'fracdiff' (Fraley et al. 2012) and 'forecast' (Hyndman et al. 2017) R packages. The fitted model is used for forecasting future values.

The BATS method, introduced by De Livera et al. (2011), combines an innovations state-space modelling framework, which incorporates Box–Cox transformations, Fourier representations with time varying coefficients, and ARMA error correction, while it is suitable for forecasting time series exhibiting complex seasonal patterns. The method is implemented through the 'forecast' R package (Hyndman et al. 2017).

The simple exponential smoothing method, introduced by Brown (1959) and described in Hyndman et al. (2008, p. 13), computes the forecast of the next period based on the forecast of the previous period, the latter adjusted using its error according to the following equation:

$$f_{t+1} = f_t + a(x_t - f_t),\ a \in (0, 1) \tag{5}$$

The estimation of the parameters of the simple exponential smoothing method is performed using procedures of the 'forecast' R package (Hyndman et al. 2017).

The Theta method, introduced by Assimakopoulos and Nikolopoulos (2000), is a special case of simple exponential smoothing with a drift parameter. The drift parameter is half the slope of the linear trend fitted to the data, as shown in Hyndman and Billah (2003). The Theta method performed well in the M-3 forecasting competition (Makridakis and Hibon 2000).

The Prophet method, introduced by Taylor and Letham (2017a) and implemented through the 'prophet' R package (Taylor and Letham 2017b), considers time series forecasting as a curve-fitting exercise, while it does not explicitly consider the temporal dependence of the time series. It simultaneously uses a decomposable time series model. The Prophet method is inspired by the nature of the time series forecasted at Facebook, which are characterized by trend, multiple seasonality and holidays (an example of similar time series in the water science is the water demand). Furthermore, this method is designed to *"forecast at scale"* and to fit to the data very fast.

2.4 Seasonality, trends and non-normality

Due to the seasonality and the possible non-normality of the data of the present study, some forecasting methods cannot be directly applied. Some other methods can automatically transform the data, without external handling.

When seasonality is observed, two possible transformations of the data before applying the forecasting methods of Table 3 are the classical additive and multiplicative seasonal decomposition (Hyndman and Athanasopoulos 2017, Chapter 6.3). The additive model is suitable when the seasonal fluctuations do not depend on the level of the time series. The multiplicative model is suitable for modelling seasonal fluctuations which are proportional to the level of the time series (Hyndman and Athanasopoulos 2017, Chapter 6.1). In the particular case of the multiplicative seasonal decomposition of the precipitation time series, we add 10 mm to each value, because the zero observed values result in zeros during the seasonal decomposition. Obviously, the inverse transform involves the subtraction of 10 mm for the forecasted values. This approach may affect the decomposition pattern; however, it is the most practical in this case.

Some methods are applied under the assumption of the normality of the data. Since the time series data are usually non-normal, an appropriate transformation is needed. The most popular transformation is the Box-Cox transformation, introduced by Box and Cox (1964) and given by the following equation for $x > 0$:

$$f_\lambda(x) = \begin{cases} (x^\lambda - 1)/\lambda \text{ if } \lambda \neq 0 \\ \ln(x) \text{ if } \lambda = 0 \end{cases} \tag{6}$$

where x denotes the variable to be transformed and λ is a parameter which is estimated here using the method of Guerrero (1993), as implemented in Hyndman et al. (2017).

In Table 4 we present the variations of the methods described in Table 3. The variations depend on whether a transformation is used and on which is this specific transformation. In Table 5 we present the transformations accounting for seasonality, while in Table 6 we present the transformations accounting for non-normality. Each variation of the method is assigned to a particular combination of the transformations presented in Table 5 and Table 6 respectively. When a transformation is applied, then the forecasts are obtained through the inverse transform. The BATS and the Prophet methods can also handle the seasonal patterns automatically. Therefore, their variations include cases in which the time series are seasonally decomposed manually, as well as cases in which seasonality is considered through an automatic scheme.

Table 4. Variations of the methods of Table 3.

s/n	Abbreviated name	Primal method (see Table 3)	Handling of seasonality (see Table 5)	Handling of non-normality (see Table 6)
1	naïve	1	1	1
2	rw_1	2	2	1
3	rw_2	2	2	2
4	rw_3	2	3	1
5	rw_4	2	3	2
6	arfima_1	3	2	1
7	arfima_2	3	2	2
8	arfima_3	3	3	1
9	arfima_4	3	3	2
10	bats_1	4	2	1
11	bats_2	4	2	2
12	bats_3	4	3	1
13	bats_4	4	3	2
14	bats_5	4	4	1
15	bats_6	4	4	2
16	ses_1	5	2	1
17	ses_2	5	2	2
18	ses_3	5	3	1
19	ses_4	5	3	2
20	theta_1	6	2	3
21	theta_2	6	3	3
22	prophet_1	7	2	3
23	prophet_2	7	3	3
24	prophet_3	7	4	3

Table 5. Choices for the handling of seasonality adopted in the present study.

s/n	Handling
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

Table 6. Choices for the handling of non-normality adopted in the present study.

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

2.5 Metrics

We compute the error and absolute error at each time step of the forecast horizon for each forecasting attempt, denoted with E_i and $(\mathrm{AE})_i$ respectively in the following equations:

$$E_i := f_i - x_i \quad (7)$$

$$(\mathrm{AE})_i := |f_i - x_i| \quad (8)$$

We build the side-by-side boxplots of the E_i values, while for each time step i of the forecast horizon we compute the median of the $(\mathrm{AE})_i$ values and form the respective figures. We further compute the Root Mean Squared Error (RMSE; Hyndman and Athanasopoulos 2017, Chapter 3.4) and the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe 1970) of each multi-step ahead forecast. The RMSE and the NSE are defined by the following equations:

$$\mathrm{RMSE} := \left(\left(\sum_{i=1}^{n}(f_i - x_i)^2\right)/n\right)^{1/2} \tag{9}$$

$$\mathrm{NSE} := 1 - \left(\sum_{i=1}^{n}(f_i - x_i)^2\right)/\left(\sum_{i=1}^{n}\left(x_i - \left(\left(\sum_{i=1}^{n} x_i\right)/n\right)\right)^2\right) \tag{10}$$

3. Results

3.1 Analysis using the temperature time series

Section 3.1 is devoted to the results of the analysis using the temperature time series. In Fig. 5 we present the side-by-side boxplots of the errors at each time step of the forecast horizon, as formed for all the temperature forecasts produced by the naïve, rw_1 and prophet_1 methods. The random walk variations create a similar image to each other (see the one of rw_1), while the same applies to the set of ARFIMA, BATS, simple exponential smoothing, Theta and Prophet variations (see the one of prophet_1). To illustrate this closeness in the forecasting performance of the methods, in Fig. 6 we present the median values of the absolute errors computed for the total of the temperature time series. The random walk variations are the least accurate at almost every step of the examined horizon with a minimum median of absolute errors around 1.3 K and a maximum around 4 K, while naïve is also worse than the rest of the automatic methods with minimum and maximum medians around 0.8 K and 4.2 K respectively. The best median performance is around 0.5 K, which is about 70% smaller than the median of the estimated standard deviations of the deseasonalized time series (see Fig. 2). We further observe that the presented time series tend rather to run in parallel than to intersect each other and, therefore, the good/bad forecasts of the various methods are rather grouped in the horizontal direction. This behaviour may be explained by the fact that the magnitude of the error of the forecast produced by a specific method largely depends on the value to be forecasted, i.e. some forecasting attempts are by nature more difficult than others. As a result, the worst median performance of each and every of the methods is observed for January 1989, a month exhibiting higher temperature than the one expected from seasonality (see Fig. 3), while the second worst for February 1989 for the opposite reason.

