Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: A multiple-case study from Greece

Georgia Papacharalampous[1,*], Hristos Tyralis[2], and Demetris Koutsoyiannis[3]

[1] Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Polytechniou 5, 157 80 Zografou, Greece; papacharalampous.georgia@gmail.com; ORCID: 0000-0001-5446-954X

[2] Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Polytechniou 5, 157 80 Zografou, Greece; montchrister@gmail.com; ORCID: 0000-0002-8932-4997

[3] Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Polytechniou 5, 157 80 Zografou, Greece; dk@itia.ntua.gr; ORCID: 0000-0002-6226-0241

* Correspondence: papacharalampous.georgia@gmail.com, tel: +30 69474 98589

Accepted in Water Resources Management, doi: 10.1007/s11269-018-2155-6

Abstract: We provide contingent empirical evidence on the solutions to three problems associated with univariate time series forecasting using machine learning (ML) algorithms by conducting an extensive multiple-case study. These problems are: (a) lagged variable selection, (b) hyperparameter handling, and (c) comparison between ML and classical algorithms. The multiple-case study is composed by 50 single-case studies, which use time series of mean monthly temperature and total monthly precipitation observed in Greece. We focus on two ML algorithms, i.e. neural networks and support vector machines, while we also include four classical algorithms and a naïve benchmark in the comparisons. We apply a fixed methodology to each individual case and, subsequently, we perform a cross-case synthesis to facilitate the detection of systematic patterns. We fit the models to the deseasonalized time series. We compare the one- and multi-step ahead forecasting performance of the algorithms. Regarding the one-step ahead forecasting performance, the assessment is based on the absolute error of the forecast of the last monthly observation. For the quantification of the multi-step ahead forecasting performance we compute five metrics on the test set (last year's monthly observations), i.e. the root mean square error, the Nash-Sutcliffe efficiency, the ratio of standard deviations, the coefficient of correlation and the index of agreement. The evidence derived by the experiments can be summarized as follows: (a) the results mostly favour using less recent lagged variables, (b) hyperparameter optimization does not necessarily lead to better forecasts, (c) the ML and classical algorithms seem to be equally competitive.

Key Words: case studies; cross-case synthesis; hyperparameter optimization; lagged variable selection; multi-step ahead forecasting; one-step ahead forecasting

1. Introduction

1.1 Background information

Machine learning (ML) algorithms are widely used for the forecasting of univariate geophysical time series as an alternative to classical algorithms. Popular ML algorithms are the rather well-established Neural Networks (NN) and the new-entrant in most scientific fields Support Vector Machines (SVM). The latter algorithm has been presented in its current form by Cortes and Vapnik (1995; see also Vapnik 1995, 1999). The large number and wide range of the relevant applications is apparent in the review papers of Maier and Dandy (2000), and Raghavendra and Deka (2014) respectively. The competence of ML algorithms in univariate time series forecasting has been empirically proven in Papacharalampous et al. (2017a), and Tyralis and Papacharalampous (2017) through extensive simulation experiments.

Nevertheless, univariate time series forecasting using ML algorithms also implies the handling of specific factors that may improve or deteriorate the performance of the algorithms, i.e. the lagged variables and the hyperparameters. In contrast to the typical regression problem, in a forecasting problem the set of predictor variables is a set of lagged variables, formed using observed past values of the process to be forecasted and, consequently, holding information about the temporal dependence. Although the amount of the available historical information taken into account increases when using a large number of lagged variables, the length of the fitting set concomitantly decreases; for more details, see Tyralis and Papacharalampous (2017). While there is a wide literature on applications of ML algorithms in hydrological univariate time series forecasting, mainly comprising single- or few-case studies that particularly focus on details about the model structure (e.g. Atiya et al. 1999; Guo et al. 2011; Hong 2008; Kumar et al. 2004; Moustris et al. 2011; Ouyang and Lu 2017; Sivapragasam et al. 2001; Wang et al. 2006), studies explicitly stating information concerning the variable selection issue, such as Belayneh et al. (2014), Nayak et al. (2004), Hung et al. (2009) and Yaseen et al. (2016), are less. Tyralis and Papacharalampous (2017) have investigated the effect of a sufficient number of lagged variable selection choices on the performance of the Breiman' s random forests algorithm (Breiman 2001) in one-step ahead univariate time series forecasting.

On the other hand, information on the hyperparameter selection is usually emphasized in the hydrological literature (e.g. Belayneh et al. 2014; Hung et al. 2009; Koutsoyiannis et al. 2008; El-Shafie et al. 2007; Tongal and Berndtsson 2017; Valipour et al. 2013; Yu et al. 2004). An example of a hyperparameter is the number of hidden nodes within a neural networks structure. Hyperparameters are distinguished from the basic parameters, because they are usually optimized or tuned with the aim to improve the performance of a ML algorithm. Hyperparameter optimization can be performed using a single validation set extracted from the fitting set or *k*-fold cross-validation, which involves multiple set divisions and tests. The optimal hyperparameter values are most frequently searched heuristically, either using grid search or random search, while ML or Bayesian methods can be adopted for this task as well (Witten et al. 2017). However, non-tuned ML models are also used in hydrology (e.g. Yaseen et al. 2016). Finally, a popular problem arising when using ML forecasting algorithms is the comparison between ML and classical algorithms. This problem is mostly examined within single-case studies (e.g. Ballini et al. 2001; Koutsoyiannis et al. 2008; Tongal and Berndtsson 2017; Valipour et al. 2013; Yu et al. 2004), as applying to lagged variable and hyperparameter selection as well.

1.2 Main contribution of this study

The main contribution of this study is the exploration in geoscience concepts of the problems presented in detail in Section 1.1 and summarized here below, together with their related research questions of focus:

- Problem 1: Lagged variable selection in time series forecasting using ML algorithms

  Research question 1: *Should we select less recent lagged variables or a large number of lagged variables in time series forecasting using ML algorithms?*

- Problem 2: Hyperparameter selection in time series forecasting using ML algorithms

  Research question 2: *Does hyperparameter optimization necessarily lead to a better performance in time series forecasting using ML algorithms?*

- Problem 3: Comparison between ML and classical algorithms

Research question 3: *Do the ML algorithms exhibit better (or worse) performance than the classical ones?*

In fact, exploration is indispensable for understanding the phenomena involved in a specific problem and, therefore, it constitutes an essential part within every theory-development process.

1.3 Research method and implementation

We adopt the multiple-case study research method (presented in detail in Yin (2003)), which embraces the examination of more than one individual cases, facilitating the observation of specific phenomena from multiple perspectives or within different contexts (Dooley 2002). For the detection of systematic patterns across the individual cases a cross-case synthesis can be performed (Larsson 1993). Given the fact that the boundaries between the phenomena and the context are not clear (thus, it is meaningful to consider a case study design, as explained in Baxter and Jack (2008)), it is important that each individual case keeps its identity within the multiple-case study, so that one can specifically focus on it. This exploration within and across the individual cases can provide interesting insights into the phenomena under investigation, as well as a form of generalization named *"contingent empirical generalization"*, while retaining the immediacy of the single-case study method (Achen and Snidal 1989).

We explore the three problems summarized in Section 1.2 by conducting an extensive multiple-case study composed by 50 single-case studies, which use temperature and precipitation time series observed in Greece. We examine these two geophysical processes, because they exhibit different properties, which may affect differently the results within the explorations. We focus on two ML algorithms, i.e. NN and SVM, for an analogous reason. Moreover, the explorations are conducted for the one- and a multi-step ahead horizons, as their corresponding forecasting attempts are not of the same difficulty. We apply a fixed methodology to each individual case. This fixed methodology provides the common basis to further perform a cross-case synthesis for the detection of systematic patterns across the individual cases. The latter is the novelty of our study.

2. Data and methods

2.1 Methodology outline

We conduct 50 single-case studies by applying a fixed methodology to each of the 50 time series presented in Section 2.2, as explained subsequently. First, we split the time series into a fitting and a test set. The latter is the last monthly observation for the one-step ahead forecasting experiments and the last year's monthly observations for the multi-step ahead forecasting experiments. Second, we fit the models to the seasonally decomposed fitting set, within the context described in Section 2.3, and make predictions corresponding to the test set. Third, we recover the seasonality in the predicted values and compare them to their corresponding observed using the metrics of Section 2.4. Finally, we perform a cross-case synthesis to demonstrate similarities and differences between the single-case studies conducted. We present the results per category of tests, which is determined by the set {set of methods, process, forecast horizon}, and further summarize them, as discussed in Section 2.4. The sets of methods are defined in Section 2.3, while the total number of categories is 20. We place emphasis on the exploration of the three problems summarized in Section 1.2, but we also present quantitative information about the produced forecasts and search for evidence regarding the existence of a possible relationship between the forecast quality, and the standard deviation (σ), coefficient of variation (cv) and Hurst parameter (H) estimates for the deseasonalized time series (available in Section 2.2). Statistical software information is summarized in Appendix A.

2.2 Time series

We use 50 time series of mean monthly temperature and total monthly precipitation observed in Greece. These time series are sourced from Lawrimore et al. (2011), and Peterson and Vose (1997) respectively. We select only those with few missing values (blocks with length equal or less than one). Subsequently, we use the Kalman filter algorithm of the `zoo` R package (Zeileis and Grothendieck 2005) for filling in the missing values. The basic information about the time series is provided in Table 1, while Figure 1 presents the locations of the stations at which the data has been recorded. We use the deseasonalized fitting sets for fitting the forecasting models, as suggested in Taieb et al. (2012) for the improvement of the forecast quality. The time series decomposition is performed exclusively on the fitting sets using the multiplicative model for the temperature time series and the additive model for the precipitation ones. The reason for this differentiation is that the use of the multiplicative model on the precipitation time series results in zero forecasts for some methods, as a result of zero precipitation observations in the summer months.

Table 1. Time series of the present study.

s/n	Process	Code	Location	Station information: ID	Latitude	Longitude	Reference	Start	End	Length (months)
1	Temperature	temp_1	Araxos	16687001	38.20	21.40	Lawrimore et al. (2011)	Jan 1951	Dec 1980	360
2		temp_2	Athens	16714000	37.97	23.72		Jan 1858	Dec 1975	1416
3		temp_3	Athens	16714000	37.97	23.72		Jan 1989	Dec 2001	156
4		temp_4	Athens	16716000	37.90	23.73		Jan 1951	Dec 2012	744
5		temp_5	Heraklion	16754000	35.33	25.18		Jan 1950	Dec 2015	792
6		temp_6	Kalamata	16726000	37.07	22.02		Jan 1956	Dec 2015	720
7		temp_7	Kerkyra	16641000	39.62	19.92		Jan 1951	Dec 2016	792
8		temp_8	Larissa	16648000	39.63	22.42		Jan 1899	Dec 2016	1416
9		temp_9	Lemnos	16650000	39.92	25.23		Jan 1951	Dec 1998	576
10		temp_10	Methoni	16734000	36.83	21.70		Jan 1951	Dec 1972	264
11		temp_11	Methoni	16734000	36.83	21.70		Jan 1975	Dec 2000	312
12		temp_12	Patra	16689000	38.25	21.73		Jan 1951	Dec 1989	468
13		temp_13	Samos	16723000	37.70	26.92		Jan 1955	Dec 1969	180
14		temp_14	Samos	16723000	37.70	26.92		Jan 1974	Dec 2003	360
15		temp_15	Souda	16746000	35.48	24.12		Jan 1961	Dec 2015	660
16		temp_16	Thessaloniki	16622000	40.52	22.97		Jan 1892	Dec 2016	1500
17		temp_17	Thessaloniki	16622001	40.52	23.02		Jan 1961	Dec 1970	120
18	Precipitation	prec_1	Agrinion	16672000	38.60	21.70	Peterson and Vose (1997)	Jan 1956	Dec 1987	384
19		prec_2	Alexandroupoli	16627000	40.80	25.90		Jan 1951	Dec 1990	480
20		prec_3	Aliartos	16674000	38.40	23.10		Jan 1907	Dec 1990	1008
21		prec_4	Anogeia	16754001	35.30	24.90		Jan 1919	Dec 1939	252
22		prec_5	Anogeia	16754001	35.30	24.90		Jan 1950	Dec 1979	360
23		prec_6	Araxos	16687000	38.20	21.40		Jan 1949	Dec 2000	624
24		prec_7	Athens	16714000	38.00	23.70		Jan 1860	Dec 1881	264
25		prec_8	Athens	16714000	38.00	23.70		Jan 1887	Dec 2005	1428
26		prec_9	Athens	16716000	37.90	23.70		Jan 1929	Dec 1945	204
27		prec_10	Fragma	16715001	38.20	23.90		Jan 1926	Dec 1990	780
28		prec_11	Heraklion	16754000	35.30	25.10		Jan 1946	Dec 1990	540
29		prec_12	Igoumenitsa	16641001	39.50	20.30		Jan 1951	Dec 1990	480
30		prec_13	Ioannina	16642000	39.70	20.80		Jan 1951	Dec 1990	480
31		prec_14	Kalamata	16726000	37.00	22.10		Jan 1956	Dec 1970	180
32		prec_15	Kalo Chorio	16756001	35.10	25.70		Jan 1950	Dec 1984	420
33		prec_16	Kastelli	16760001	35.20	25.30		Jan 1949	Dec 1976	336
34		prec_17	Kerkyra	16641000	39.60	19.90		Jan 1952	Dec 1996	540
35		prec_18	Kythira	16743000	36.30	23.00		Jan 1951	Dec 1973	276
36		prec_19	Kos	16742000	36.80	27.10		Jan 1958	Dec 1990	396
37		prec_20	Kozani	16632000	40.30	21.80		Jan 1955	Dec 1987	396
38		prec_21	Larissa	16648000	39.60	22.40		Jan 1951	Dec 1997	564
39		prec_22	Lemnos	16650001	39.90	25.30		Jan 1951	Dec 2000	600
40		prec_23	Methoni	16734000	36.80	21.70		Jan 1951	Dec 1991	492
41		prec_24	Milos	16738000	36.70	24.50		Jan 1951	Dec 1990	480
42		prec_25	Mytilene	16667000	39.10	26.60		Jan 1952	Dec 1990	468
43		prec_26	Naxos	16732000	37.10	25.50		Jan 1955	Dec 1971	204
44		prec_27	Patra	16689000	38.20	21.70		Jan 1901	Dec 1984	1008
45		prec_28	Sitia	16757000	35.20	26.10		Jan 1960	Dec 1983	288
46		prec_29	Skyros	16684000	38.90	24.60		Jan 1955	Dec 1987	396
47		prec_30	Thessaloniki	16622000	40.60	23.00		Jan 1931	Dec 1997	804
48		prec_31	Thessaloniki	16622002	40.50	22.90		Jan 1961	Dec 1970	120
49		prec_32	Trikala	16645001	39.60	21.80		Jan 1951	Dec 1990	480
50		prec_33	Tripoli	16710000	37.50	22.40		Jan 1951	Dec 1985	420

