Stochastic Environmental Research and Risk Assessment (2023) 37:219–231
https://doi.org/10.1007/s00477-022-02290-3

ORIGINAL PAPER

Short-term forecast improvement of maximum temperature by state-space model approach: the study case of the TO CHAIR project

F. Catarina Pereira[1] · A. Manuela Gonçalves[1] · Marco Costa[2]

Accepted: 19 July 2022 / Published online: 3 August 2022

Abstract
In the context of "TO CHAIR" project, this work aims to improve the accuracy of short-term forecasts of maximum air temperature obtained from the https://weatherstack.com/ website. The proposed methodology is based on a state-space representation that incorporates the latent process, the state, which is estimated recursively using the Kalman filter. The proposed model linearly and stochastically relates the forecasts from the website (as a covariate) to the observations of the maximum temperature recorded at the study site. The specification of the state-space model is performed using the maximum likelihood method under the assumption of normality of errors, where empirical confidence intervals are presented. In addition, this work also presents a treatment of outliers based on the ratios between the observed maximum temperature and the website forecasts.

Keywords State-space models · Temperature · Kalman filter · Time series · Data assimilation

1 Introduction

Today we live in a society where climate change has brought to the fore a global concern of the population particularly regarding water availability, since it is a non-renewable natural resource essential to life. This global issue can be seen in the investments made to find water outside planet Earth—for instance, the recent discovery of water on the Moon and Mars (United Nations 2021).

Population growth, pollution, and climate change are the major factors contributing to the dwindling of the planet's water resources. Furthermore, this natural resource has been heavily exploited due to economic growth, and water mismanagement has also become a global concern.

According to the United Nations World Water Development Report (United Nations 2021), agriculture consumes about 69% of the world's available fresh water, which is practically used for irrigation, and this proportion can reach 95% in some developing countries. The temperature increases that have occurred in recent years have contributed to the decrease of available water. Therefore, it is crucial to find the best technical solutions to improve water management, particularly in irrigation systems.

This work is carried out in the context of project "TO CHAIR—The Optimal Challenges in Irrigation" (https://systec.fe.up.pt/projects/FCT-TOCHAIR/) and aims to ensure efficient water management in irrigation systems through mathematical/statistical modeling. This project's main challenge is to study how to manage irrigation problems as an optimal control problem: the daily irrigation problem of minimizing water consumption. For that, it is necessary to measure and forecast meteorological variables in real time at each monitoring area of irrigation.

One solution for obtaining forecasts of weather variables for a specific location is to consider self-service-based weather information web services that are available. But however friendly and affordable this solution is, these

The authors have contributed equally to this work.

✉ Marco Costa
marco@ua.pt

F. Catarina Pereira
id9976@alunos.uminho.pt

A. Manuela Gonçalves
mneves@math.uminho.pt

[1] Department of Mathematics and Center of Mathematics, University of Minho, Campus de Gualtar of University of Minho, 4710-057 Braga, Portugal

[2] Águeda School of Technology and Management and Center for Research and Development in Mathematics and Applications, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

forecasts contain increased uncertainties because they are the result of spatial interpolation procedures, such as the geostatistical interpolation technique known as Kriging or other (Costa and Alpuim 2011; Berndt and Haberlandt 2018). Although these platforms may adopt diverse methodologies, the forecasts for a given location are derived from the observations obtained from a network of previously existing fixed meteorological stations that have an irregular distribution and density in space.

This work aims both to establish a statistical modeling approach based on meteorological variables and in state-space modeling and to improve the forecasts obtained from an easily accessible online website. In this context, forecast improvement aims to improve the quality indicators usually considered, such as the square root of the mean square error (RMSE), among others, as mentioned later.

This approach aims to establish a model that can model and calibrate or correct in real time the predictions obtained from the website. It is intended that a dynamic model calibrates the forecasts that are assumed to have a forecast error increased by the interpolation error of the website by incorporating the values observed at the study site through a portable weather station.

The framework proposed in this work can be seen as a method of Data Assimilation (DA), since the proposed approach combines forecasts across time and from different sources. DA is in general a sequential time-stepping procedure, in which forecasts from a source are compared with newly received observations to produce optimal forecasts. These methods have been widely applied in several areas, but most often in meteorology, where meteorological forecasts are improved by combining forecasts and observations (Navon 2009), or by combining several remote sensing data sources that might be used for automated prediction of forest variables, such as growing stock volume (Lindgren et al. 2021).