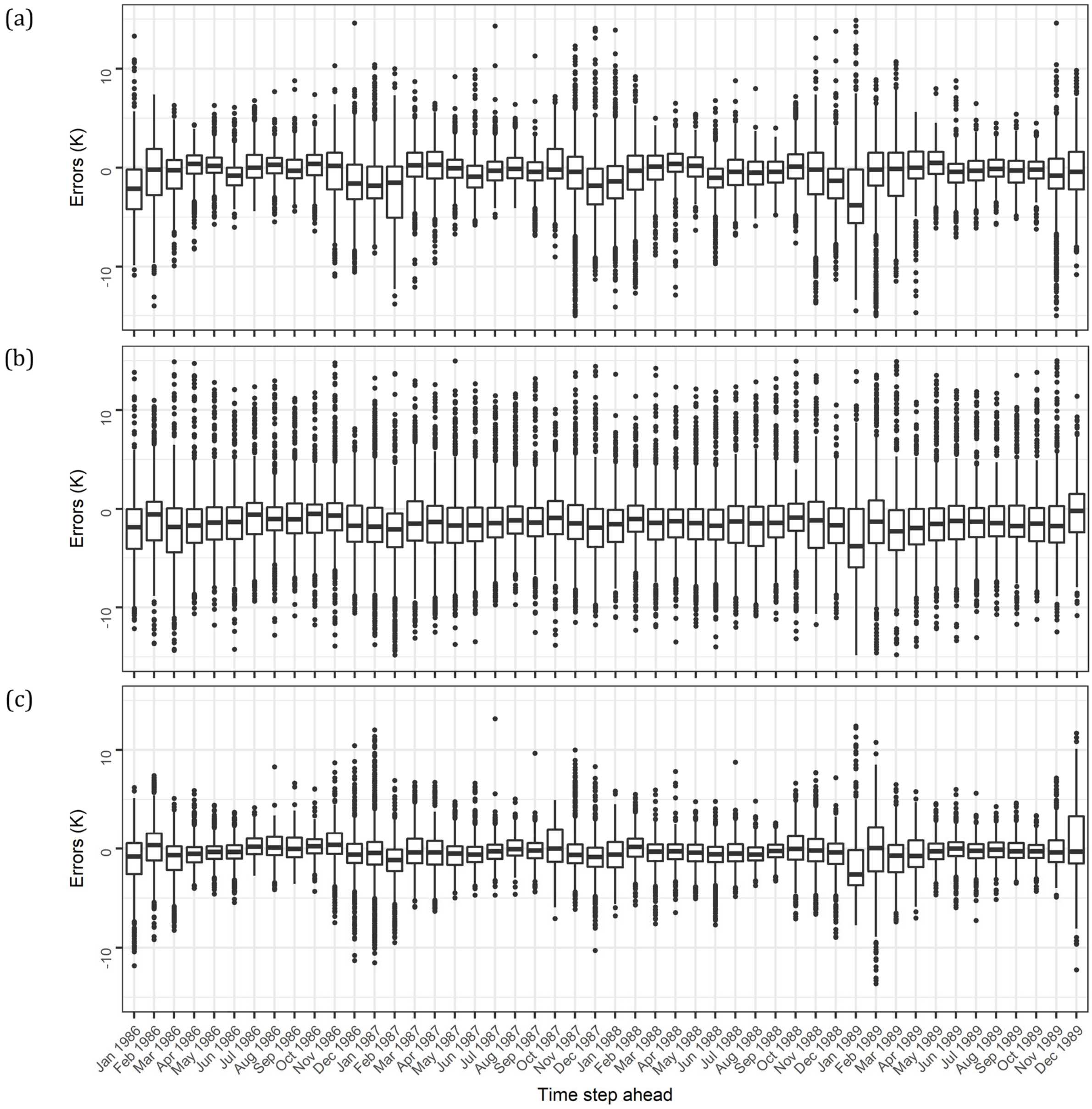

Fig. 5. Errors at each time step of the forecast horizon for the total of temperature time series and the (a) naïve, (b) rw_1 and (c) prophet_1 methods. The outliers with absolute value larger than 15 K are omitted.

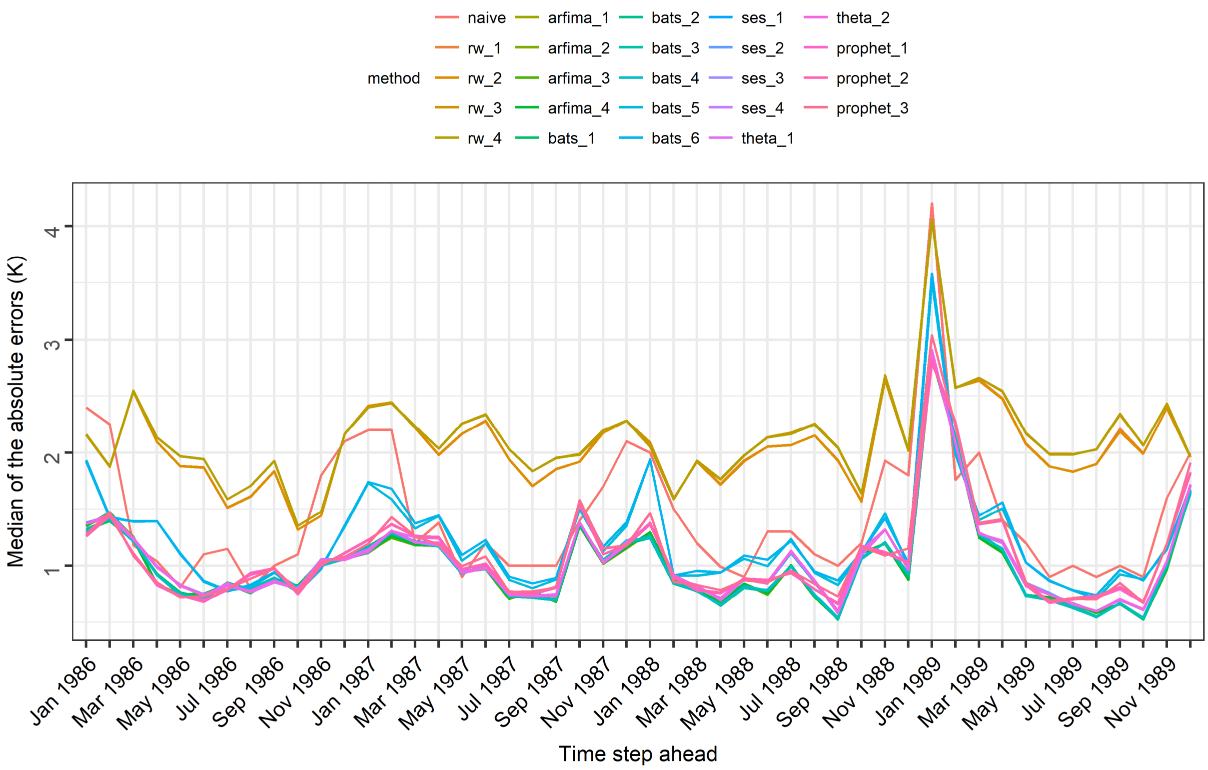

Fig. 6. Medians of the absolute errors at each time step of the forecast horizon for the total of the temperature time series.

In Fig. 7 and 8 we present the median values of the absolute errors computed for each of the groups of stations of Table 1. Since most of the temperature time series are observed in North America, the results corresponding to this geographical region affect the total (presented in Fig. 6) to a significant extend. In fact, Fig. 6 is more similar to Fig. 7(a) than to Fig. 7(b, c) or Fig. 8. However, the medians of the absolute errors are larger in North America than worldwide. In the remaining geographical regions the performance of the random walk methods are closer to the performance of the rest automatic methods, while for the specific cases of North Europe and Siberia the random walk variations are better than naïve as well. The best average performances are measured for Oceania with a minimum median of absolute errors around 0.25 K and a maximum around 2.1 K, while the respective values for Asia (except Siberia) (around 0.30 K and 4 K) are also better than the overall.

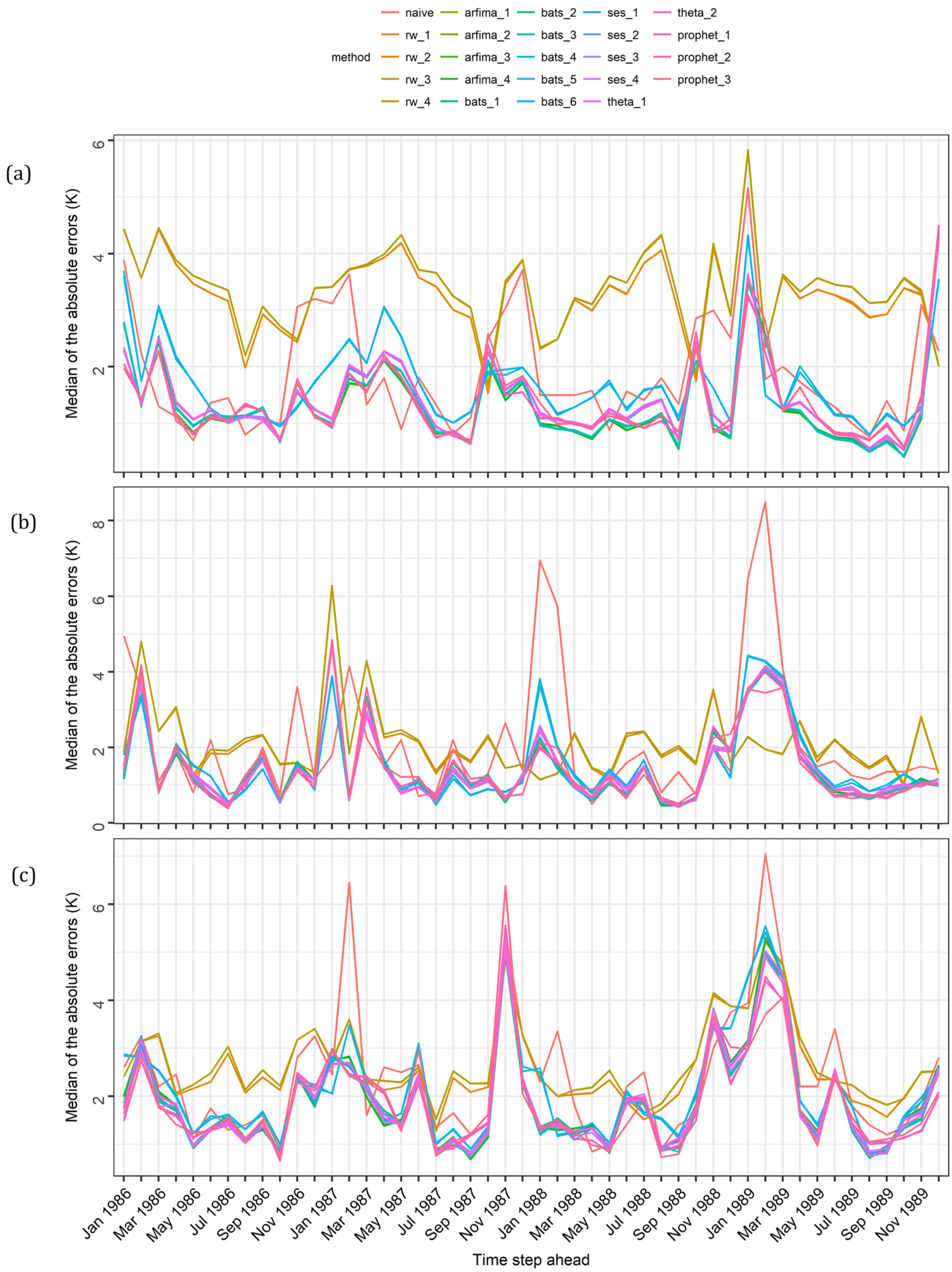

Fig. 7. Medians of the absolute errors at each time step of the forecast horizon for the temperature time series observed in: (a) North America, (b) North Europe and (c) Siberia.