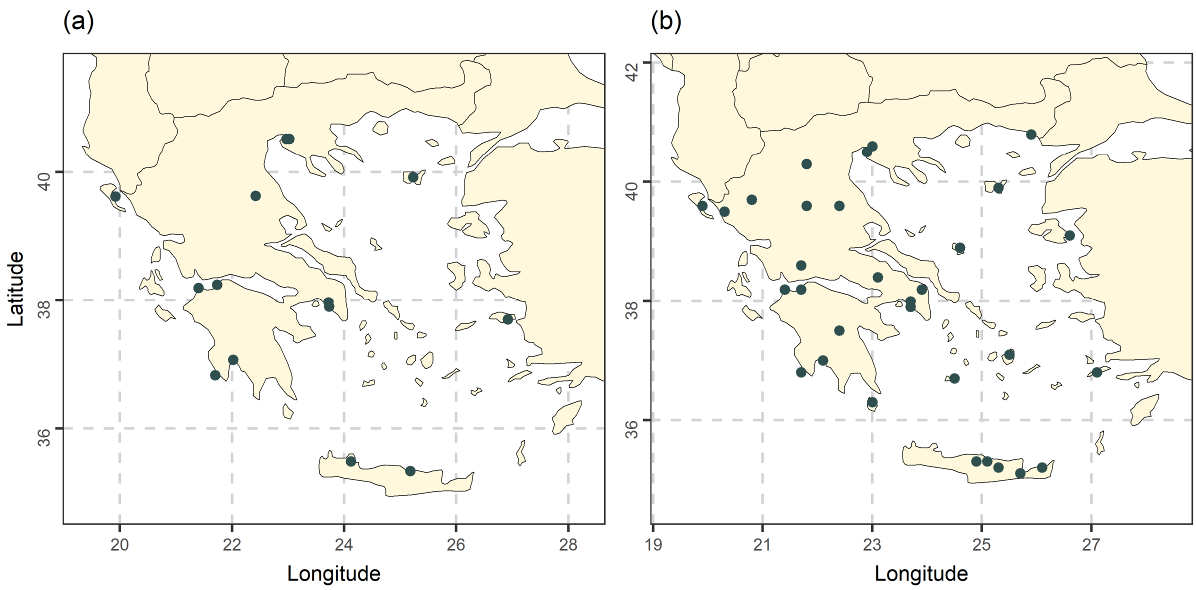


Figure 1. Maps of the locations of the (a) temperature and (b) precipitation stations; their sources are Lawrimore et al. (2011), and Peterson and Vose (1997) respectively.

We also apply the time series decomposition models to the entire time series to deseasonalize them. We then estimate the mean (μ), σ and H parameters of the Hurst-Kolmogorov process for each of the seasonally decomposed entire time series using the maximum likelihood estimator (Tyralis and Koutsoyiannis 2011) implemented via the `HKprocess` R package (Tyralis 2016). We further estimate the coefficient of variation (cv), which is defined by Equation 1. The μ, σ, cv and H estimates are presented in Tables 2 and 3. The Hurst parameter is assumed to be informative about the magnitude of long-range dependence observed in geophysical time series.

$$\text{cv} := \sigma/\mu \quad (1)$$

Table 2. Mean (μ), standard deviation (σ), coefficient of variation (cv) and Hurst parameter (H) estimates for the deseasonalized temperature time series.

Time series	μ estimate (°C)	σ estimate (°C)	cv estimate	H estimate
temp_1	17.95	1.25	0.07	0.66
temp_2	17.86	1.93	0.11	0.67
temp_3	18.51	1.81	0.10	0.68
temp_4	18.70	1.62	0.09	0.65
temp_5	18.97	1.18	0.06	0.69
temp_6	17.90	1.42	0.08	0.74
temp_7	17.75	1.47	0.08	0.67
temp_8	15.91	2.75	0.17	0.64
temp_9	16.36	2.11	0.13	0.74
temp_10	18.24	1.07	0.06	0.59
temp_11	17.83	1.20	0.07	0.61
temp_12	17.71	1.41	0.08	0.69
temp_13	18.21	1.46	0.08	0.64
temp_14	18.38	1.64	0.09	0.64
temp_15	18.63	1.47	0.08	0.71
temp_16	16.21	2.59	0.16	0.67
temp_17	16.13	2.16	0.13	0.48

Table 3. Mean (μ), standard deviation (σ), coefficient of variation (cv) and Hurst parameter (H) estimates for the deseasonalized precipitation time series.

Time series	μ estimate (mm)	σ estimate (mm)	cv estimate	H estimate
prec_1	81.09	56.61	0.70	0.47
prec_2	46.50	37.30	0.80	0.56
prec_3	55.52	42.14	0.76	0.53
prec_4	93.61	78.01	0.83	0.57
prec_5	95.62	74.42	0.78	0.48
prec_6	57.59	43.65	0.76	0.54
prec_7	33.44	30.45	0.91	0.56
prec_8	32.79	29.44	0.90	0.53
prec_9	29.65	27.87	0.94	0.53
prec_10	47.30	37.03	0.78	0.53
prec_11	40.02	35.27	0.88	0.50
prec_12	88.81	66.22	0.75	0.56
prec_13	94.36	60.85	0.64	0.57
prec_14	66.19	45.58	0.69	0.46
prec_15	42.12	35.65	0.85	0.50
prec_16	60.14	47.45	0.79	0.52
prec_17	92.53	65.00	0.70	0.56
prec_18	47.10	39.39	0.84	0.52
prec_19	58.63	53.36	0.91	0.57
prec_20	43.94	32.23	0.73	0.54
prec_21	36.46	30.90	0.85	0.54
prec_22	40.84	36.72	0.90	0.55
prec_23	60.59	44.00	0.73	0.50
prec_24	35.08	32.84	0.94	0.47
prec_25	56.00	49.39	0.88	0.51
prec_26	27.61	22.43	0.81	0.53
prec_27	60.23	44.64	0.74	0.52
prec_28	40.39	35.38	0.88	0.46
prec_29	38.55	32.86	0.85	0.56
prec_30	37.15	27.98	0.75	0.54
prec_31	35.24	24.94	0.71	0.55
prec_32	62.91	47.51	0.76	0.61
prec_33	68.45	44.77	0.65	0.47

2.3 Forecasting algorithms and methods

We focus on two ML forecasting algorithms, i.e. NN and SVM. The NN algorithm is the `mlp` algorithm of the `nnet` R package (Venables and Ripley 2002), while the SVM algorithm is the `ksvm` algorithm of the `kernlab` R package (Karatzoglou et al. 2004). These algorithms implement a single-hidden layer Multilayer Perceptron (MLP), and the Radial Basis kernel "Gaussian" function with C = 1 and epsilon = 0.1 respectively. Their application is made using the `CasesSeries`, `fit` and `lforecast` functions of the `rminer` R package (Cortez 2010, 2016). We also include four classical algorithms, i.e. the Autoregressive order one model (AR(1)), an algorithm from the family of Autoregressive Fractionally Integrated Moving Average models (auto_ARFIMA), the exponential smoothing state space algorithm with Box-Cox transformation, ARMA errors, Trend and

Seasonal Components (BATS) and the Theta algorithm, and a naïve benchmark in the comparisons. The latter sets each monthly forecast equal to its corresponding last year's monthly value. We apply the classical algorithms using the `forecast` R package (Hyndman and Khandakar 2008; Hyndman et al. 2017) and, specifically, five functions included in the latter, namely the `Arima`, `arfima`, `bats`, `forecast` and `thetaf` functions. The auto_ARFIMA algorithm applies the Akaike Information Criterion with a correction for finite sample sizes (AICc) for the estimation of the p, d, q values of the ARFIMA(p,d,q) model, while both the AR(1) and auto_ARFIMA algorithms implement the maximum likelihood method for the estimation of the ARMA parameters. The auto_ARFIMA algorithm considers the long-range dependence observed in the time series through the d parameter. The AR(1), auto_ARFIMA and BATS algorithms apply Box-Cox transformation to the input data before fitting a model to them. All the algorithms used herein are well-grounded in the literature; thus, in their presentation we place emphasis on implementation information.

While the classical methods are simply defined by the classical algorithm, the ML methods are defined by the set {ML algorithm, hyperparameter selection procedure, lags}. We compare 21 regression matrices, each using the first n time lags, n = 1, 2, …, 21, and two procedures for hyperparameter selection, i.e. predefined hyperparameters (default values of the algorithms) or defined after optimization. The symbol * in the name of a ML method is hereafter used to denote that the model's hyperparameters have been optimized. The hyperparameter optimization is performed with the grid search method using a single validation set (last 1/3 of the deseasonalized fitting set). The hyperparameters optimized are the number of hidden nodes and the number of variables randomly sampled as candidates at each split of the NN and SVM models respectively. For the NN* method the hyperparameter optimization procedure is described subsequently. First, we fit 16 different NN models (defined by the grid values 0, …, 15) to the fist 2/3 of the deseasonalized fitting set. Second, we use these models to produce forecasts corresponding to the validation set. Third, we select the one exhibiting the smallest root mean square error (RMSE) on the validation set. To produce the forecast corresponding to the test set we further fit the selected model to the whole deseasonalized fitting set. For the SVM* method the procedure is the same, except that the candidate models are five (defined by the grid values 1, … 5). Hereafter, we consider that the ML models are used with predefined hyperparameters and that the regression matrix is built using only the first lag, unless mentioned differently. We use the sets of methods defined in Table 4. Each of them has a specific utility within our experiments, which is also reported in Table 4. A secondary utility of set of methods no 5 is the investigation of the existence of a possible relationship between the forecast quality and the parameter estimates for the deseasonalized time series.

Table 4. Sets of methods and their main utility within this study.

s/n	Set of methods	Number of included methods	Main utility
1	{NN given a regression matrix formed using the first n lags, n = 1, 2, …, 21}	21	Exploration of Problem 1 for the NN algorithm
2	{SVM given a regression matrix formed using the first n lags, n = 1, 2, …, 21}	21	Exploration of Problem 1 for the SVM algorithm
3	{NN, NN*}	2	Exploration of Problem 2 for the NN algorithm
4	{SVM, SVM*}	2	Exploration of Problem 2 for the SVM algorithm
5	{Naïve, AR(1), auto_ARFIMA, BATS, Theta, NN, SVM}	7	Exploration of Problem 3 for the NN and SVM algorithms

2.4 Metrics and summary statistics

The one-step ahead forecasting performance is assessed by computing the absolute error (AE) of the forecast, while the multi-step ahead forecasting performance by computing the RMSE, the Nash-Sutcliffe efficiency (NSE), the ratio of standard deviations (rSD), the index of agreement (d) and the coefficient of correlation (Pr). Subsequently, we provide the definitions of the five latter metrics. For these definitions we consider a time series of N values. Let us also consider a model fitted to the first $N - n$ values of this specific time series and subsequently used to make predictions corresponding to the last n values. Let $x_1, x_2, \ldots, x_n$ represent the last n values and $f_1, f_2, \ldots, f_n$ represent the forecasts.

The RMSE metric is defined by

$$\text{RMSE} := \left(\left(\textstyle\sum_{i=1}^{n}(f_i - x_i)^2\right)/n\right)^{1/2} \quad (2)$$

It can take values between 0 and +∞. The closer to 0 it is, the better the forecast.

Let $\bar{x}$ be the mean of the observations, which is defined by

$$\bar{x} := (1/n) \textstyle\sum_{i=1}^{n} x_i \quad (3)$$

The NSE metric is defined by (Nash and Sutcliffe 1970)

$$\text{NSE} := 1 - \left(\textstyle\sum_{i=1}^{n}(f_i - x_i)^2 / \sum_{i=1}^{n}(x_i - \bar{x})^2\right) \quad (4)$$

It can take values between −∞ and 1. The closer to 1 it is, the better the forecast, while

NSE values above 0 indicate acceptable forecasts.

Let s_x be the standard deviation of the observations, which is defined by

$$s_x := ((1/(n-1)) \sum_{i=1}^{n} (x_i - \bar{x})^2)^{1/2} \quad (5)$$

Let $\bar{f}$ be the mean of the forecasts and s_f be the standard deviation of the forecasts, which are defined by Equations (6) and (7) respectively.

$$\bar{f} := (1/n) \sum_{i=1}^{n} f_i \quad (6)$$

$$s_f := ((1/(n-1)) \sum_{i=1}^{n} (f_i - \bar{f})^2)^{1/2} \quad (7)$$

The rSD metric is defined by (Zambrano-Bigiarini 2017a)

$$\text{rSD} := s_f / s_x \quad (8)$$

It can take values between 0 and +∞. The closer to 1 it is, the better the forecast.

The Pr metric is defined by (Krause et al. 2005)

$$\text{Pr} := (\sum_{i=1}^{n} (x_i - \bar{x}) (f_i - \bar{f})) / (\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (f_i - \bar{f})^2)^{1/2} \quad (9)$$

It can take values between −1 and 1. The closer to 1 it is, the better the forecast.

The d metric is defined by (Krause et al. 2005)

$$d := 1 - (\sum_{i=1}^{n} (f_i - x_i)^2 / \sum_{i=1}^{n} (|f_i - \bar{x}| + |x_i - \bar{x}|)^2) \quad (10)$$

It can take values between 0 and 1. The closer to 1 it is, the better the forecast.

To summarize the results of the multiple-case study we compute some summary statistics for the values of each metric, i.e. the minimum, median and maximum, separately for each algorithm. For the ML ones, these summary statistics are computed by aggregating the total of the values of each metric computed for methods that are based on each specific ML algorithm (tested for the exploration of Problems 1, 2 or 3). We also compute the linear regression coefficient (LRC) for each method per category of tests. This summary statistic can be used to measure the dependence of the forecasts f_j on their corresponding target values x_j, when this dependence is expressed by the following linear regression model:

$$f_j = (\text{LRC})\, x_j + b \quad (11)$$

It can take values between −∞ and +∞. The closer to 1 it is, the better the forecasts. The subscript j in the above notations indicates the serial number of each of the pairs {forecast, target value} formed for a specific category of tests.

3. Results and discussion

In Section 3 we present and discuss the results of our multiple-case study. We place emphasis on the qualitative presentation of the results, because of its importance in the exploration of the research questions of Section 1.2. Especially the heatmap visualization adopted herein allows the examination of each single-case study alone and in comparison to the rest simultaneously. Quantitative information, derived by our multiple-case study and particularly significant for the case of Greece, is also presented. Regarding this type of information, the present study could be viewed as an expansion of Moustris et al. (2011). The latter study has focused on four long precipitation time series observed in Alexandroupoli, Athens, Patra and Thessaloniki (a subset of the time series examined within our multiple-case study), with the aim to present forecasts for the monthly maximum, minimum, mean and cumulative precipitation totals using NN methods.