This approach is developed for the maximum air temperature in order to improve the accuracy of the forecasts obtained from the website https://weatherstack.com/. This variable was selected because it is a meteorological variables that most interferes in the evapotranspiration process (Rodrigues and Braga 2021) and it is useful to calculate crop reference evapotranspiration (ETo) (Allan et al. 1998). Although the maximum temperature is an extreme variable and therefore has specific characteristics (Leadbetter et al. 1983), the approach proposed in this work based on the DA method avoids the generality of the problems associated with this type of variables.

In the literature there are several research studies based on temperature data analysis using different methodologies. For instance, Ksiksi and Al Blooshi (2019) employed the ARIMA modeling approach coupled with STI (Standardized Temperature Index) to predict temperature anomalies using average monthly temperatures across four United Arab Emirates bio-climatic regions: Abu Dhabi, Al-Ain, Dubai, and Sharjah. An application of TBATS and regression models with correlated errors methods to the minimum air temperature were considered in Gonçalves et al. (2021) to deal with time series with complex seasonal patterns. Citakoglu (2021) investigated four different machine learning-based approaches in estimating monthly maximum temperature, average temperature, and minimum temperature between 1974 and 2020, recorded in 275 meteorological stations in Turkey. Glynis et al. (2021) studied the temporal evolution of near-surface air temperature extremes, having found that these meteorological variables showed a much more complex behavior than usually modeled by classical statistics. In this study, Glynis et al. (2021) states that the climacogram and the K-moments present several advantages compared to autocovariance and the power-spectrum and as a variant of probability weighted moments. The analysis and modeling of extreme environmental series can be a real challenge, because in many cases, when observations are subject to greater uncertainties, the solutions obtained by the Kalman filter may suffer from conditional biases, i.e., a quantity that shows how the estimates differ on the average from a given true observation, and thus negatively influence the estimation and prediction of extremes (Seo et al. 2018).

In this work, a linear state-space model, which belongs to a general and flexible class of stochastic models (Woody et al. 2020), is applied to study the time series of maximum air temperature. We selected this class of models due to both their efficiency from a stochastic point of view and their Markovian nature, which allows working with samples of variable size, where the predictions are recursively updated as new observations are incorporated into the model, updating it and making it more efficient by producing the improved predictions and forecasts.

In the literature there are several papers that analyze meteorological time series through state-space models. In particular, Costa and Monteiro (2019) analyzed long-term time series of monthly average air temperature of three Portuguese cities—Lisbon, Oporto, and Coimbra—through periodic mixed linear state-space model associated with the Kalman filter, which incorporates fixed effects (seasonality with fixed level effects) and stochastic effects (dynamic slopes throughout the months of the year). Parameter estimation was done using the maximum likelihood method. The mean quarterly temperature and rainfall time series were studied by Prokosch et al. (2019) to analyze the relationship between temperature and body mass in a population of "mountain wagtails" in KwaZulu-Natal, South Africa, between 1976 and 1999, where the time series were decomposed into trend and seasonal effects using a state-space model. Adedotun et al. (2020) modeled

the temperature and rainfall dataset using the state-space model associated with the univariate Kalman filter to both identify the pattern of the trend movement in the occurrence of climatic change and to evaluate the forecasting power of the proposed models. Parameter estimation was also conducted using the maximum likelihood method.

In many applications (environmental, economic, etc.) whose series tend to have complex seasonality or strong temporal correlation, the normality of the errors is not always guaranteed or is rejected in the model validation procedure by analyzing the normality of the residuals. Thus, in the literature there are alternative non-parametric estimation methods to parameterize the state-space models, such as the quasi-likelihood and asymptotic quasi-likelihood methods proposed in Alzghool and Lin (2008), Bayesian approaches (Kitagawa 2021), estimators for univariate state-space models based on the generalized method of moments that do not depend on the distribution of errors proposed in Costa and Alpuim (2010) and subsequently generalized to multivariate models in Gonçalves and Costa (2013), and deep learning-based methods (Rangapuram et al. 2018) for which, according to Aghelpour et al. (2021), time series and machine learning models have shown a good performance in estimating and predicting meteorological variables.

The paper is organized as follows: in Sect. 2 we present a brief description of the model, the Kalman filter, parameter estimation, and how to compute h-steps predictions for the state vector. Section 3 presents an exploratory analysis of the series under study, as well as the imputation of missing values and the treatment of outliers. In Sect. 4, the proposed model is applied to the data. Finally, Sect. 5 presents the conclusions.