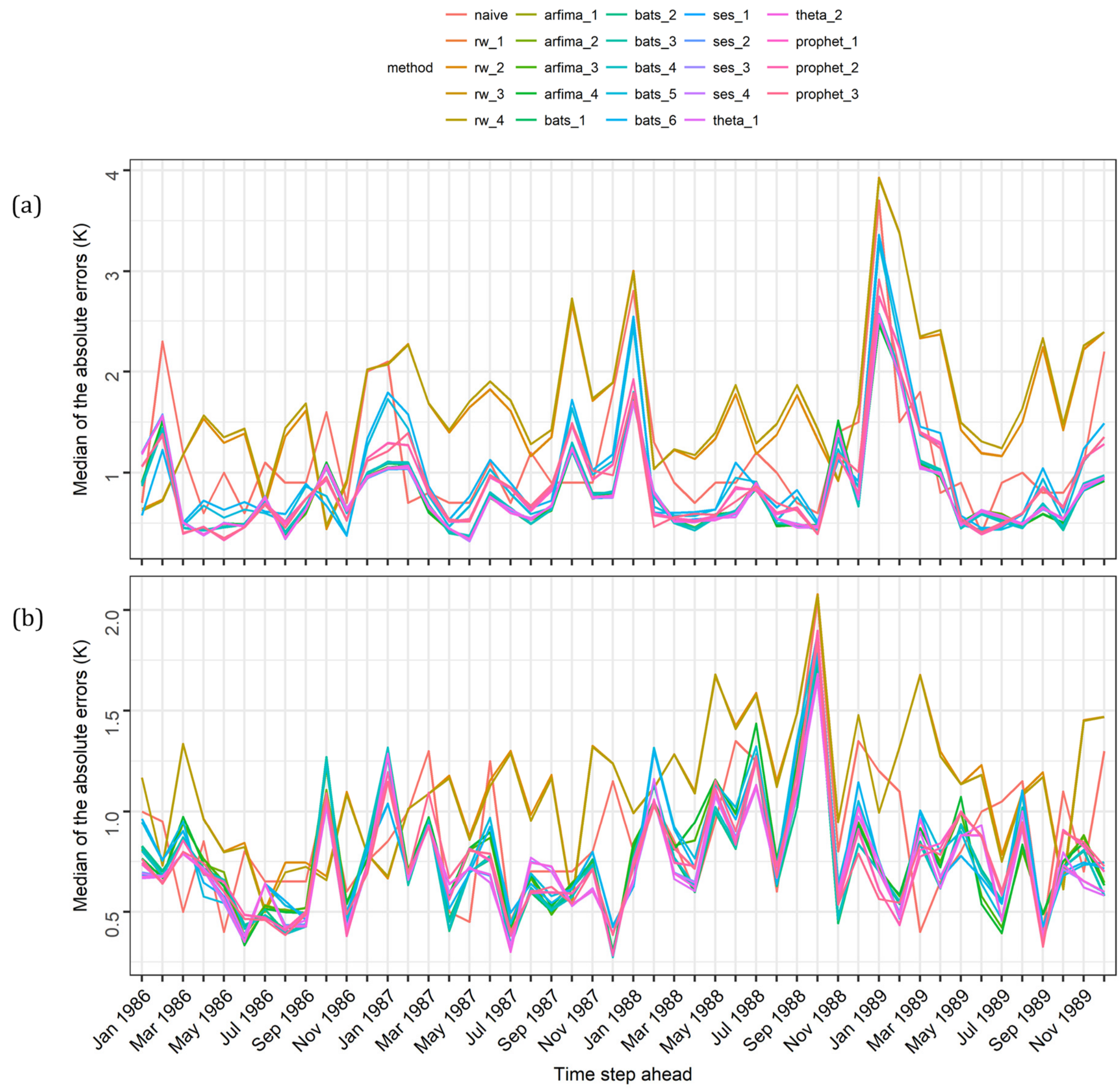

Fig. 8. Medians of the absolute errors at each time step of the forecast horizon for the temperature time series observed in: (a) Asia (except Siberia) and (b) Oceania.

In Tables 7 and 8 we present the medians of the measured RMSE and NSE values of the global dataset and for each of the groups of stations of Table 1, while the side-by-side boxplots of the RMSE values are presented in Fig. 9. These numerical results can facilitate a comparison on a common basis of the methods regarding their performance in the experiments using the temperature time series. In terms of RMSE, for the total of the temperature time series the use of a random walk variation (offering a median of 2.60 K or 2.66 K) instead of naïve (offering a median of 2.29 K) leads to about 14-16% less accurate forecasts, while the use of the remaining automatic methods (offering median values between 1.62 K and 1.86 K) to about 19-29% more accurate forecasts. For the time series observed in North America, North Europe, Siberia, Asia (except Siberia) and

Oceania these latter percentages are 18-30%, 30-32%, 19-25%, 17-31% and 10-15% respectively. The median values of the latter methods are close to the median of the estimated standard deviations of the deseasonalized time series (see Fig. 2). On the other hand, all the variations of the ARFIMA, BATS, simple exponential smoothing, Theta and Prophet methods except for the bats_5, bats_6 and prophet_3 are rather competitive to each other, while each of these categories of methods exhibits better or worse performance in comparison to the rest depending on the examined sample of time series. For example, prophet_1 exhibits the smallest median RMSE for the temperature forecasts for North Europe and Siberia, while offering 13-32% (depending on the examined set of time series) more accurate results than naïve. The variations belonging to each of the {arfima_1, arfima_2, arfima_3, arfima_4}, {bats_1, bats_2, bats_3, bats_4}, {ses_1, ses_2, ses_3, ses_4}, {theta_1, theta_2}, {prophet_1, prophet_2} sets differ in their performance at maximum about 1%, while the results do not suggest any specific combination of choices for the external handling of seasonality and non-normality as best for each algorithm. Nevertheless, the handling of the seasonality through the BATS and Prophet forecasting algorithms leads to less accurate forecasts than the external one, especially for the former algorithm. These facts are illustrated in Fig. 9 as well. Finally, all NSE values of Table 8 indicate a good forecasting performance for the total of the methods.

Table 7. Medians of the RMSE values (K) of the forecasts for each group of temperature stations. The best performance (when rounding to more than four digits) for each model is in bold.

Method	Globe	North America	North Europe	Siberia	Asia (except Siberia)	Oceania
naïve	**2.29**	**2.70**	**3.14**	**3.49**	**1.61**	**1.19**
rw_1	**2.60**	3.59	2.84	3.22	**1.99**	1.43
rw_2	2.60	**3.57**	**2.84**	**3.22**	1.99	1.43
rw_3	2.66	3.65	2.84	3.26	2.02	**1.42**
rw_4	2.66	3.64	2.84	3.25	2.02	1.42
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

Table 8. Medians of the NSE values of the forecasts for each group of temperature stations. The best performance (when rounding to more than four digits) for each model is in bold.

Method	Globe	North America	North Europe	Siberia	Asia (except Siberia)	Oceania
naïve	**0.91**	**0.90**	**0.79**	**0.93**	**0.95**	**0.90**
rw_1	0.89	0.81	0.83	0.94	**0.93**	0.89
rw_2	**0.89**	**0.81**	**0.83**	**0.94**	0.93	0.89
rw_3	0.89	0.80	0.82	0.94	0.93	0.89
rw_4	0.89	0.80	0.82	0.94	0.93	**0.89**
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**