3.1 Exploration of Problem 1

Section 3.1 is devoted to the exploration of Problem 1. In Figures 2 and 3 we visualize the one and twelve-step ahead temperature forecasts respectively, produced for this exploration for the NN and SVM algorithms, in comparison to their corresponding target values. We observe that, for a specific target value, the forecasts are more scattered (in the vertical direction) for the NN algorithm than they are for the SVM algorithm. This fact indicates that the performance of the SVM algorithm is affected less than the performance of the NN algorithm by changes in the lagged regression matrix used in the fitting process. The effect under discussion may result in more or less accurate NN forecasts (laying closer or farther from the 1:1 line included in the scatterplots of Figures 2 and 3) than the ones produced by the SVM algorithm. Evidence that the NN algorithm is more prone to changes in the regression matrix than the SVM one is provided by the tests conducted using the precipitation time series as well. In Figure 4 we present the twelve-step ahead precipitation forecasts in comparison to their corresponding target values.

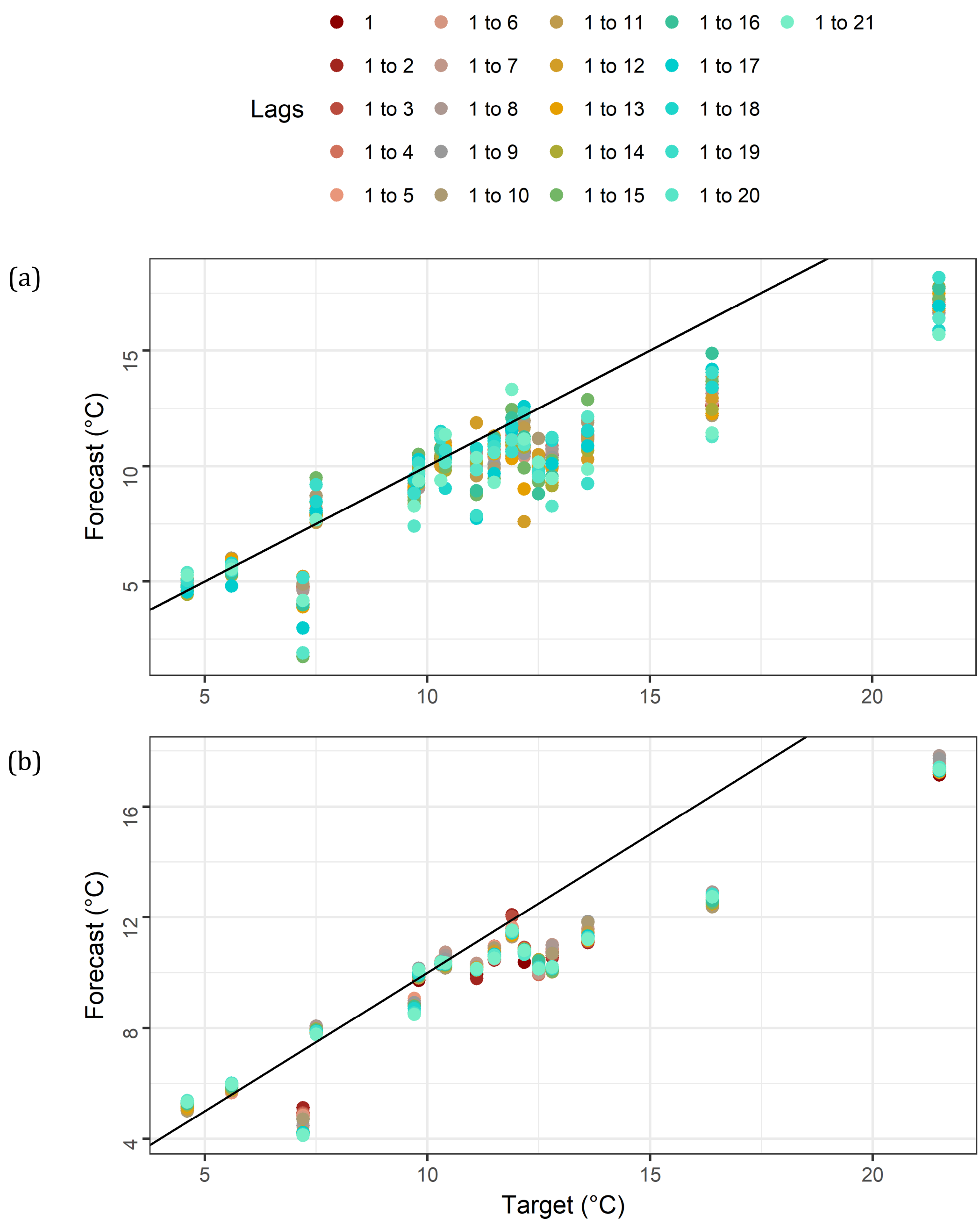


Figure 2. One-step ahead temperature forecasts, produced for the exploration of Problem 1 for the (a) NN and (b) SVM algorithms, in comparison to their corresponding target values.

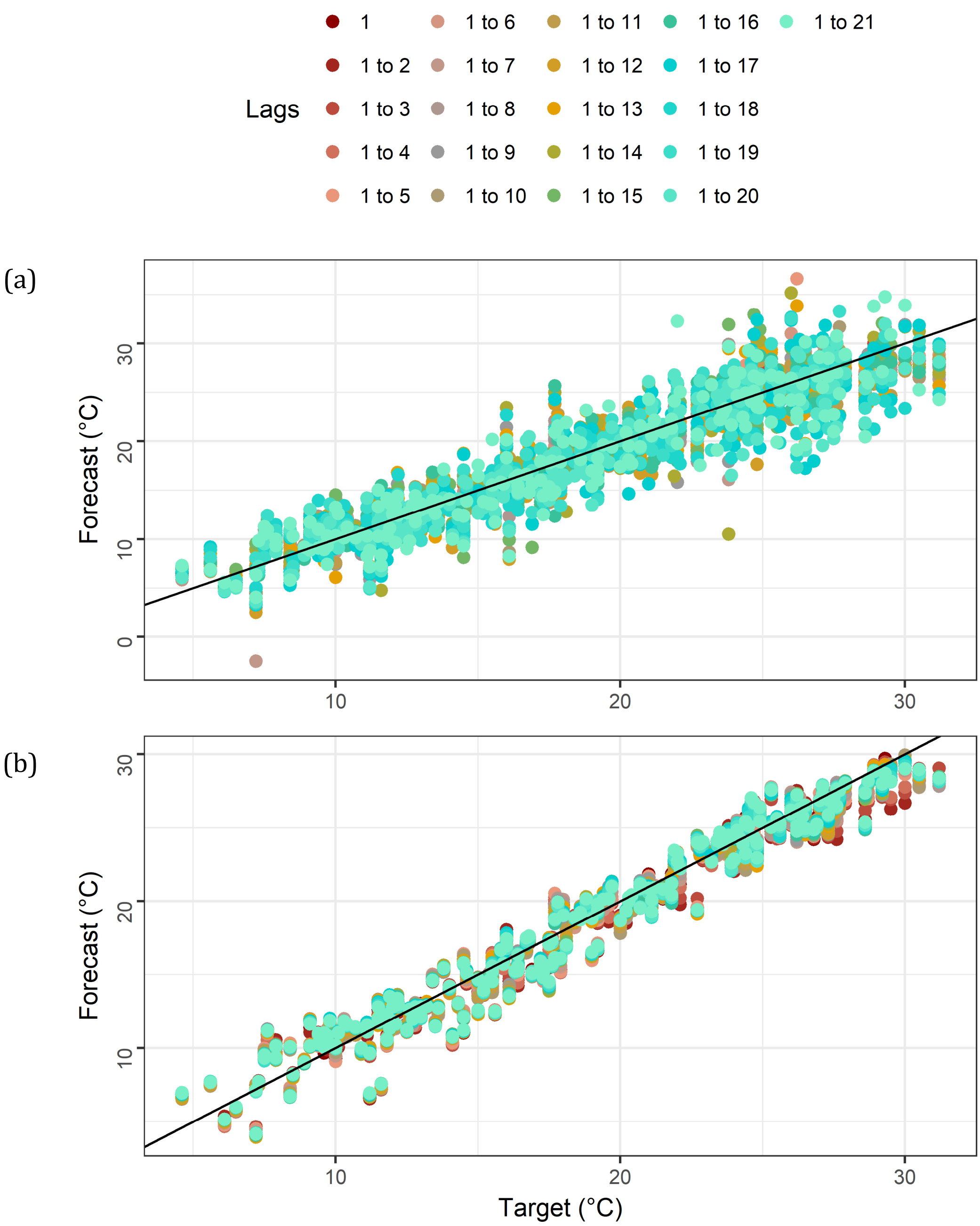


Figure 3. Twelve-step ahead temperature forecasts, produced for the exploration of Problem 1 for the (a) NN and (b) SVM algorithms, in comparison to their corresponding target values.

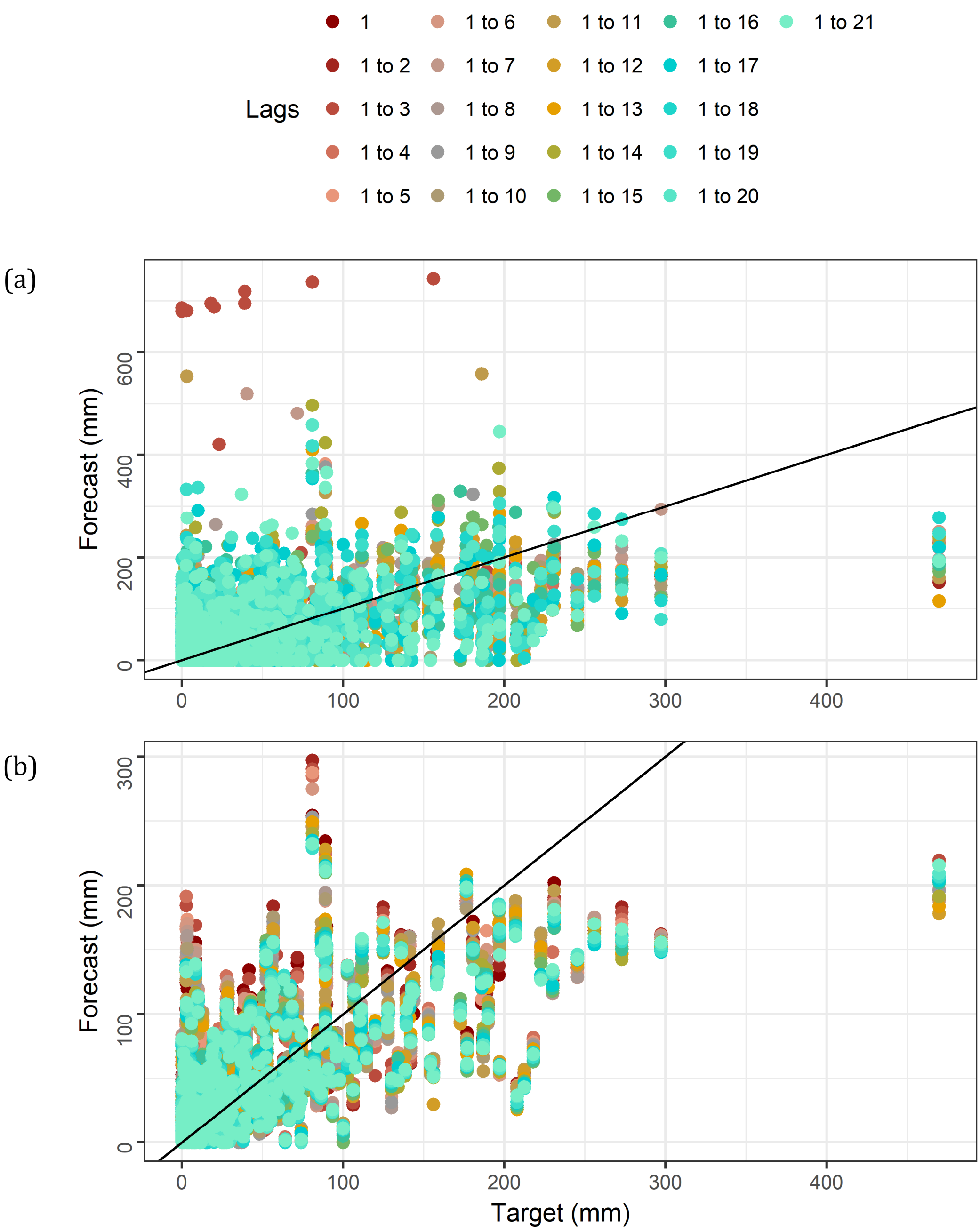


Figure 4. Twelve-step ahead precipitation forecasts, produced for the exploration of Problem 1 for the (a) NN and (b) SVM algorithms, in comparison to their corresponding target values.

More importantly, in Figures 5 and 6 we comparatevely present the AE, RMSE, NSE and d values computed for the temperature forecasts, produced for the exploration of Problem 1 for the NN and SVM algorithms, for each individual case examined. By the examination of these two figures we observe the following:

(a) There are variations in the results across the individual cases, to an extent that it is impossible to decide on a best or worst method. Therefore, no evidence is provided by the respective categories of tests that any of the compared lagged regression matrices systematically leads to better forecasts than the rest, either for the NN or the SVM algorithms.

(b) The heatmaps formed for the SVM algorithm are smoother in the row direcion than those formed for the NN algorithm, a fact rather expected from Figures 2 and 3. In other words, the variations within each single-case study are of small magnitude for the case of the SVM algorithm, while they are significant for the NN algorithm.

(c) For the SVM algorithm there are no systematic patterns and the small variations seem to be rather random.

(d) For the NN algorithm and especially for the twelve-step ahead forecasts the left parts of the heatmaps are smoother with no white cells. Alternatively worded, it seems that is is more likely that the forecasts are better when using less recent lagged variables in conjuction with this algorithm.

Observation (a) is particularly important, because it reveals that the forecast quality is subject to limitations. Each forecasting method has some specific theoretical properties and, due to the latter, it performs better or worse than other forecasting methods, depending on the case examined. Even forecasting methods based on the same algorithm can produce forecasts with very different quality, as indicated by the results obtained for the NN algorithm. Observation (d), on the other hand, provides some interesting evidence, which however is contigent and, therefore, should be further investigated within larger forecast-comparing studies, such as Tyralis and Papacharalampous (2017). Furthermore, in Figure 7 we present the AE and RMSE values computed for the precipitation forecasts, produced for the exploration of Problem 1 for the NN and SVM algorithms, within each single-case study. Observations (a) and (b) apply here as well. Moreover, both the ML algorithms, seem to perform rather better, to a small extent though, when given a lagged regression matrix using less recent lags.