2 Methodology

State-space models are very flexible models due to their ability to integrate several data features, and to recursively update predictions in real time as new observations become available. Because of their formulation, it is possible to analyze dynamic phenomena that vary significantly over time. According to Petris et al. (2009), they are much more flexible than ARIMA models in dealing with non-stationary time series and modeling structural changes and are generally easier to interpret.

2.1 The time-varying linear state-space model

In general, the time-varying linear state-space model is characterized by two equations: the observation and state equations. In this work, the proposed model is given by

$$Y_t = \beta_t W_{t,(h)} + e_t, \tag{1}$$

$$\beta_t = \mu + \phi(\beta_{t-1} - \mu) + \varepsilon_t, \tag{2}$$

where Eq. (1) is the observation equation, which relates to the observed variable $W_{t,(h)}$, which is the h-days ahead forecasts coming from the https://weatherstack.com/ website collected at time $t-h$, with the unobserved variable β_t, called the state vector, through a linear relation. Equation (2) translates the stochastic model underlying the state vector. Y_t is the observed maximum temperature on the farm at day t, with $t = 1, \ldots, n$.

The process $\{\beta_t\}$ follows a stationary autoregressive process of order one, that is, $\{\beta_t\} \sim \text{AR}(1)$ and $|\phi| < 1$. It is assumed that $E(\beta_t) = \mu$ and $var(\beta_t) = \sigma_\varepsilon^2/(1-\phi^2)$. Moreover, processes $\{e_t\}$ and $\{\varepsilon_t\}$ are uncorrelated white noise processes, i.e., $E(e_t) = 0$, $E(e_t e_s) = 0$, $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_s) = 0$, and $E(e_t \varepsilon_s) = 0$, $\forall t, s$. In view of statistical inferences, it is assumed that errors processes are normally distributed, i.e., $e_t \sim N(0, \sigma_e^2)$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

The parameters of the model $\Theta = \{\mu, \phi, \sigma_e^2, \sigma_\varepsilon^2\}$ can be known or must be estimated for each time horizon $h = 1, 2, \ldots, 6$. In this case, for the specification of the state-space model it is necessary to estimate Θ according to an estimation method presented in Sect. 2.3. As the state process β_t is a latent process, it must be predicted based on the observed variables Y_t through the Kalman filter algorithm.

2.2 The Kalman filter

The most common procedure to predict the state vector is the Kalman filter algorithm proposed by Kalman (1960), which is a recursive algorithm that, at each time, computes the optimal predictions of the state vector, in the sense that it has the minimum mean square error (MSE), based on the information available up to instant t, if normality is verified. Nevertheless, according to Harvey (2009), even when the normality assumption of errors is not verified, the Kalman filter still returns optimal predictions within the class of all linear predictors, i.e., the estimates are relatively robust to small deviations from normality (Kokic et al. 2011). However, these optimal properties of the Kalman filter predictors can only be guaranteed when all the parameters of state-space model are known (Rodríguez and Ruiz 2012; Costa and Monteiro 2016).

Let $\widehat{\beta}_{t|t-1}$ denote the predictor of β_t based on the observations $Y_1, Y_2, \ldots, Y_{t-1}$ and $P_{t|t-1}$ its MSE, this is, $P_{t|t-1} = E[(\widehat{\beta}_{t|t-1} - \beta_t)^2]$. Considering the model defined by Eqs. (1) and (2), the linear Kalman filter algorithm is expressed by a set of iterative equations presented in Fig. 1. In step 1 it is computed $\widehat{Y}_{t|t-1}$, the one-step ahead forecast by the linearity of Eq. (1).

When Y_t is observed in time t, steps 3 and 4 allow updating the forecast $\widehat{\beta}_{t|t-1}$ based on the one-step ahead forecast's error $Y_t - \widehat{Y}_{t|t-1}$ and Kalman gain (step 3). With the best prediction of state in time t, $\widehat{\beta}_{t|t}$, the forecast of state to time $t+1$, $\widehat{\beta}_{t+1|t}$ is computed in step 5.

The success of the Kalman filter lies in the fact that it is a real-time estimation procedure, updating and improving the predictions of the state vector when new observations become available. Moreover, it can be applied to both stationary and non-stationary processes, depending on whether the autoregressive parameter ϕ is inside or outside the unit circle, respectively.