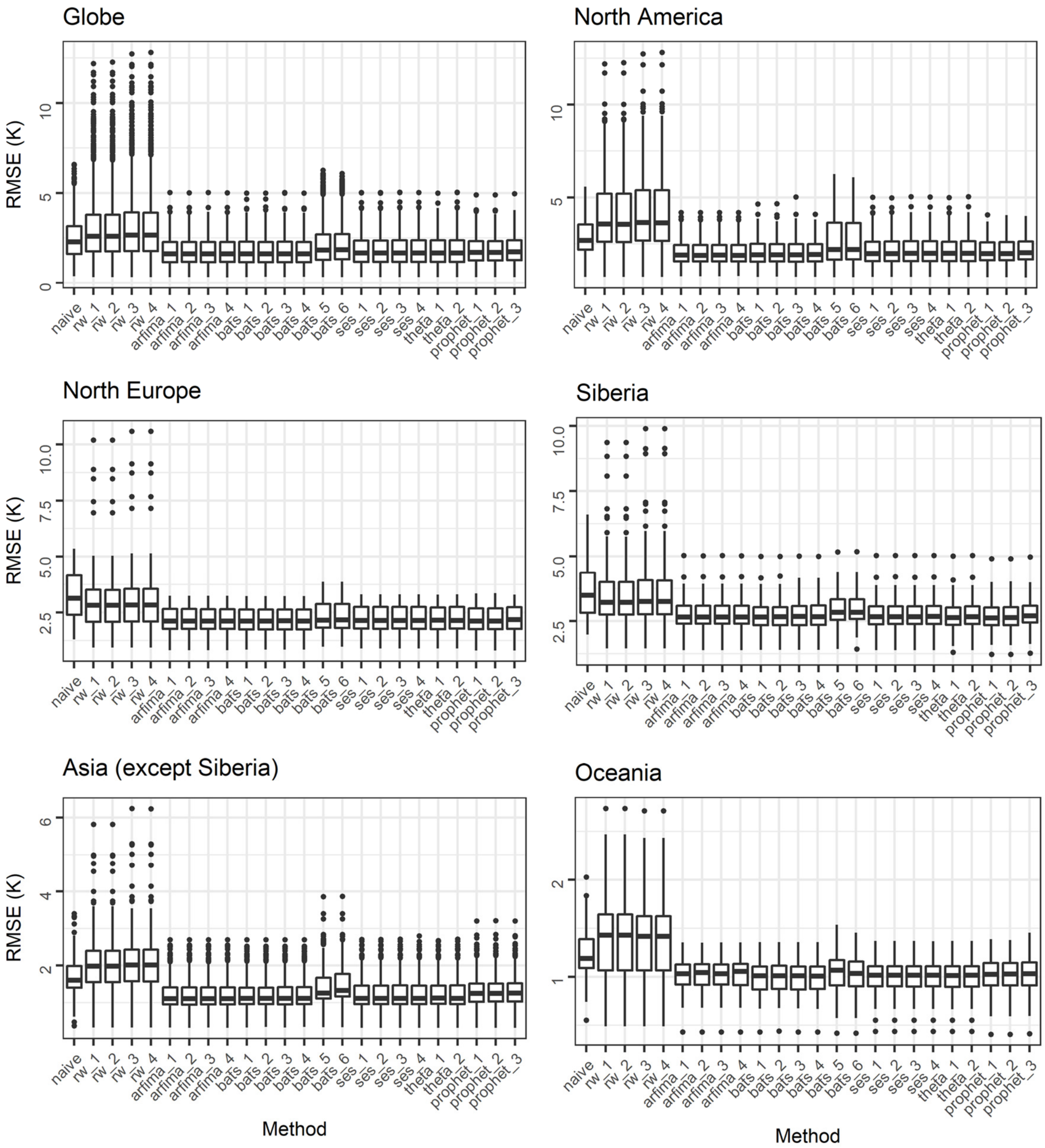

Fig. 9. RMSE for the temperature time series.

3.2 Analysis using the precipitation time series

Section 3.2 is devoted to the results of the analysis using the precipitation time series. In Fig. 10 we present the median values of the absolute errors computed for the total of the precipitation time series. Here as well, the random walk variations are the least accurate at almost every step of the examined horizon with a minimum median around 25 mm and a maximum around 48 mm. The naïve method is also worse than the rest with minimum and maximum medians around 20 mm and 39 mm respectively. The best average performance is around 15 mm, which is about 70% smaller than the median of the estimated standard deviations of the deseasonalized time series (see Fig. 2), the same as applying to the temperature forecasts. Moreover, in Fig. 11 and 12 we present the median values of the absolute errors computed for each of the groups of stations of Table 2. First, we observe that Fig. 10 approximates more to Fig. 12(c) than to Fig. 11 or Fig. 12(a, b). This is a rather expected outcome, since most of the totally examined precipitation time series originate from Australia. Nevertheless, the medians of the absolute errors are larger in this geographical region than worldwide. Furthermore, in Australia and North Europe there is not a clear pattern of seasonality in the presented medians, while for the case of North America there is one, but only for the years 1988 and 1989. In North Africa and East Asia, on the contrary, the medians between April and October are clearly higher than for the rest of the months with a peak in August (or July), indicating a larger difficulty in their corresponding forecasting attempts. This is due to a more regular precipitation variability in these geographical regions. The same applies, to a smaller extend, to the case of South Africa, for which the precipitation variables between October and April are found to be the least predictable. Finally, in North Africa, South Africa and East Asia some medians of the absolute errors are equal or very close to zero. All the above-stated facts may be largely explained in Fig. 4.

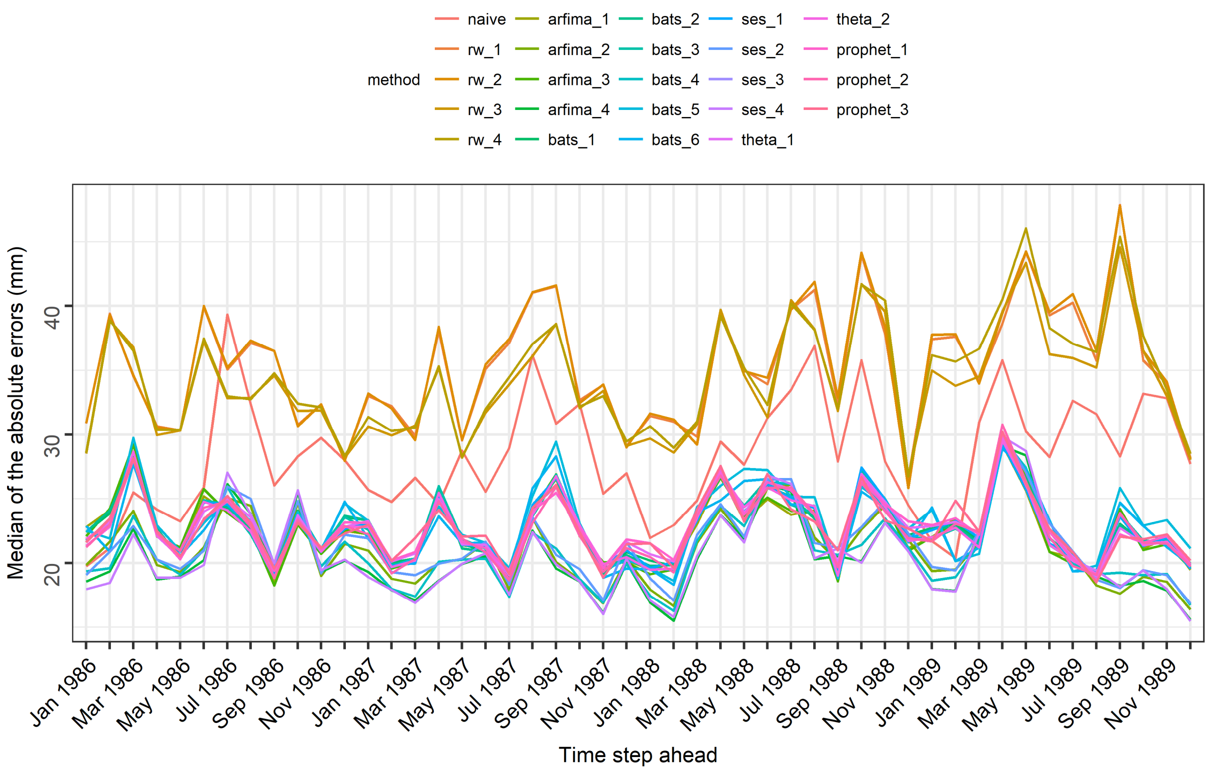

Fig. 10. Medians of the absolute errors at each time step of the forecast horizon for the total of the precipitation time series.