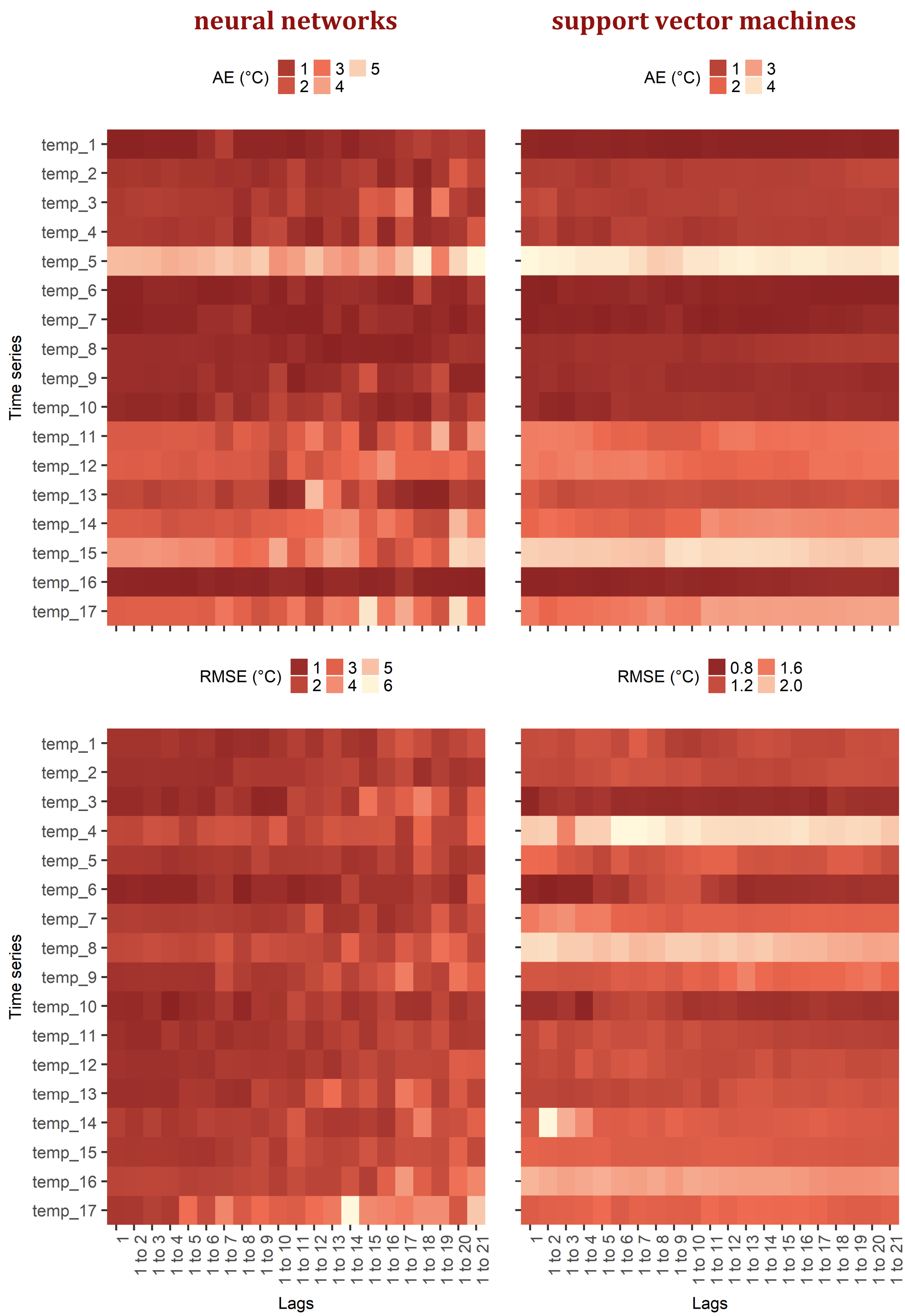


Figure 5. Cross-case synthesis for the exploration of Problem 1 for the NN and SVM algorithms using the temperature time series (part 1).

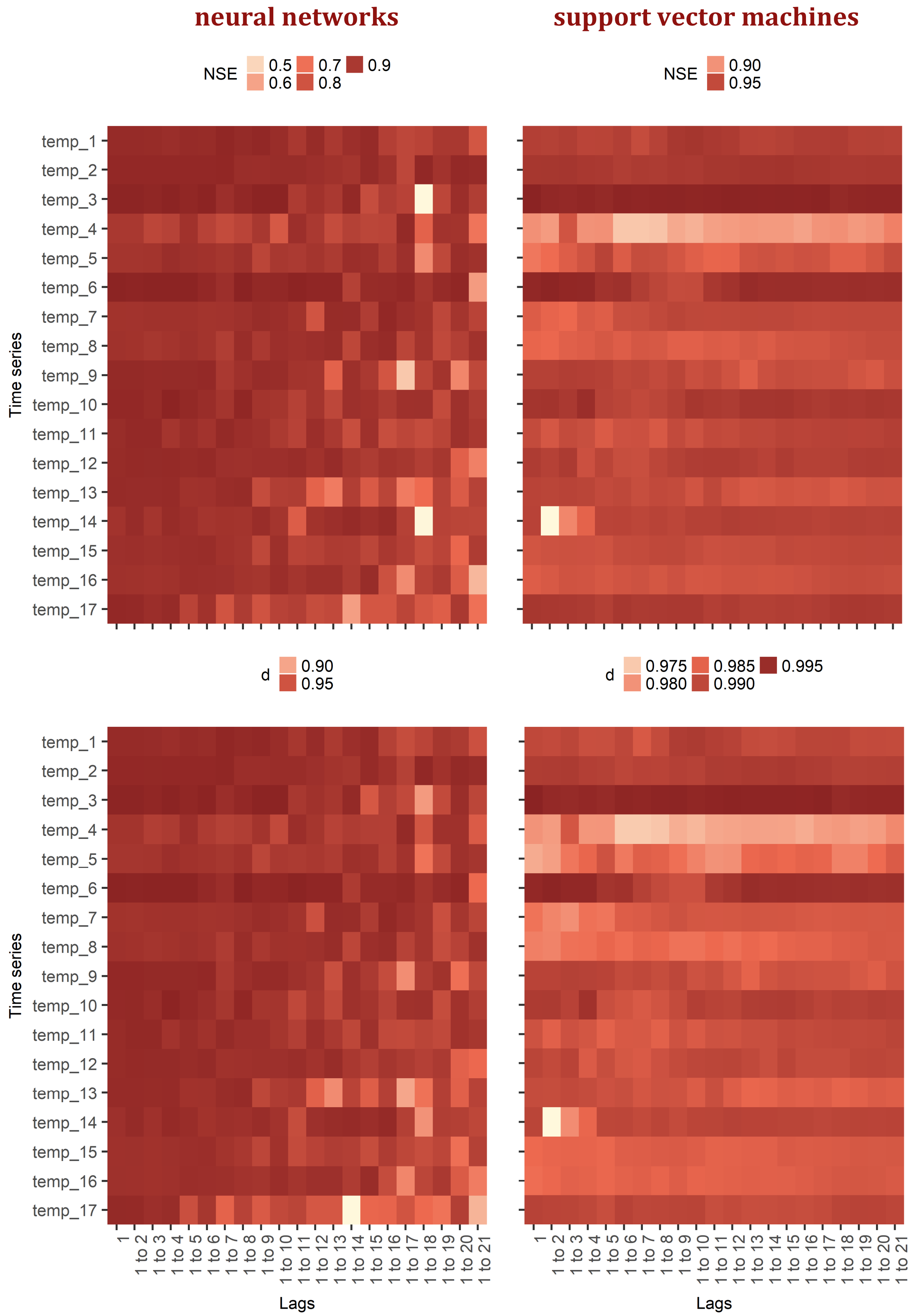


Figure 6. Cross-case synthesis for the exploration of Problem 1 for the NN and SVM algorithms using the temperature time series (part 2).

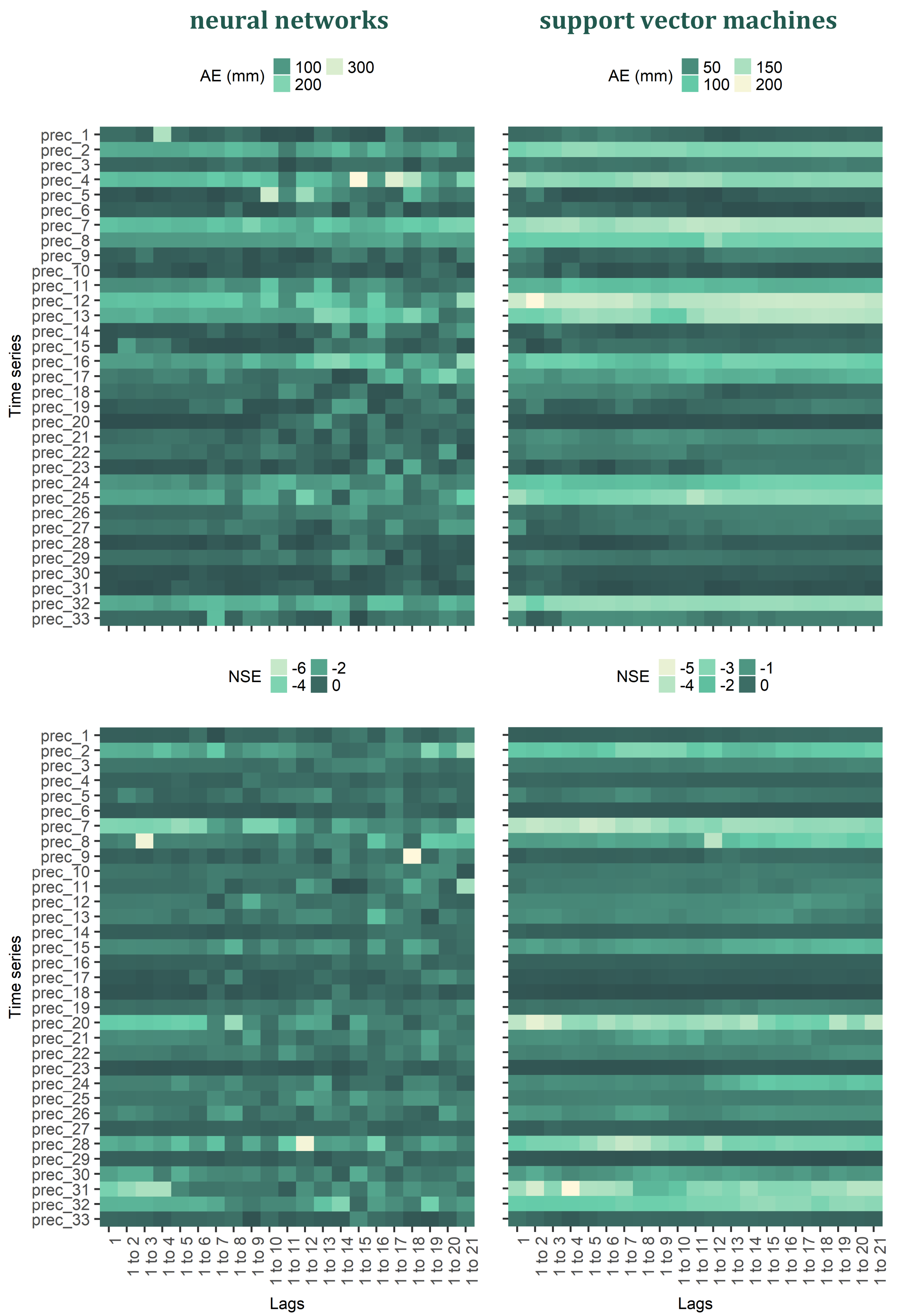


Figure 7. Cross-case synthesis for the exploration of Problem 1 for the NN and SVM algorithms using the precipitation time series.

3.2 Exploration of Problem 2

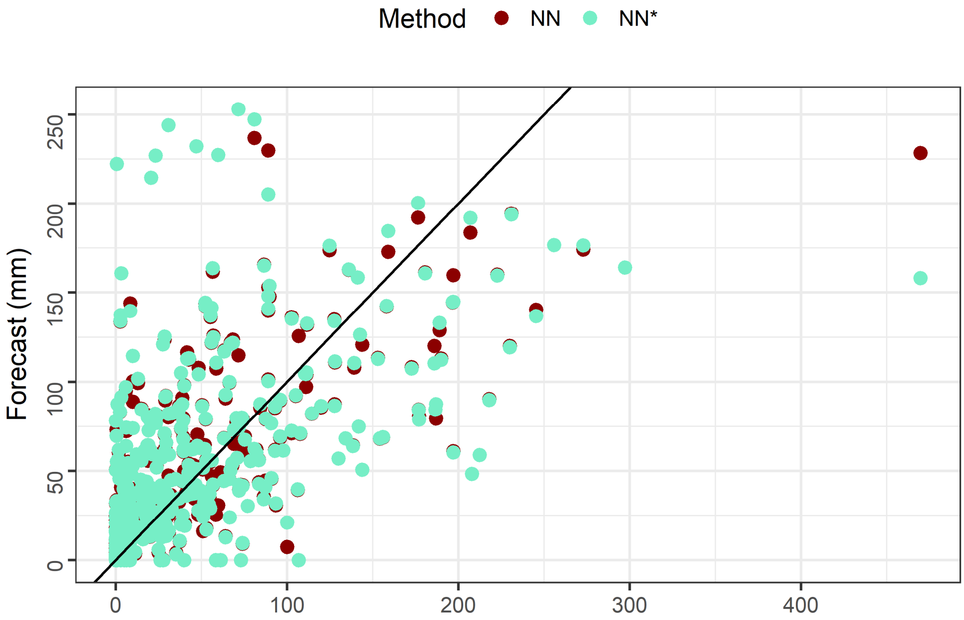
Section 3.2 is devoted to the exploration of Problem 2. In Figure 8 we present the twelve-step ahead precipitation forecasts, produced for this exploration for the NN and SVM algorithms, in comparison to their corresponding target values. Figure 8 could be studied alongside with Figure 4, providing contingent evidence that hyperparameter optimization affects less the performance of these two ML algorithms than lagged variable selection does. The latter observation applies more to the NN algorithm. Furthermore, in Figure 9 we comparatively present the AE, RMSE, rSD and d values computed for the one- and twelve-step ahead temperature forecasts, produced for the exploration of Problem 2, within each single-case study. By the examination of Figure 9 we observe the following:

(a) Here as well, none of the compared methods seems to be systematically better across the individual cases examined. In other words, the results do not systematically favour any of the two tested hyperparameter selection procedures and, therefore, we can state that hyperparameter optimization does not necessarily lead to better forecasts than the use of the default values of the algorithms.

(b) For both the ML algorithms the observed variations within each of the single-case studies are of smaller magnitude for the one-step ahead forecasts than they are for the twelve-step ahead ones.

(c) For the case of the NN algorithm the twelve-step ahead forecasts seem to be rather better when hyperparameter optimization precedes the fitting process, while the opposite applies to the case of the SVM algorithm.

Finally, in Figure 10 we present the AE, NSE, rSD and d values computed for the one- and twelve-step ahead precipitation forecasts, produced for the exploration of Problem 2, within each single-case study. Observation (a) also applies to the precipitation forecasts, while the variations can be significant for both the one- and twelve-step ahead forecasts. For the latter it seems that hyperparameter optimization mostly leads to less accurate forecasts. This may be explained by the fact that the default values of the algorithms are usually set based on tests performed by their developers or in the scientific literature, so that the performance of the algorithms is mostly maximized for a variety of problems.



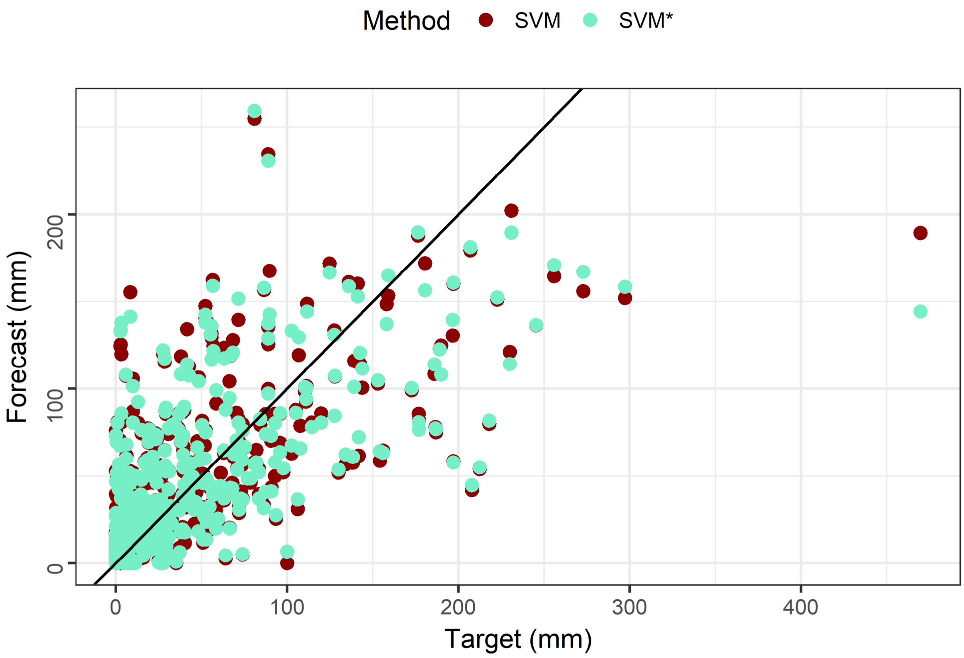


Figure 8. Twelve-step ahead precipitation forecasts, produced for the exploration of Problem 2 for the NN and SVM algorithms, in comparison to their corresponding target values.

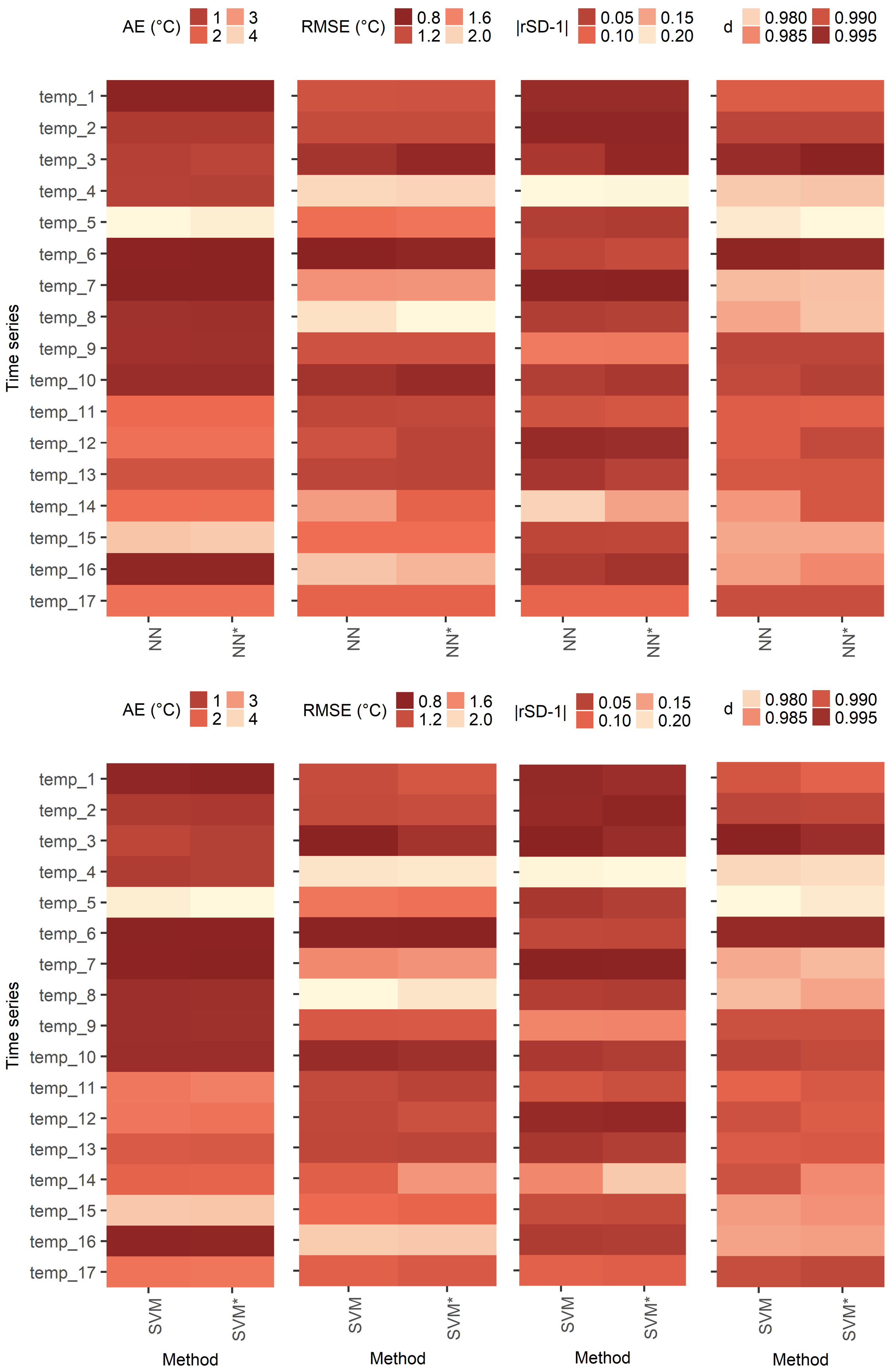


Figure 9. Cross-case synthesis for the exploration of Problem 2 for the NN and SVM algorithms using the temperature time series.

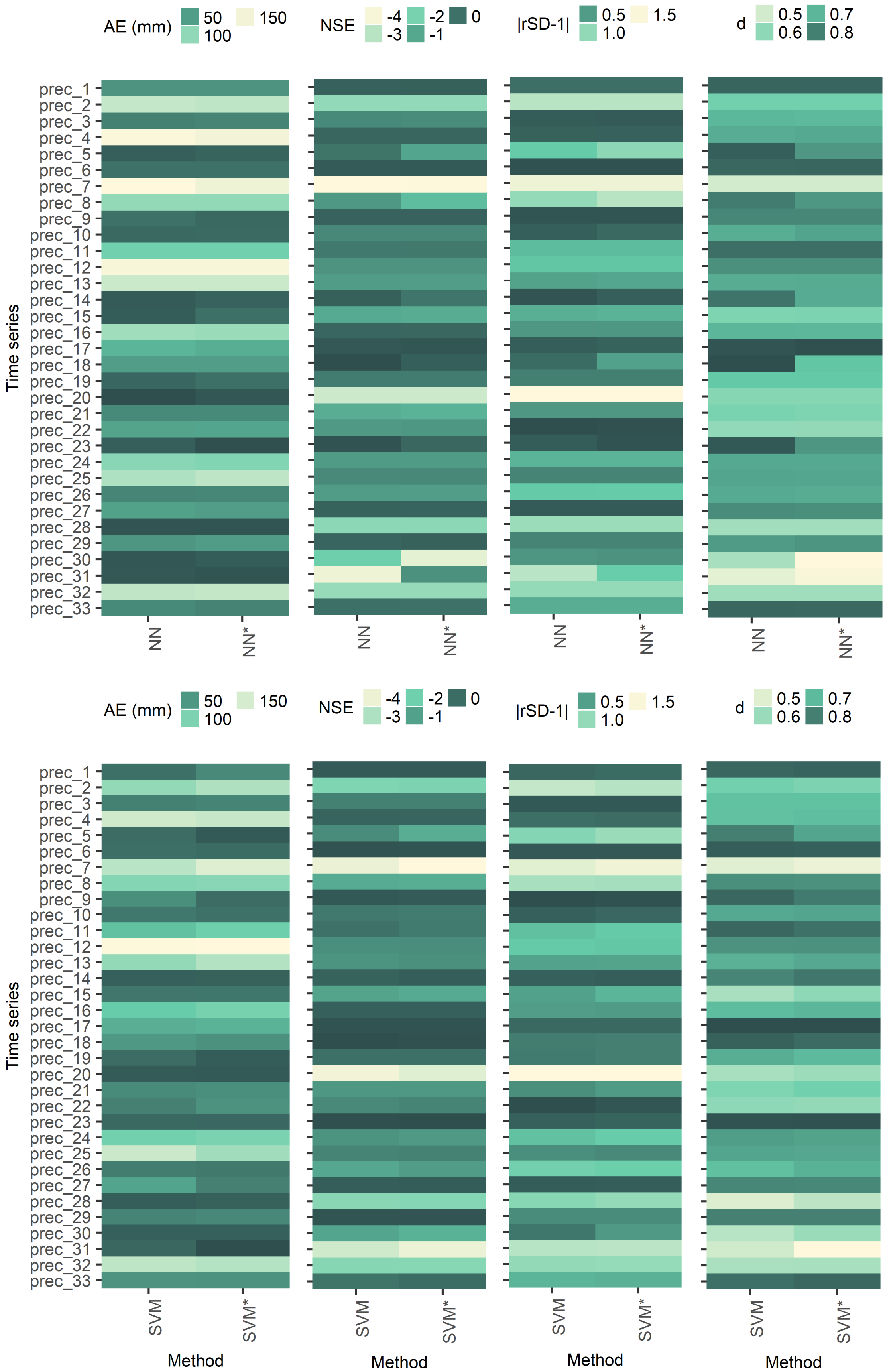


Figure 10. Cross-case synthesis for the exploration of Problem 2 for the NN and SVM algorithms using the precipitation time series.

3.3 Exploration of Problem 3

Section 3.3 is devoted to the exploration of Problem 3. In Figure 11 we present the one- and twelve-step ahead temperature forecasts, produced for this exploration, in comparison to their corresponding target values, while in Figure 12 we present an analogous visualization for the precipitation forecasts serving the same purpose. Moreover, in Figures 13 and 14 we comparatively present all the metric values computed for the temperature forecasts and the AE, RMSE and d values computed for the precipitation forecasts respectively within each single-case study. By the examination of these four figures we observe the following:

(a) Here as well, the results of the single-case studies vary significantly.

(b) The best method within a specific single-case study depends on the criterion of interest. In fact, even within a specific single-case study, we cannot decide on one best (or worst) method regarding all the criteria set simultaneously.

(c) Observations (a) and (b) apply equally to the ML and the classical methods. In fact, it seems that both categories can rarther perform equally well, under the same limitations.

(d) We observe that the Naïve benchmark, competent as well, frequently produces far different forecasts than those produced by the ML or classical algorithms.

If we further compare Figures 11(a), 11(b) and 12 with Figures 2, 3 and 4 respectively, we observe that the performance of the NN algorithm (when given the 21 regression matrices examined in the present study) can vary more than the performance of the here compared ML and classical methods. This observation does not apply to the case of the SVM algorithm. Finally, we note that the exploration presented in Section 3.3 and Papacharalampous et al. (2017a) effectively complement each other. In fact, the former illustrates and provides evidence on important points by presenting real-world results, while the latter confirms the evidence derived by the former by conducting simulation experiments of large scale. Both illustration and confirmation are integral parts of every theory-building process.

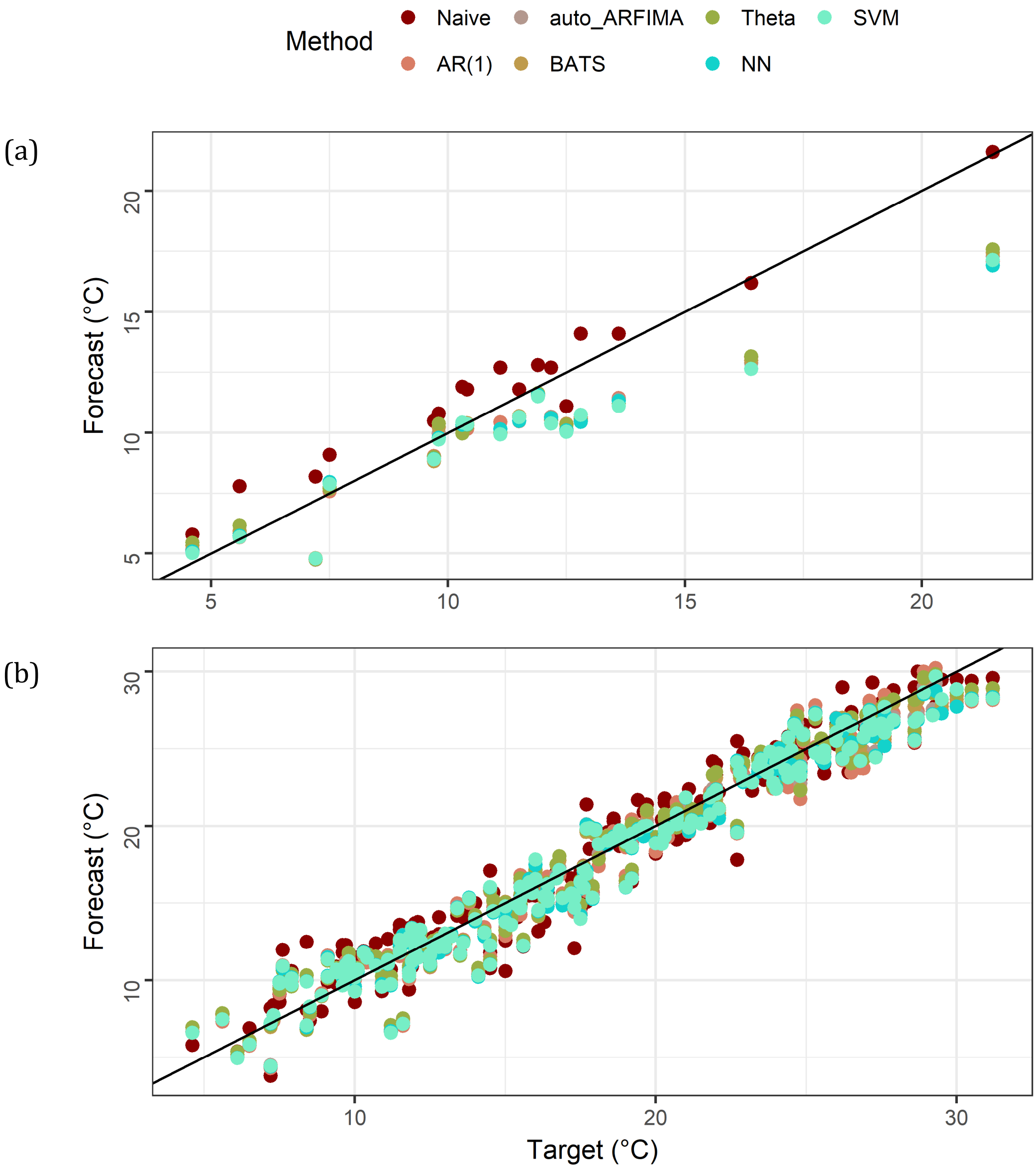


Figure 11. (a) One- and (b) twelve-step ahead temperature forecasts, produced for the exploration of Problem 3, in comparison to their corresponding target values.