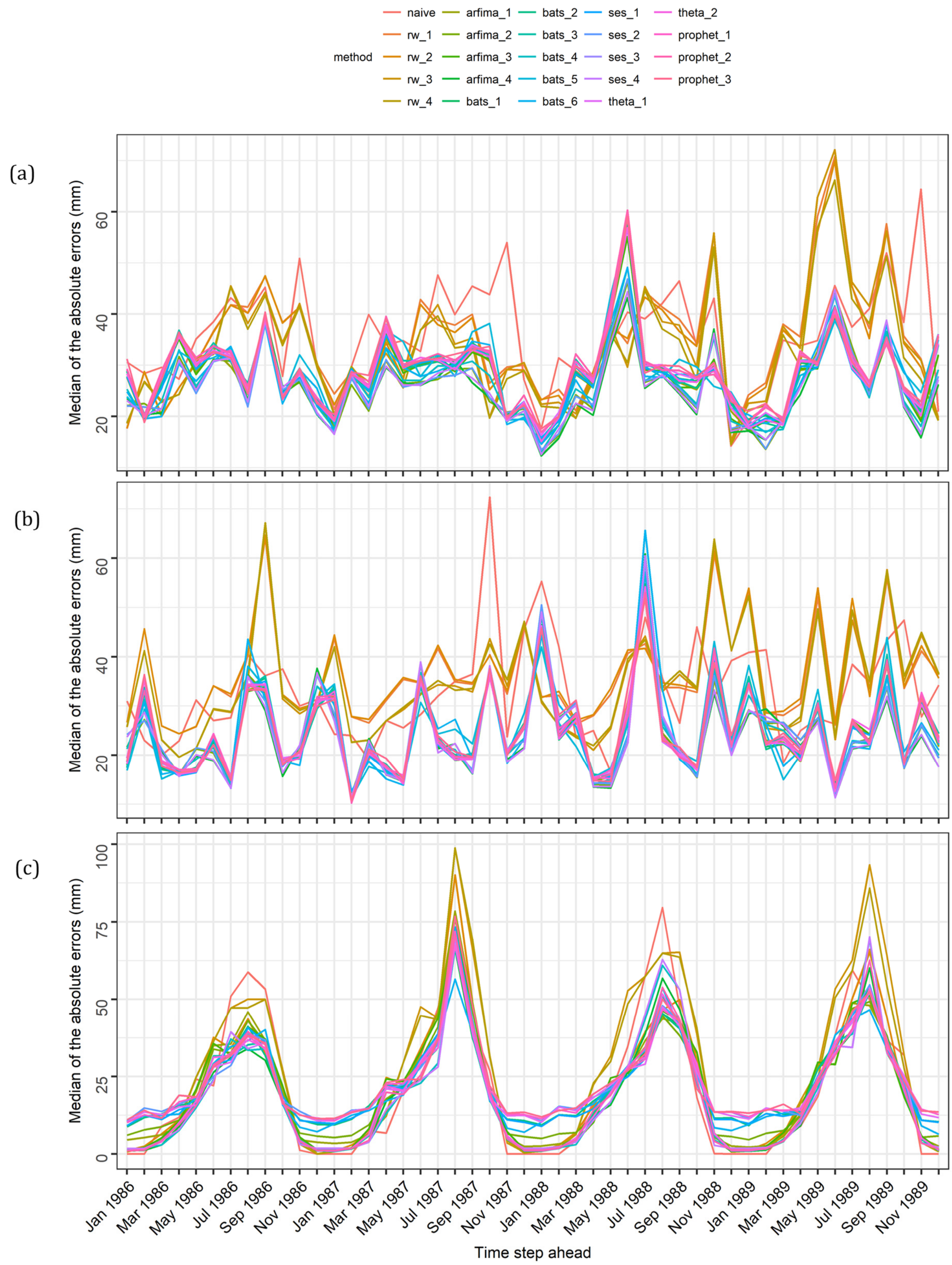

Fig. 11. Medians of the absolute errors at each time step of the forecast horizon for the precipitation time series observed in: (a) North America, (b) North Europe and (c) North Africa.

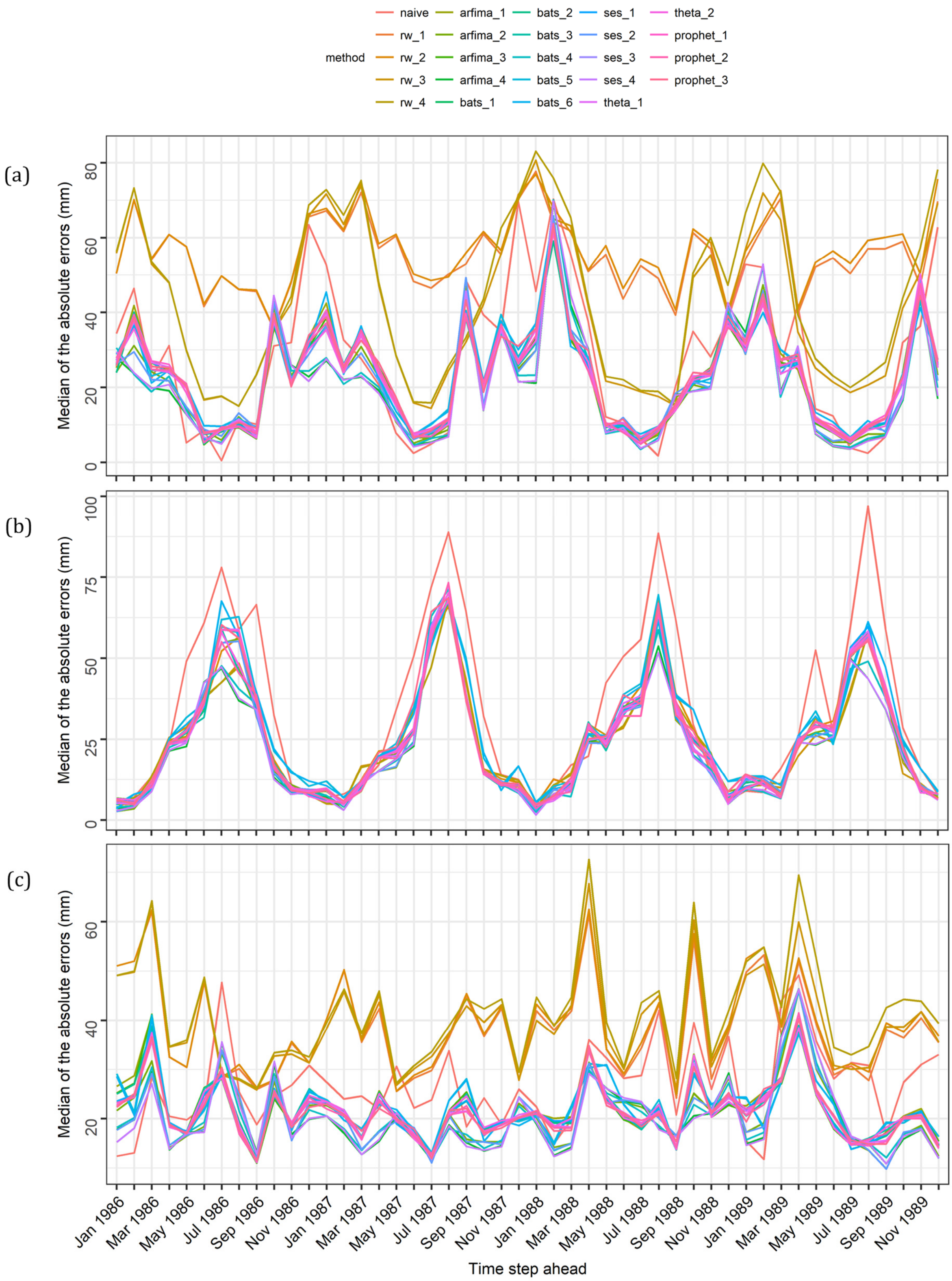

Fig. 12. Medians of the absolute errors at each time step of the forecast horizon for the precipitation time series observed in: (a) South Africa, (b) East Asia and (c) Australia.

In Fig. 13 and 14 we present in more detail the medians of the absolute errors for two special cases, i.e. those of North and South Africa respectively. In these figures we individually compare the {rw_1, rw_2, rw_3, rw_4}, {arfima_1, arfima_2, arfima_3, arfima_4}, {bats_1, bats_2 bats_3, bats_4, bats_5, bats_6}, {ses_1, ses_2, ses_3, ses_4}, {theta_1, theta_2}, {prophet_1, prophet_2, prophet_3} sets of variations. As illustrated in Fig. 13, in North Africa the tested choices for handling the seasonality seem to result in different seasonality patterns in the median values of the absolute errors, a fact not applying to the choices for handling the non-normality. Particularly for this specific geographical region the use of the additive model results in larger absolute errors than the multiplicative model from October to April and to smaller absolute errors for the rest of the year. These differences are more visible for the BATS, simple exponential smoothing, Theta and Prophet variations, but also exist for the random walk and ARFIMA ones. In South Africa two discrete seasonality patterns are observed only for the random walk variations.

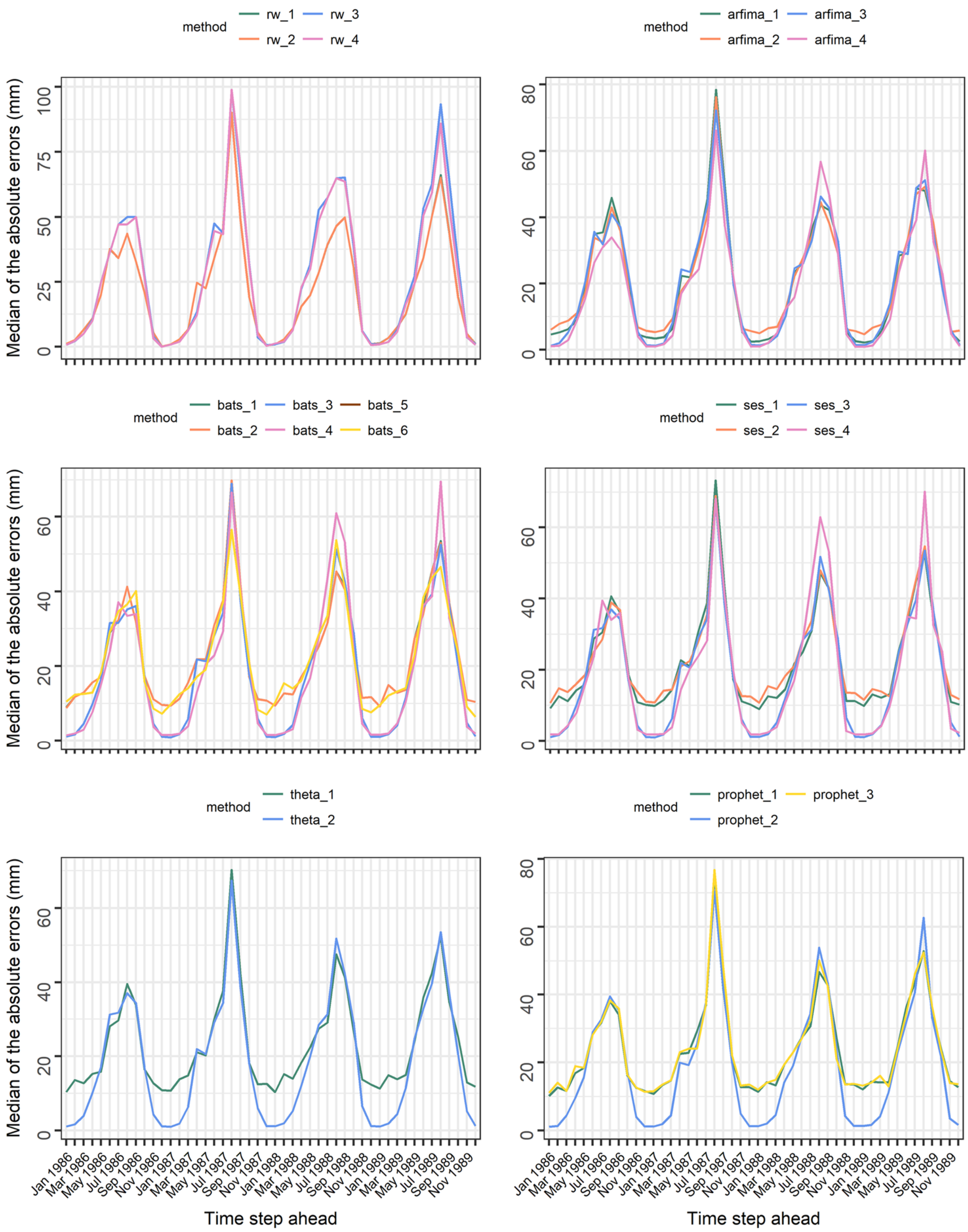

Fig. 13. Medians of the absolute errors at each time step of the forecast horizon for the precipitation time series observed in North Africa: comparison among the methods using the same model.

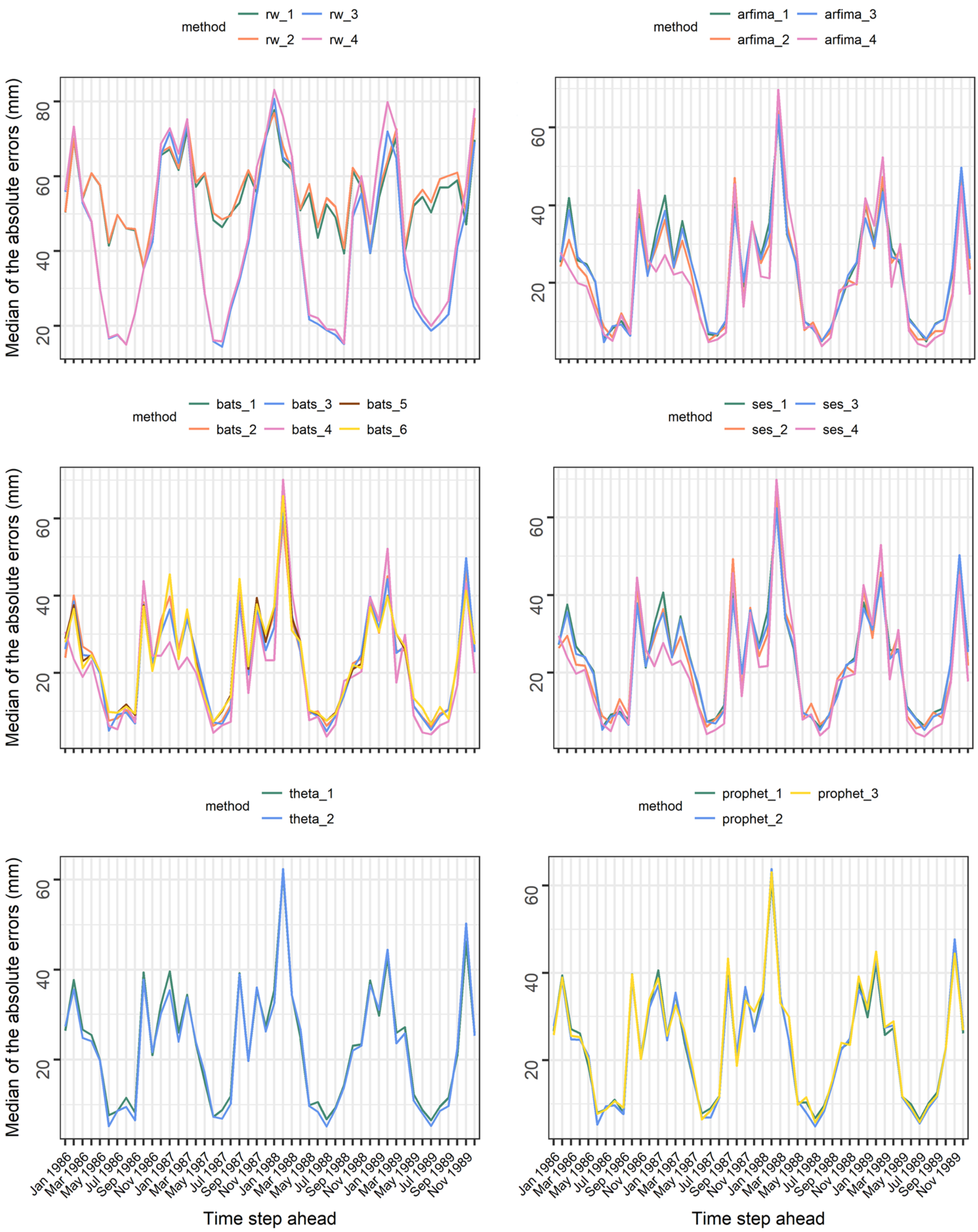

Fig. 14. Medians of the absolute errors at each time step of the forecast horizon for the precipitation time series observed in South Africa: comparison among the methods using the same model.

Additionally, in Tables 9 and 10 we present the medians of the measured RMSE and NSE values of the global dataset and for each of the groups of stations of Table 2, while the side-by-side boxplots of the RMSE values are presented in Fig. 15. In the latter we notice the similarity in the performance of the ARFIMA, BATS, simple exponential smoothing, Theta and Prophet variations, which is also reported for the analysis of Section 3.1. Some (absolute and relative) differences in the forecasting performance of the methods are also evident. For example, for the case of North Africa rw_1 and rw_2 are more accurate than rw_3 and rw_4, while the opposite applies to the case of North Europe albeit to a smaller extend. By the examination of Table 9 we observe that for the total of the precipitation time series the use of all the automatic methods apart from the random walk variations (offering median values between 41.67 mm and 42.39 mm) instead of naïve (offering a median value of 53.74 mm) leads to about 21-22% more accurate forecasts in terms of RMSE. For the time series observed in North America, North Europe and East Asia these percentages are 26-29%, 22-24% and 32-38% respectively, while for those observed in North Africa, South Africa and Australia 18-25%, 15-18% and 19-22% respectively.

Here as well, the Prophet model is competitive to ARFIMA, BATS, simple exponential smoothing and Theta models, offering from 16% up to 38% (depending on the examined set of time series) better results than the naïve method, while exhibiting the smallest RMSE amongst all the methods for the precipitation forecasts for East Asia. The median values of the best-performing automatic methods are close to the median of the estimated standard deviations of the deseasonalized time series (see Fig. 2), as also applying to the temperature forecasts. Furthermore, the variations belonging to each of the {arfima_1, arfima_2, arfima_3, arfima_4}, {bats_1, bats_2, bats_3, bats_4}, {ses_1, ses_2, ses_3, ses_4}, {theta_1, theta_2}, {prophet_1, prophet_2} sets cannot be ranked either using the results of the experiments on precipitation time series, while the use of bats_5, bats_6 and prophet_3 has mostly (but not in all cases) led to less accurate forecasts. Regarding the different seasonality patterns illustrated in Fig. 13, these do not result in some dramatic difference in the numerical results in contrast to those illustrated in Fig. 14 for the case of the random walk variations. Finally, the NSE values of Table 10 are far worse than those corresponding to the temperature forecasts. Still, most of them are greater than zero and, thus, indicate acceptable performances, while for the geographical regions of East Asia and North Africa the performances could be characterized as moderate.

Table 9. Medians of the RMSE values (mm) of the forecasts for each group of precipitation stations. The best performance for each model is in bold.

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**
rw_1	**53.65**	**55.35**	48.46	**47.82**	69.97	**48.69**	**51.70**
rw_2	54.04	55.39	48.50	47.82	70.17	48.70	53.18
rw_3	56.00	56.68	**46.89**	61.11	**66.18**	54.13	55.36
rw_4	56.53	56.53	47.47	58.88	67.33	55.22	57.46
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

Table 10. Medians of the NSE values of the forecasts for each group of precipitation stations. The best performance for each model is in bold.