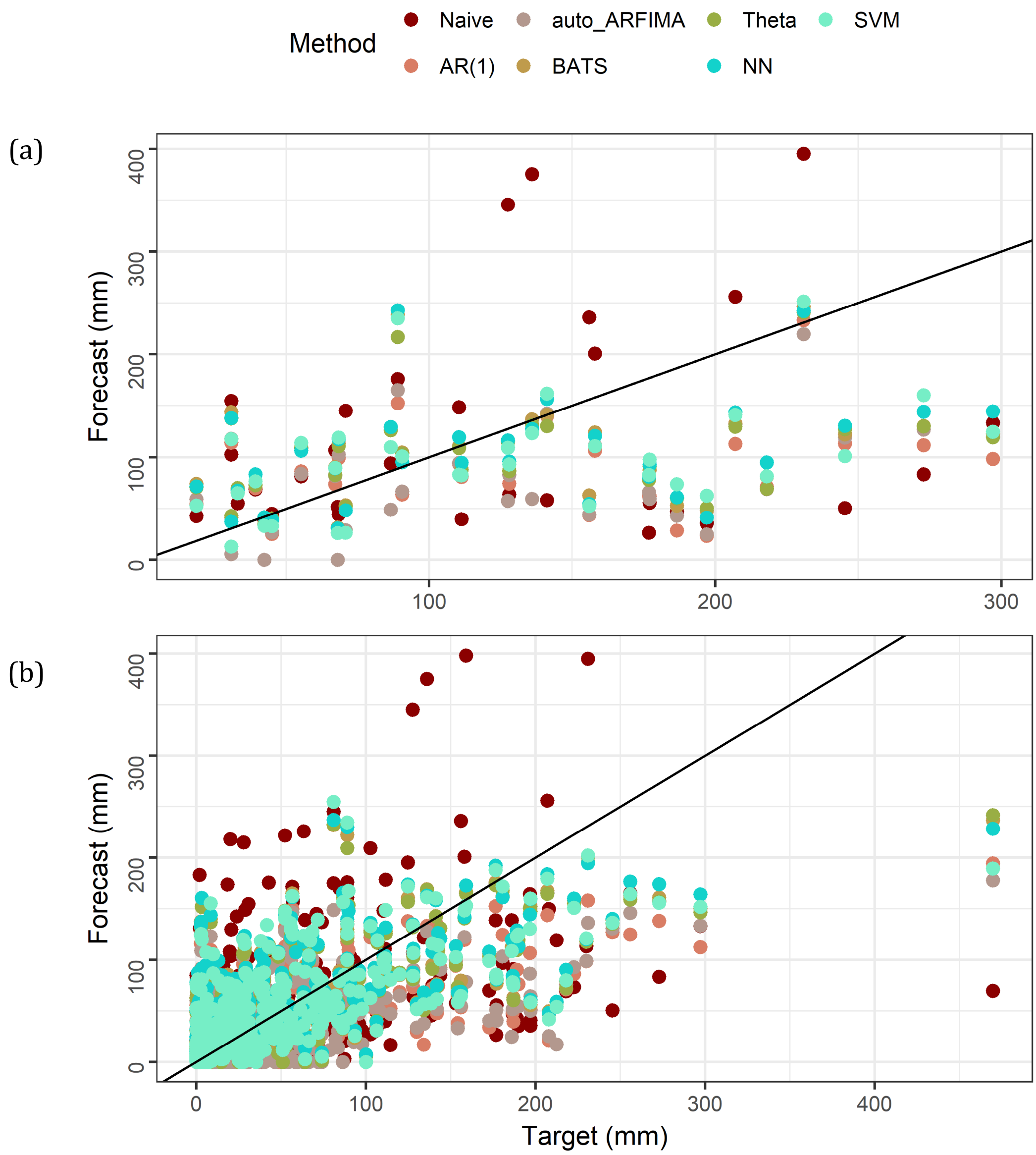


Figure 12. (a) One- and (b) twelve-step ahead precipitation forecasts, produced for the exploration of Problem 3, in comparison to their corresponding target values.

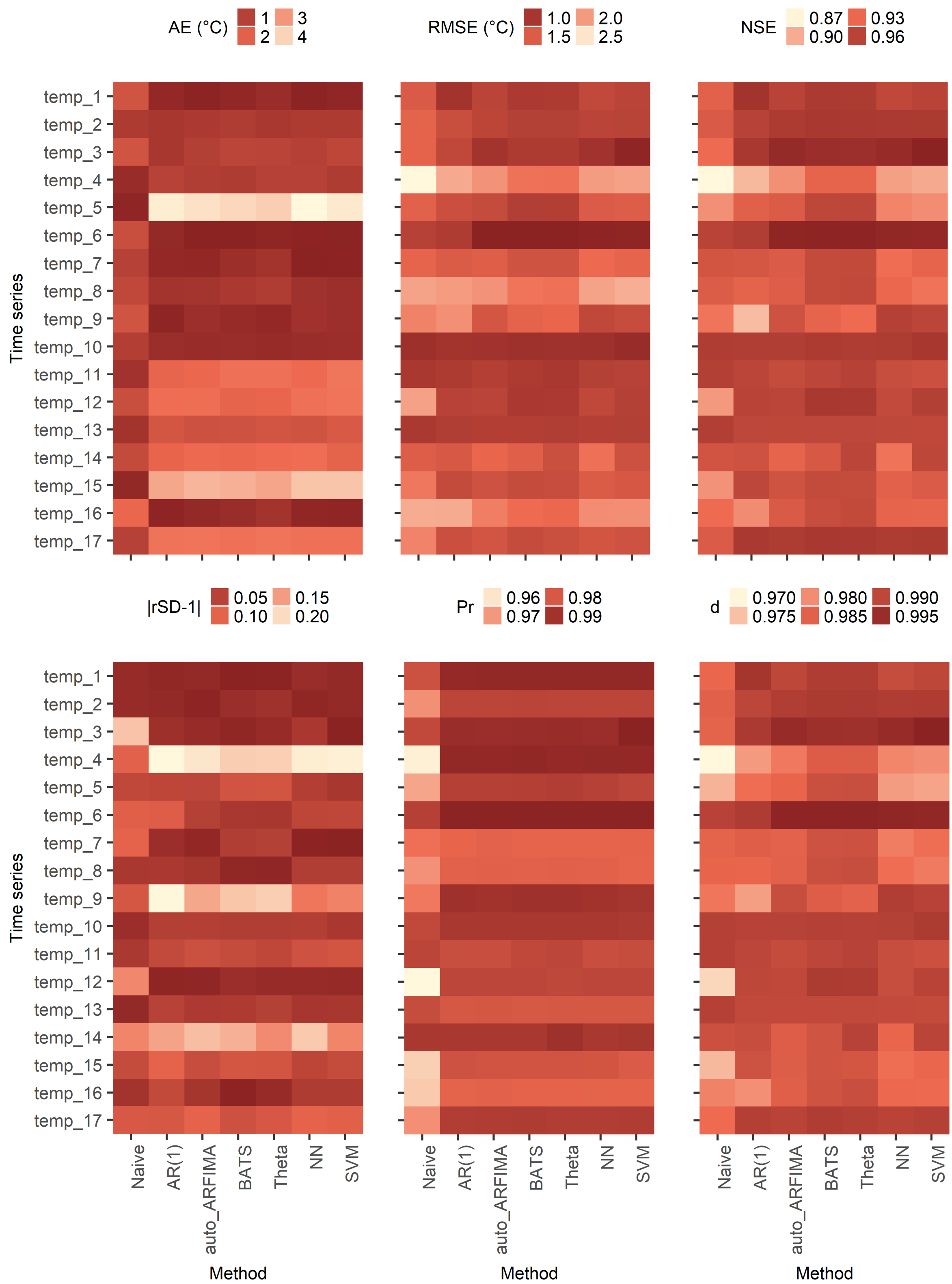


Figure 13. Cross-case synthesis for the exploration of Problem 3 using the temperature time series.

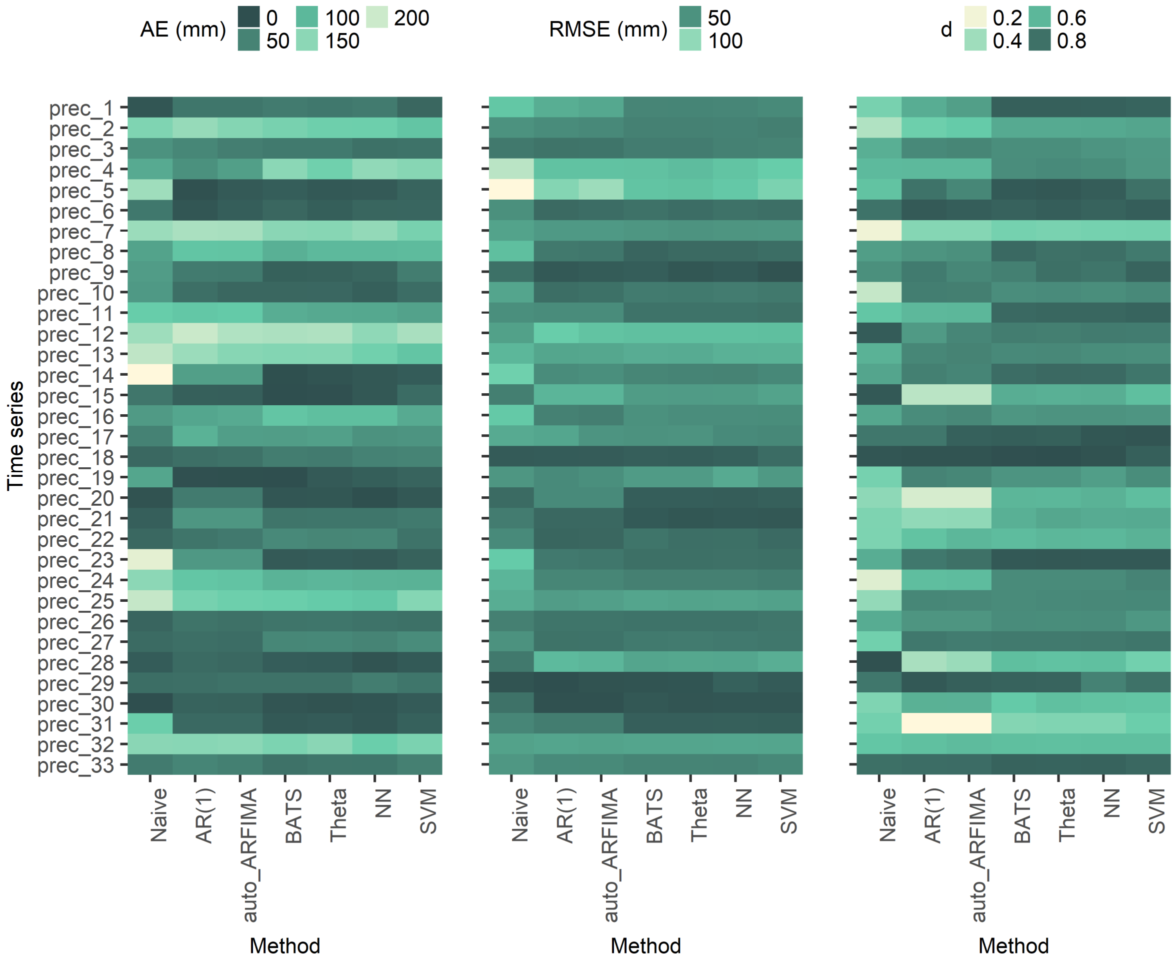


Figure 14. Cross-case synthesis for the exploration of Problem 3 using the precipitation time series.

3.4 Additional information

Section 3.4 is devoted to some additional worth-discussed information derived by our multiple-case study. In fact, the results produced mainly for the exploration of Problems 1, 2 and 3 can also be examined from different points of view, which are considered of secondary importance within this study. In Tables 5 and 6 we present the summary statistics of the metric values, separately for each algorithm, and in Table 7 the LRC values for each category of tests. This information stands as a summary of the quantitative information provided by our multiple-case study and, together with Figures 2-14, can facilitate the below discussion in a satisfactory manner. Regarding an overall assessment of the algorithms, they are all found to mostly have a better average-case forecasting performance than the Naïve benchmark, with the NN algorithm being the worst. This is due to the reported high effect of the lagged regression matrix on the performance of this algorithm. On the contrary, the SVM algorithm has a better average-case performance, (almost) as good as the one of the best-performing classical algorithms, i.e. BATS, Theta and auto_ARFIMA.

Table 5. Summary statistics of the metric values computed for the temperature forecasts. The values reported for the NN and SVM algorithms are computed for the total of the NN and SVM methods implemented in this study respectively.