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**
rw_1	**−0.17**	−0.28	−0.14	**0.68**	−0.20	**0.49**	**−0.33**
rw_2	−0.17	**−0.24**	−0.15	0.68	−0.20	0.49	−0.34
rw_3	−0.20	−0.34	**−0.09**	0.49	**0.01**	0.42	−0.36
rw_4	−0.20	−0.30	−0.11	0.51	−0.02	0.44	−0.37
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

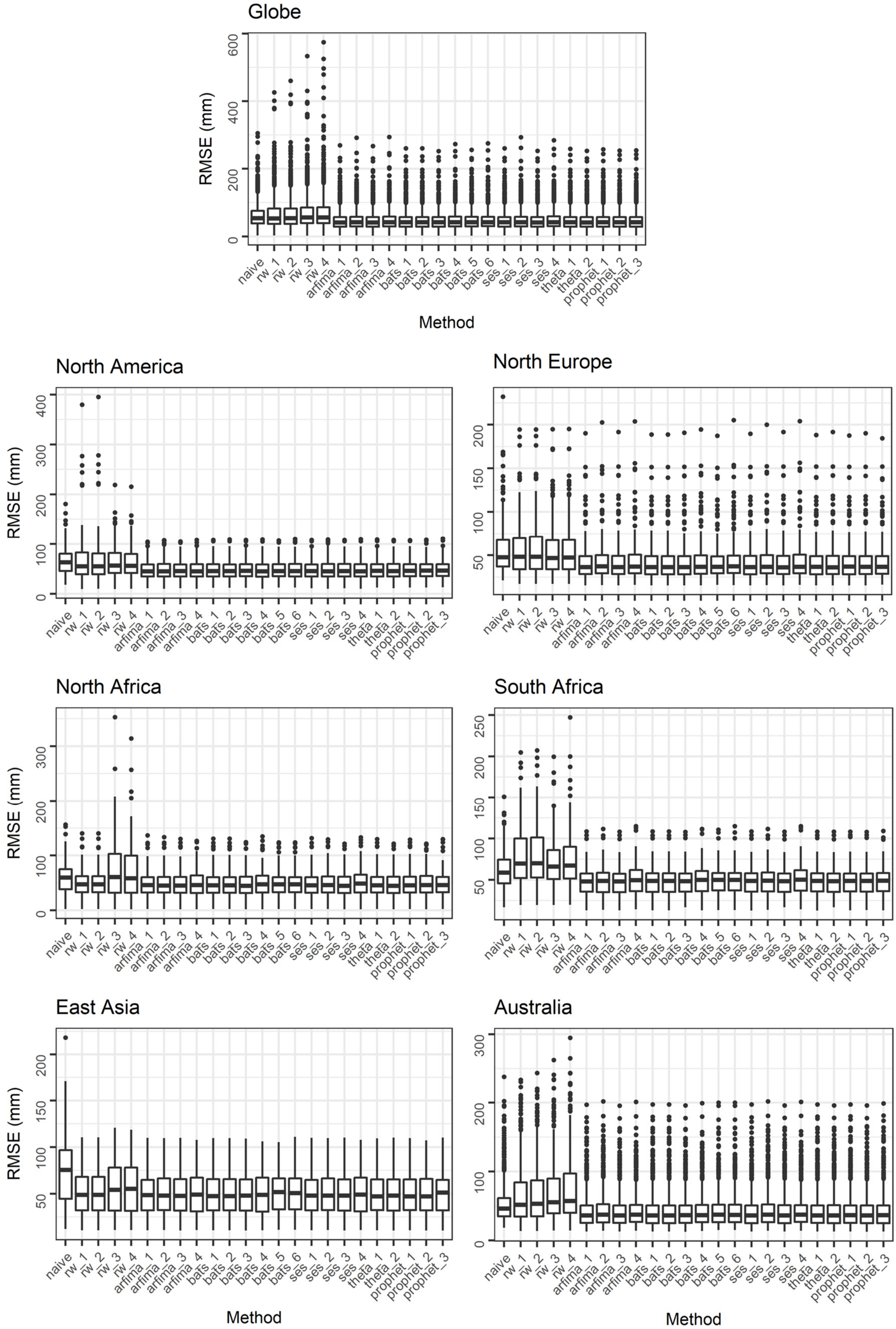

Fig. 15. RMSE for the precipitation time series.

4. Summary and discussion

We investigate the predictability of monthly temperature and precipitation by applying seven automatic univariate time series forecasting methods to 985 and 1552 monthly time series of temperature and precipitation respectively. The methods include a naïve one based on the monthly values of the last year, while the rest are based on the random walk (with drift), ARFIMA, BATS, simple exponential smoothing, Theta and Prophet models. Prophet is a recently introduced model inspired by the nature of time series forecasted at Facebook and it has not been applied to hydrometeorological time series before. The ARFIMA model, on the other hand, is widely used in a non-automatic way in the hydrological literature, while the rest of the models have been rarely implemented in hydrology, e.g. in Papacharalampous et al. (2017a,c), although they are very common in the forecasting literature. In the latter studies, no investigation is provided on how different choices of handling the seasonality and non-normality affect the performance of the models. This investigation constitutes one of the main aims of the present study (therefore, proper variations of the methods are examined), together with the quantification of the performance of the selected models on monthly hydrometeorological time series and the comparison of the Prophet model to the rest. The used time series are 480 months long with no missing values, observed between January 1950 and December 1989 in stations covering a significant part of the Earth's surface and, therefore, including various real-world process behaviours. The models are fitted in the first 36 years of data (432 months) and subsequently tested in performing multi-step ahead forecasts for the last four years of data (48 months). The results are summarized in global scores, while their examination by group of stations leads to five individual scores for temperature and six for precipitation. The groups are formed according to the geographical vicinity of the stations.

The results indicate that all the examined methods apart from the naïve and random walk ones are accurate enough to be used in long-term forecasting applications. Even the simple exponential smoothing 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. Seasonality is also the reason why patterns of error evolution, investigated in Papacharalampous et al. (2017c), are not revealed within the experiments of the present study, although the forecasting horizon is long enough here as well. The above-stated qualitative outcome is consistent with the 92 single-case studies of Papacharalampous et al. (2017a) and the 50 single-case studies of Papacharalampous et al. (2017b), which use monthly streamflow data and monthly temperature and precipitation data respectively. In these studies, the seasonality term is estimated using the multiplicative model. 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. The minimum and maximum medians of the absolute errors of the temperature forecasts are found to be around 0.25 K and 8.2 K respectively. Furthermore, a zero median of the absolute errors is computed for the precipitation forecasts produced for the dry months in geographical regions with relatively regular variability in precipitation, while the maximum median computed is around 100 mm. These values could be viewed in comparison with the minimum and maximum medians of absolute errors for annual temperature and precipitation, as derived in a study of Papacharalampous, Tyralis and Koutsoyiannis (manuscript in review, 2018) using two real-world datasets of 297 time series in total, which are about 0.23 K and 1.10 K, and 68 mm and 189 mm respectively. Moreover, the computed RMSE values range between 1.01 K and 3.65 K for temperature, and 36.16 mm and 70.17 mm for precipitation, while the respective NSE values are 0.79 and 0.98 for temperature, and -0.55 and 0.71 for precipitation.

Excluding the naïve method and the variations using the random walk model, 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, simple exponential smoothing, 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. This higher degree of accuracy is non-ignorable and particularly important in a long run perspective. Importantly, the Prophet model is found to offer from 13% up to 32% and from 16% up to 38% better results than the naïve method for the temperature and precipitation time series respectively. Moreover, the minimum and maximum NSE medians for the ARFIMA, BATS, simple exponential smoothing, Theta and Prophet models are 0.89 and 0.98 for temperature, and -0.04 and 0.71 for precipitation. The former NSE values indicate good forecasting performances and the latter acceptable to moderate. The higher predictability of the monthly temperature compared to the monthly precipitation is expected already from the comparison of their corresponding standard deviation values of the seasonally decomposed time series, which have a median around 1.7 K and 42 mm 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.

5. Conclusions

We investigate the predictability of monthly temperature and precipitation, and simultaneously assess the multi-step ahead performance of seven automatic univariate time series forecasting methods by applying the latter to the largest sample of hydrometeorological time series ever used for such purposes. The implemented methods are a naïve one based on the monthly values of the last year, as well as random walk (with drift), ARFIMA (acronym for AutoRegressive Fractionally Integrated Moving Average), BATS (acronym for Box–Cox transform, ARMA errors, Trend, and Seasonal components), simple exponential smoothing, Theta and Prophet. The latter is a recently introduced model, inspired by the nature of time series forecasted at Facebook and never applied to geophysical processes in the past, while most of the remaining methods are rarely used in hydrology. Proper variations of the methods are examined to further investigate how different choices of handling the seasonality and non-normality affect the performance of the models. The results indicate that (a) the last five models perform well, better than the naïve and random walk methods, (b) monthly temperature and precipitation can be forecasted to a level of accuracy which can barely be improved using other methods, (c) the externally applied classical seasonal decomposition results mostly in better forecasts compared to the automatic seasonal decomposition and (d) the Prophet forecasting method is competitive, especially when it is combined with externally applied classical seasonal decomposition.

Appendix A Statistical software

The analyses and visualizations have been performed in R Programming Language (R Core Team 2017). We have used the following contributed R packages: 'devtools' (Wickham and

Chang 2017), 'forecast' (Hyndman et al. 2017), 'fracdiff' (Fraley et al. 2012), 'gdata' (Warnes et al. 2017), 'ggplot2' (Wickham 2016), 'HKprocess' (Tyralis 2016), 'knitr' (Xie, 2014; 2015; 2017), 'maps' (Brownrigg et al. 2017), 'prophet' (Taylor and Letham 2017b) and 'readr' (Wickham et al., 2017).

Acknowledgements: We thank Prof. Krzysztof Kochanek for the invitation to submit a manuscript to the special issue of the Acta Geophysica "New Insights in Statistical Hydrology". We also thank two anonymous reviewers, whose comments have led to the improvement of this paper.

Conflict of interest: The authors declare no conflict of interest.