Metric	Algorithm	Summary statistic		
		Minimum	Median	Maximum
AE (°C)	Naïve	0.10	1.00	2.20
	AR(1)	0.08	0.66	4.41
	auto_ARFIMA	0.02	0.88	4.22
	BATS	0.00	0.86	4.07
	Theta	0.11	1.00	3.92
	NN	0.00	0.98	5.79
	SVM	0.01	0.90	4.52
RMSE (°C)	Naïve	0.92	1.60	2.62
	AR(1)	0.96	1.32	2.12
	auto_ARFIMA	0.74	1.28	1.95
	BATS	0.74	1.14	1.75
	Theta	0.74	1.14	1.73
	NN	0.63	1.70	6.05
	SVM	0.73	1.31	2.30
NSE	Naïve	0.87	0.94	0.97
	AR(1)	0.89	0.96	0.97
	auto_ARFIMA	0.91	0.95	0.98
	BATS	0.93	0.96	0.99
	Theta	0.93	0.96	0.99
	NN	0.44	0.93	0.99
	SVM	0.85	0.95	0.99
rSD	Naïve	0.87	1.01	1.18
	AR(1)	0.90	1.01	1.22
	auto_ARFIMA	0.90	1.01	1.21
	BATS	0.92	1.00	1.19
	Theta	0.92	0.99	1.19
	NN	0.89	1.01	1.24
	SVM	0.89	1.02	1.24
Pr	Naïve	0.96	0.97	0.99
	AR(1)	0.98	0.99	0.99
	auto_ARFIMA	0.98	0.99	0.99
	BATS	0.98	0.99	0.99
	Theta	0.98	0.99	0.99
	NN	0.79	0.98	1.00
	SVM	0.97	0.99	0.99
d	Naïve	0.97	0.98	0.99
	AR(1)	0.98	0.99	0.99
	auto_ARFIMA	0.98	0.99	1.00
	BATS	0.99	0.99	1.00
	Theta	0.98	0.99	1.00
	NN	0.86	0.98	1.00
	SVM	0.97	0.99	1.00

Table 6. Summary statistics of the metric values computed for the precipitation forecasts. The values reported for the NN and SVM algorithms are computed for the total of the NN and SVM methods implemented in this study respectively.

Metric	Algorithm	Summary statistic		
		Minimum	Median	Maximum
AE (mm)	Naïve	0	72	239
	AR(1)	2	52	199
	auto_ARFIMA	1	45	178
	BATS	0	41	175
	Theta	2	40	178
	NN	0	51	340
	SVM	0	39	206
RMSE (mm)	Naïve	17	52	147
	AR(1)	15	46	94
	auto_ARFIMA	16	45	105
	BATS	17	41	76
	Theta	18	41	75
	NN	17	47	588
	SVM	11	41	101
NSE	Naïve	−13.20	−0.21	0.48
	AR(1)	−46.17	−0.90	0.64
	auto_ARFIMA	−46.17	−1.01	0.61
	BATS	−4.46	−0.35	0.69
	Theta	−5.07	−0.30	0.70
	NN	−7.55	−0.42	0.86
	SVM	−5.44	−0.44	0.76
rSD	Naïve	0.35	1.05	3.59
	AR(1)	0.55	1.60	4.10
	auto_ARFIMA	0.56	1.55	4.10
	BATS	0.53	1.47	2.53
	Theta	0.53	1.46	2.71
	NN	0.19	1.10	2.60
	SVM	0.48	1.38	2.71
Pr	Naïve	−0.09	0.46	0.93
	AR(1)	0.09	0.62	0.92
	auto_ARFIMA	0.09	0.62	0.93
	BATS	0.21	0.60	0.91
	Theta	0.24	0.60	0.91
	NN	−0.74	0.54	0.96
	SVM	−0.37	0.62	0.92
d	Naïve	0.20	0.59	0.89
	AR(1)	0.17	0.70	0.89
	auto_ARFIMA	0.17	0.73	0.89
	BATS	0.46	0.73	0.90
	Theta	0.47	0.73	0.90
	NN	0.01	0.67	0.97
	SVM	0.25	0.71	0.93

Table 7. LRC values computed for each category of tests.

Set of methods (see Table 4)	Process	One-step ahead forecasts		Twelve-step ahead forecasts	
		Minimum	Maximum	Minimum	Maximum
1	Temperature	0.62	0.79	0.88	0.97
2		0.70	0.75	0.93	0.96
3		0.69	0.70	0.94	0.94
4		0.70	0.70	0.94	0.95
5		0.69	0.88	0.94	0.96
1	Precipitation	0.00	0.43	0.41	0.56
2		0.21	0.29	0.49	0.52
3		0.25	0.27	0.48	0.52
4		0.25	0.29	0.49	0.51
5		0.21	0.29	0.40	0.52

The reported values of the summary statistics, as well as Figures 2, 3, 4, 8, 11 and 12, reveal that the temperature forecasts are remarkably better than the precipitation ones. This may be explained by the cv estimates presented in Tables 2 and 3. Finally, in Figure 15 we visualize the AE values computed for the one-step ahead temperature forecasts, produced using the set of methods no 5 of Table 4, in comparison to their corresponding σ, cv and H estimates for the deseasonalized time series (presented in Table 2), while in Figures 16 and 17 we present an analogous visualization for the AE values computed for the one-step ahead precipitation forecasts and the RMSE values computed for the twelve-step ahead precipitation forecasts respectively, produced for the exploration of Problem 3. The estimated parameters for the deseasonalized precipitation time series are presented in Table 3. These figures are representative of the conducted investigation of the existence of a possible relationship between the forecast quality and the estimated parameters for the deseasonalized time series and provide no evidence of such existence either for temperature or precipitation. This fact may be related to our methodological framework and, in particular, to the way that we handle seasonality to produce better forecasts.

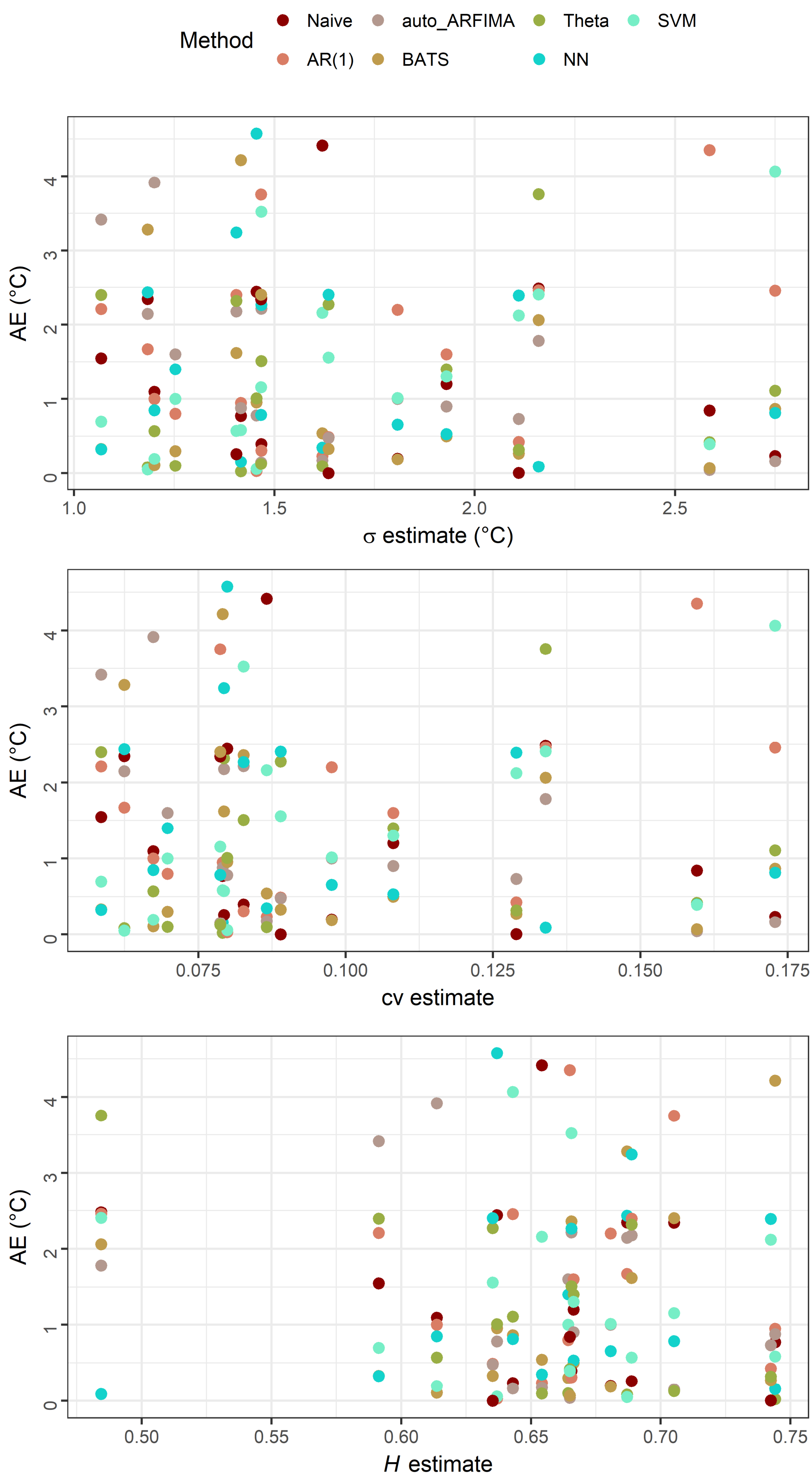


Figure 15. AE values of the one-step ahead temperature forecasts, produced by set of methods no 5 (see Table 4), in comparison to the σ, cv and H estimates.

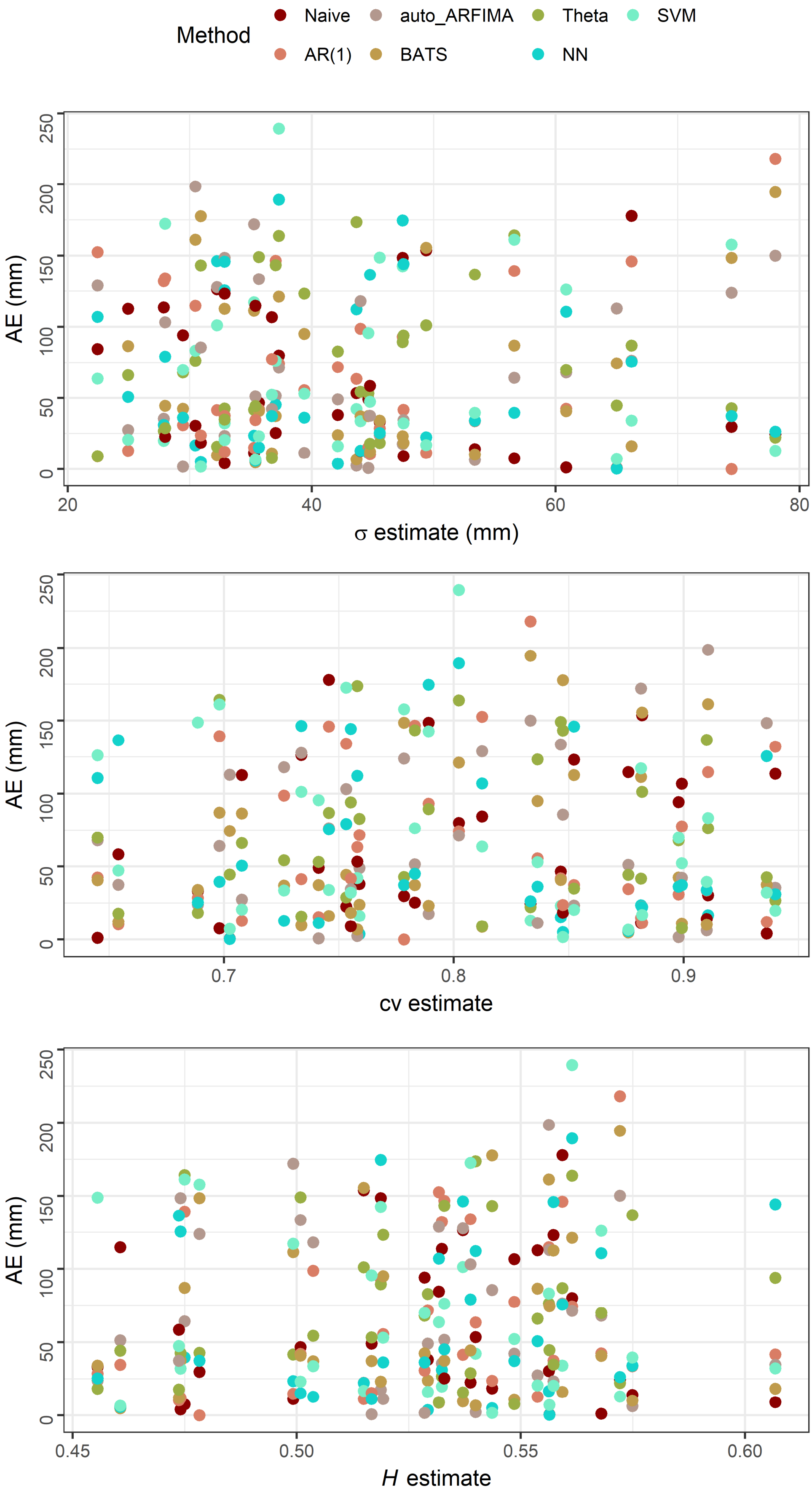


Figure 16. AE values of the one-step ahead precipitation forecasts, produced by set of methods no 5 (see Table 4), in comparison to the σ, cv and H estimates.

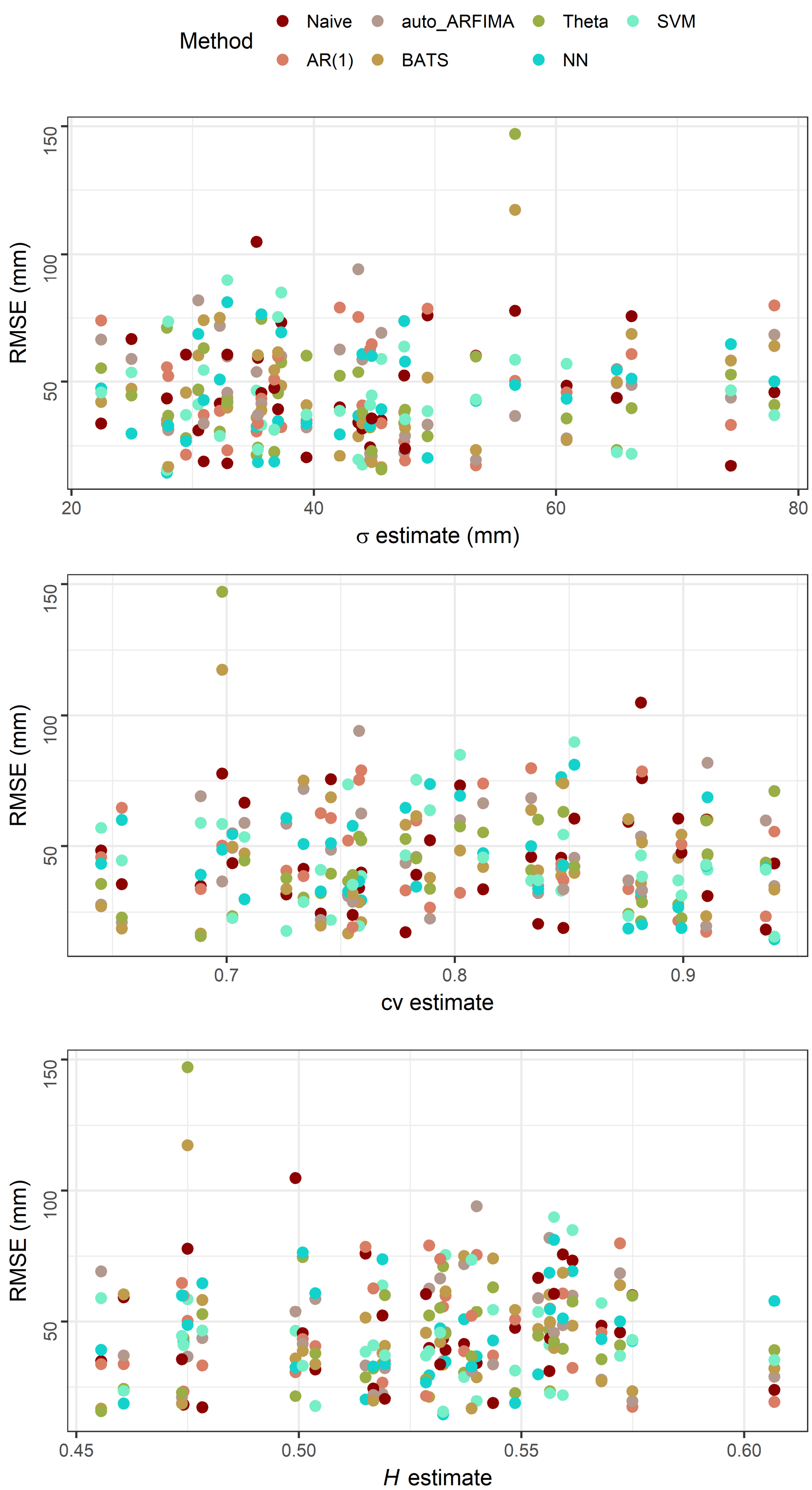


Figure 17. RMSE values of the twelve-step ahead precipitation forecasts, produced by set of methods no 5 (see Table 4), in comparison to the σ, cv and H estimates.

4. Summary and conclusions

We have examined 50 mean monthly temperature and total monthly precipitation time series observed in Greece by applying a fixed methodology to each of them and, subsequently, by performing a cross-case synthesis. The main aim of this multiple-case study is the exploration of three problems associated with univariate time series forecasting using machine learning algorithms, i.e. the (a) lagged variable selection, (b) hyperparameter selection, and (c) comparison between machine learning and classical algorithms. We also present quantitative information about the quality of the forecasts (particularly important for the case of Greece) and search for evidence regarding the existence of a possible relationship between the forecast quality, and the standard deviation, coefficient of variation and Hurst parameter estimates for the deseasonalized time series (used for model-fitting). We have focused on two machine learning algorithms, i.e. neural networks and support vector machines, while we have also included four classical algorithms and a naïve benchmark in the comparisons. We have assessed the one- and twelve-step ahead forecasting performance of the algorithms.

The findings suggest that forecasting methods based on the same machine learning algorithm may exhibit very different performance, to an extent mainly depending on the algorithm and the individual case. In fact, the neural networks algorithm can produce forecasts of many different qualities for a specific individual case, in contrast to the support vector machines one. The performance of the former algorithm seems to be more affected by the selected lagged variables than by the adopted hyperparameter selection procedure (use of predefined hyperparameters or defined after optimization). While no evidence is provided that any of the compared lagged regression matrices systematically leads to better forecasts than the rest, either for the neural networks or the support vector machines algorithms, the results mostly favour using less recent lagged variables. Furthermore, for the algorithms used in the present study hyperparameter optimization does not necessarily lead to better forecasts than the use of the default hyperparameter values of the algorithms. Regarding the comparisons performed between machine learning and classical algorithms, the results indicate that methods from both categories can perform equally well, under the same limitations. The best method depends on the case examined and the criterion of interest, while it can be either machine learning or classical. Some information of secondary importance derived by our experiments is subsequently reported. The average-case performance of the algorithms used to produce one- and twelve-step ahead monthly temperature forecasts ranges between 0.66 °C and 1.00 °C, and 1.14 °C and 1.70 °C, in terms of absolute error and root mean square error respectively. For the monthly precipitation forecasts the respective values are 39 mm and 72 mm, and 41 mm and 52 mm. Finally, no evidence is provided by our multiple-case study that there is any relationship between the forecast quality and the estimated parameters for the deseasonalized time series.

Appendix A Statistical software

The analyses and visualizations have been performed in R Programming Language (R Core Team 2017) by 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), `hydroTSM` (Zambrano-Bigiarini 2017b), `kernlab` (Karatzoglou et al. 2004), `knitr` (Xie 2014, 2015, 2017), `maps` (Brownrigg et al. 2017), `nnet` (Venables and Ripley 2002), `readr` (Wickham et al. 2017), `rminer` (Cortez 2010, 2016), `tidyr` (Wickham and Henry 2017) and `zoo` (Zeileis and Grothendieck 2005).

Acknowledgements

A previous shorter version of the paper has been presented in the 10th World Congress of EWRA "Panta Rei" Athens, Greece, 5-9 July, 2017 under the title *"Forecasting of geophysical processes using stochastic and machine learning algorithms"* (Papacharalampous et al. 2017b). We thank the Scientific and Organizing Committees for selecting this research. We also thank the Guest Editor and two anonymous reviewers of Water Resources Management for the time they have devoted to our work.

References

[1] Achen CH, Snidal D (1989) Rational deterrence theory and comparative case studies. World Politics 41(2):143–169. https://doi.org/10.2307/2010405

[2] Atiya AF, El-Shoura SM, Shaheen SI, El-Sherif MS (1999) A comparison between neural-network forecasting techniques-case study: river flow forecasting. IEEE Transactions on Neural Networks 10(2):402–409. https://doi.org/10.1109/72.750569

[3] Ballini R, Soares S, Andrade MG (2001) Multi-step-ahead monthly streamflow forecasting by a neurofuzzy network model. Joint 9th IFSA World Congress and 20th NAFIPS International Conference:992–997. https://doi.org/10.1109/NAFIPS.2001.944740

[4] Baxter P, Jack S (2008) Qualitative case study methodology: Study design and implementation for novice researchers. The Qualitative Report 13(4):544–559

[5] Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. Journal of Hydrology 508:418–429. https://doi.org/10.1016/j.jhydrol.2013.10.052
[6] Breiman L (2001) Random Forests. Machine Learning 45(1):5–32. https://doi.org/10.1023/A:1010933404324
[7] Brownrigg R, Minka TP, Deckmyn A (2017) maps: Draw Geographical Maps. R package version 3.2.0. https://CRAN.R-project.org/package=maps
[8] Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20(3):273–297. https://doi.org/10.1007/BF00994018
[9] Cortez P (2010) Data mining with neural networks and support vector machines using the R/rminer tool. In: Perner P (ed) Advances in Data Mining. Applications and Theoretical Aspects. Springer Berlin Heidelberg, pp 572–583. https://doi.org/10.1007/978-3-642-14400-4_44
[10] Cortez P (2016) rminer: Data Mining Classification and Regression Methods. R package version 1.4.2. https://CRAN.R-project.org/package=rminer
[11] Dooley LM (2002) Case study research and theory building. Advances in Developing Human Resources 4(3):335–354. https://doi.org/10.1177/1523422302043007
[12] El-Shafie A, Taha MR, Noureldin A (2007) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. Water Resources Management 21(3):533–556. https://doi.org/10.1007/s11269-006-9027-1
[13] 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
[14] Guo J, Zhou J, Qin H, Zou Q, Li Q (2011) Monthly streamflow forecasting based on improved support vector machine model. Expert Systems with Applications 38(10):13073–13081. https://doi.org/10.1016/j.eswa.2011.04.114
[15] Hong WC (2008) Rainfall forecasting by technological machine learning models. Applied Mathematics and Computation 200(1):41–57. https://doi.org/10.1016/j.amc.2007.10.046
[16] Hung NQ, Babel MS, Weesakul S, Tripathi NK (2009) An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrology and Earth System Sciences 13:1413–1425. https://doi.org/10.5194/hess-13-1413-2009
[17] 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
[18] Hyndman RJ, Khandakar Y (2008) Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 27(3):1–22. https://doi.org/10.18637/jss.v027.i03
[19] Karatzoglou A, Smola A, Hornik K, Zeileis A (2004) kernlab - An S4 Package for Kernel Methods in R. Journal of Statistical Software 11(9):1–20. https://doi.org/10.18637/jss.v011.i09
[20] Koutsoyiannis D, Yao H, Georgakakos A (2008) Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. Hydrological Sciences Journal 53(1):142–164. https://doi.org/10.1623/hysj.53.1.142
[21] Krause P, Boyle DP, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 5:89–97
[22] Kumar DN, Raju KS, Sathish T (2004) River flow forecasting using recurrent neural networks. Water Resources Management 18(2):143–161. https://doi.org/10.1023/B:WARM.0000024727.94701.12

[23] Larsson R (1993) Case survey methodology: Quantitative analysis of patterns across case studies. Academy of Management Journal 36(6):1515–1546. https://doi.org/10.2307/256820
[24] 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. Journal of Geophysical Research: Atmospheres 116(D19121). https://doi.org/10.1029/2011JD016187
[25] Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling and Software 15(1):101–124. https://doi.org/10.1016/S1364-8152(99)00007-9
[26] Moustris KP, Larissi IK, Nastos PT, Paliatsos AG (2011) Precipitation forecast using artificial neural networks in specific regions of Greece. Water Resources Management 25(8):1979–1993. https://doi.org/10.1007/s11269-011-9790-5
[27] Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology 10(3):282–290. https://doi.org/10.1016/0022-1694(70)90255-6
[28] Nayak PC, Sudheer KP, Ranganc DM, Ramasastrid KS (2004) A neuro-fuzzy computing technique for modeling hydrological time series. Journal of Hydrology 291(1–2):52–66. https://doi.org/10.1016/j.jhydrol.2003.12.010
[29] Ouyang Q, Lu W (2017) Monthly rainfall forecasting using echo state networks coupled with data preprocessing methods. Water Resources Management 32(2):659–674. https://doi.org/10.1007/s11269-017-1832-1
[30] 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
[31] Papacharalampous GA, Tyralis H, Koutsoyiannis D (2017b) Forecasting of geophysical processes using stochastic and machine learning algorithms. European Water 59:161–168
[32] Peterson TC, Vose RS (1997) An overview of the Global Historical Climatology Network temperature database. Bulletin of the American Meteorological Society 78(12):2837–2849. https://doi.org/10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2
[33] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
[34] Raghavendra NS, Deka PC (2014) Support vector machine applications in the field of hydrology: a review. Applied Soft Computing 19:372–386. https://doi.org/10.1016/j.asoc.2014.02.002
[35] Sivapragasam C, Liong SY, Pasha MFK (2001) Rainfall and runoff forecasting with SSA-SVM approach. Journal of Hydroinformatics 3(3):141–152
[36] 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 Systems with Applications 39(8):7067–7083. https://doi.org/10.1016/j.eswa.2012.01.039
[37] Tongal H, Berndtsson R (2017) Impact of complexity on daily and multi-step forecasting of streamflow with chaotic, stochastic, and black-box models. Stochastic Environmental Research and Risk Assessment 31(3):661–682. https://doi.org/10.1007/s00477-016-1236-4

[38] Tyralis H (2016) HKprocess: Hurst-Kolmogorov Process. R package version 0.0-2. https://CRAN.R-project.org/package=HKprocess
[39] Tyralis H, Koutsoyiannis D (2011) Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. Stochastic Environmental Research and Risk Assessment 25(1):21–33. https://doi.org/10.1007/s00477-010-0408-x
[40] Tyralis H, Papacharalampous GA (2017) Variable selection in time series forecasting using random forests. Algorithms 10(4):114. https://doi.org/10.3390/a10040114
[41] 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. Journal of Hydrology 476(7):433–441. https://doi.org/10.1016/j.jhydrol.2012.11.017
[42] Vapnik VN (1995) The nature of statistical learning theory, first edition. Springer-Verlag New York. https://doi.org/10.1007/978-1-4757-3264-1
[43] Vapnik VN (1999) An overview of statistical learning theory. IEEE Transactions on Neural Networks 10(5):988–999. https://doi.org/10.1109/72.788640
[44] Venables WN, Ripley BD (2002) Modern Applied Statistics with S, fourth edition. Springer-Verlag New York. https://doi.org/10.1007/978-0-387-21706-2
[45] Wang W, Van Gelder PH, Vrijling JK, Ma J (2006) Forecasting daily streamflow using hybrid ANN models. Journal of Hydrology 324(1–4):383–399. https://doi.org/10.1016/j.jhydrol.2005.09.032
[46] 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
[47] Wickham H (2016) ggplot2. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4
[48] 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
[49] Wickham H, Henry L (2017) tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. R package version 0.7.2. https://CRAN.R-project.org/package=tidyr
[50] 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
[51] Witten IH, Frank E, Hall MA, Pal CJ (2017) Data mining: Practical machine learning tools and techniques, fourth edition. Elsevier Inc. ISBN:978-0-12-804291-5
[52] Xie Y (2014) knitr: A comprehensive tool for reproducible research in R. In: Stodden V, Leisch F, Peng RD (eds) Implementing Reproducible Computational Research. Chapman and Hall/CRC
[53] Xie Y (2015) Dynamic documents with R and knitr, 2nd edition. Chapman and Hall/CRC
[54] Xie Y (2017) knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.17. https://CRAN.R-project.org/package=knitr
[55] Yaseen ZM, Allawi MF, Yousif AA, Jaafar O, Hamzah FM, El-Shafie A (2016) Non-tuned machine learning approach for hydrological time series forecasting. Neural Computing and Applications 30(5):1479–1491. https://doi.org/10.1007/s00521-016-2763-0
[56] Yin RK (2003) Case study research: Design and methods, third edition. Sage Publications, Inc

[57] Yu X, Liong SY, Babovic V (2004) EC-SVM approach for real-time hydrologic forecasting. Journal of Hydroinformatics 6(3):209–223
[58] Zambrano-Bigiarini M (2017a) hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series. R package version 0.3-10. https://CRAN.R-project.org/package=hydroGOF
[59] Zambrano-Bigiarini M (2017b) hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling. R package version 0.5-1. https://github.com/hzambran/hydroTSM
[60] Zeileis A, Grothendieck G (2005) zoo: S3 infrastructure for regular and irregular time series. Journal of Statistical Software 14(6):1–27. https://doi.org/10.18637/jss.v014.i06