References

Andrawis RR, Atiya AF, El-Shishiny H (2011) Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. Int J Forecasting 27(3):672–688. https://doi.org/10.1016/j.ijforecast.2010.09.005
Armstrong JS, Fildes R (2006) Making progress in forecasting. Int J Forecasting 22(3):433–441. https://doi.org/10.1016/j.ijforecast.2006.04.007
Assimakopoulos V, Nikolopoulos K (2000) The theta model: a decomposition approach to forecasting. Int J Forecasting 16(4):521–530. https://doi.org/10.1016/S0169-2070(00)00066-2
Bărbulescu A (2016) Studies on Time Series Applications in Environmental Sciences. Springer International Publishing, Cham, Switzerland. https://doi.org/10.1007/978-3-319-30436-6
Box GEP, Cox DR (1964) An Analysis of Transformations. J Roy Stat Soc B Met 26(2):211–252
Box GEP, Jenkins GM (1968) Some recent advances in forecasting and control. J R Stat Soc C-Appl 17(2):91–109. https://doi.org/10.2307/2985674
Brown RG (1959) Statistical forecasting for inventory control. McGraw-Hill, New York
Brownrigg R, Minka TP, Deckmyn A (2017) maps: Draw Geographical Maps. R package version 3.2.0. https://CRAN.R-project.org/package=maps
Carlson RF, MacCormick AJA, Watts DG (1970) Application of linear random models to four annual streamflow series. Water Resour Res 6(4):1070–1078. https://doi.org/10.1029/WR006i004p01070
Chatfield C (1988) What is the 'best' method of forecasting?. J Appl Stat 15(1):19–38. https://doi.org/10.1080/02664768800000003
Chevillon G (2007) Direct multi-step estimation and forecasting. J Econ Surv 21(4):746–785. https://doi.org/10.1111/j.1467-6419.2007.00518.x
De Gooijer JG, Hyndman RJ (2006) 25 years of time series forecasting. Int J Forecasting 22(3):443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001
De Gooijer JG, Klein A (1992) On the cumulated multi-step-ahead predictions of vector autoregressive moving average processes. Int J Forecasting 7(4):501–513. https://doi.org/10.1016/0169-2070(92)90034-7
De Gooijer JG, Kumar K (1992) Some recent developments in non-linear time series modelling, testing, and forecasting. Int J Forecasting 8(2):135–156. https://doi.org/10.1016/0169-2070(92)90115-P

De Livera AM, Hyndman RJ, Snyder RS (2011) Forecasting time series with complex seasonal patterns using exponential smoothing. J Am Stat Assoc 106(496):1513–1527. https://doi.org/10.1198/jasa.2011.tm09771

Fraley C, Leisch F, Maechler M, Reisen V, 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

Franses PH, Legerstee R (2010) A unifying view on multi-step forecasting using an autoregression. J Econ Surv 24(3):389–401. https://doi.org/10.1111/j.1467-6419.2009.00581.x

Guerrero VM (1993) Time-series analysis supported by power transformations. J Forecasting 12(1):37–48. https://doi.org/10.1002/for.3980120104

Green KC, Armstrong JS (2007) Global warming: Forecasts by scientists versus scientific forecasts. Energ Environ-UK 18(7):997–1021. https://doi.org/10.1260/095830507782616887

Green KC, Armstrong JS, Soon W (2009) Validity of climate change forecasting for public policy decision making. Int J Forecasting 25(4):826–832. https://doi.org/10.1016/j.ijforecast.2009.05.011

Haslett J, Raftery AE (1989) Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource. J R Stat Soc C-Appl 38(1):1–50. https://doi.org/10.2307/2347679

Hyndman RJ, Athanasopoulos G (2017) Forecasting: principles and practice. OTexts: Melbourne, Australia. http://otexts.org/fpp2/

Hyndman RJ, Billah B (2003) Unmasking the Theta method. Int J Forecasting 19(2):287–290. https://doi.org/10.1016/S0169-2070(01)00143-1

Hyndman RJ, Khandakar Y (2008) Automatic time series forecasting: the forecast package for R. J Stat Softw 27(3):1–22. https://doi.org/10.18637/jss.v027.i03

Hyndman RJ, Koehler AB, Ord JK, Snyder RD (2008) Forecasting with exponential smoothing: The state space approach. Springer-Verlag Berlin Heidelberg, pp 3-7. https://doi.org/10.1007/978-3-540-71918-2

Hyndman RJ, O'Hara-Wild M, Bergmeir C, Razbash S, Wang E (2017) forecast: Forecasting functions for time series and linear models. R package version 8.2. https://CRAN.R-project.org/package=forecast

Lawrimore JH, Menne MJ, Gleason BE, Williams CN, Wuertz DB, Vose RS, Rennie J (2011) An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. J Geophys Res-Atmos 116(D1912). https://doi.org/10.1029/2011JD016187

Makridakis S, Hibon M, Lusk E, Belhadjali M (1987) Confidence intervals: An empirical investigation of the series in the M-competition. Int J Forecasting 3(3–4):489–508. https://doi.org/10.1016/0169-2070(87)90045-8

Makridakis S, Hibon M (2000) The M3-Competition: results, conclusions and implications. Int J Forecasting 16(4):451–476. https://doi.org/10.1016/S0169-2070(00)00057-1

Mills TC (2011) The Foundations of Modern Time Series Analysis. Palgrave Macmillan, UK. https://doi.org/10.1057/9780230305021

Montanari A, Rosso R, Taqqu MS (1997) Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. Water Resour Res 33(5):1035–1044. https://doi.org/10.1029/97WR00043

Montanari A, Rosso R, Taqqu MS (2000) A seasonal fractional ARIMA Model applied to the Nile River monthly flows at Aswan. Water Resour Res 36(5):1249–1259. https://doi.org/10.1029/2000WR900012

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. J Hydrol 10(3):282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Papacharalampous GA, Tyralis H, Koutsoyiannis D (2017a) Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. Preprints 2017100133. https://doi.org/10.20944/preprints201710.0133.v1

Papacharalampous GA, Tyralis H, Koutsoyiannis D (2017b) Forecasting of geophysical processes using stochastic and machine learning algorithms. European Water 59:161-168

Papacharalampous GA, Tyralis H, Koutsoyiannis D (2017c) Error Evolution in Multi-Step Ahead Streamflow Forecasting for the Operation of Hydropower Reservoirs. Preprints 2017100129. https://doi.org/10.20944/preprints201710.0129.v1

Pemberton J (1987) Exact least squares multi-step prediction from nonlinear autoregressive models. J Time Ser Anal 8(4):443-448. https://doi.org/10.1111/j.1467-9892.1987.tb00007.x

Peterson TC, Vose RS (1997) An Overview of the Global Historical Climatology Network Temperature Database. B Am Meteorol Soc 78:2837–2849. https://doi.org/10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2

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

Singh M, Singh R, Shinde V (2011) Application of software packages for monthly stream flow forecasting of Kangsabati River in India. Int J Comput Appl 20(3):7–14

Sivakumar B (2017) Chaos in hydrology: Bridging determinism and stochasticity. Springer, Netherlands. https://doi.org/10.1007/978-90-481-2552-4

Stoica P, Nehorai A (1989) On multistep prediction error methods for time series models. J Forecasting 8(4):357–368. https://doi.org/10.1002/for.3980080402

Taieb SB, Atiya AF (2016) A Bias and Variance Analysis for Multistep-Ahead Time Series Forecasting. IEEE T Neur Net Lear 27(1):62–76. https://doi.org/10.1109/TNNLS.2015.2411629

Taieb SB, Bontempi G, Atiya AF, Sorjamaa A (2012) A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert Syst Appl 39 (8), 7067–7083. https://doi.org/10.1016/j.eswa.2012.01.039

Taylor SJ, Letham B (2017a) Forecasting at scale. Am Stat. https://doi.org/10.1080/00031305.2017.1380080

Taylor SJ, Letham B (2017b) prophet: Automatic Forecasting Procedure. R package version 0.2. https://CRAN.R-project.org/package=prophet

Tyralis H (2016) HKprocess: Hurst-Kolmogorov Process. R package version 0.0-2. https://CRAN.R-project.org/package=HKprocess

Tyralis H, Koutsoyiannis D (2011) Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. Stoch Env Res Risk A 25(1):21–33. https://doi.org/10.1007/s00477-010-0408-x

Tyralis H, Koutsoyiannis D (2014) A Bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. Clim Dynam 42(11):2867–2883. https://doi.org/10.1007/s00382-013-1804-y

Valipour M, Banihabib ME, Behbahani SMR (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J Hydrol 476(7):433–441. https://doi.org/10.1016/j.jhydrol.2012.11.017

Warnes GR, 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

Wei WWS (2006) Time Series Analysis, Univariate and Multivariate Methods, second edition. Pearson Addison Wesley, Boston

Wickham H (2016) ggplot2. Springer International Publishing, Houston, Texas. https://doi.org/10.1007/978-3-319-24277-4
Wickham H, Chang W (2017) devtools: Tools to Make Developing R Packages Easier. R package version 1.13.4. https://CRAN.R-project.org/package=devtools
Wickham H, Hester J, Francois R, Jylänki J, Jørgensen M (2017) readr: Read Rectangular Text Data. R package version 1.1.1. https://CRAN.R-project.org/package=readr
Xie Y (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In: Stodden V, Leisch F, Peng RD (eds) Implementing Reproducible Research. Chapman and Hall/CRC, Boca Raton, Florida, pp 3-32
Xie Y (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC, Boca Raton, Florida
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