2.3 Parameter estimation

In practice, some or even all parameters $\Theta = \{\mu, \phi, \sigma_e^2, \sigma_\varepsilon^2\}$ of the state-space model Eqs. (1)–(2) are unknown and must be estimated (Durbin and Koopman 2001).

By assuming the normality of noise processes, the estimation of the unknown parameters is usually performed using the maximum likelihood method. So, in general, under the assumption of normality and independence of the noise processes and the initial state β_1, the log-likelihood depends of Θ, given the data $(Y_1, Y_2, \ldots, Y_n)$, and can be written through conditional distributions, given by:

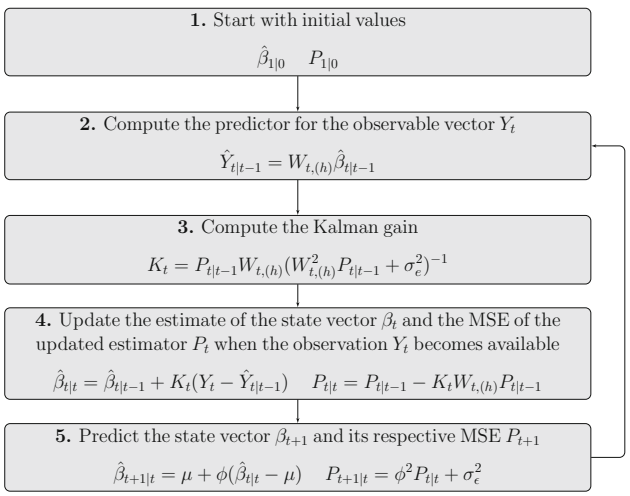

Fig. 1 The Kalman filter algorithm for the model Eqs. (1)–(2)

$$\begin{aligned}\log L(\Theta; Y_1, Y_2, \ldots, Y_n) &= \sum_{t=1}^{n} \log f_{\widehat{Y}_{t|t-1}}(Y_t \mid Y_1, Y_2, \ldots, Y_{t-1}) \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\log(\Sigma_t) \\ &\quad -\frac{1}{2}\sum_{t=1}^{n}\eta_t^2\Sigma_t^{-1}\end{aligned}$$

where $\Sigma_t = W^2_{t,(h)}P_{t|t-1} + \sigma_e^2$ and $\eta_t = Y_t - \widehat{Y}_{t|t-1}$.

Parameter estimation is done by maximizing the log-likelihood function (3). However, since this function is highly non-linear, the optimization is done through numerical methods, such as the Expectation–Maximization algorithm or the Newton–Raphson method. Generally, these algorithms start from an initial set of values Θ_0, realize a set of iterations, and a new value for the log-likelihood function is calculated for each iteration. When the log-likelihood function's increments are smaller than a small value δ defined according to a certain tolerance level, the algorithm ends.

2.4 Forecast correction procedure

After fitting the models to time horizons $h = 1, 2, \ldots, 6$, that is, after the training procedure, the goal is to correct or "calibrate" the website's forecasts $W_{t,(h)}$ regarding the maximum temperature at time $t+h$ based on the information available at time t, i.e., the observed temperature Y_t by the portable station. To correct the forecasts to improve their accuracy, a procedure based on both the Kalman filter predictions and the AR(1) properties is designed.

The forecast of Y_{t+h} at time t can be performed directly from $\widehat{\beta}_{t+h|t}$ taking the linearity of Eq. 1 into account. In fact, taking the expected value of the observation equation at time $t+h$ it comes

$$\begin{aligned}\widehat{Y}_{t+h|t} &= E(Y_{t+h} \mid Y_1, \ldots, Y_t) \\ &= E\left(\beta_{t+h}W_{t+h,(h)} + e_{t+h} \mid Y_1, \ldots, Y_t\right) \\ &= \widehat{\beta}_{t+h|t}W_{t+h,(h)},\end{aligned}$$

and its mean square error is given by

$$\begin{aligned}\Sigma_{t+h|t} &= var\left(\beta_{t+h}W_{t+h,(h)} + e_{t+h}\right) \\ &= W^2_{t+h,(h)}P_{t+h|t} + \sigma_e^2.\end{aligned}$$

The forecast of the state for time $t+h$ given the present, t, $\widehat{\beta}_{t+h|t}$ can be computed through filtered prediction of the state at time t, $\widehat{\beta}_{t|t}$ given by the Kalman filter algorithm (Fig. 1). In fact, according the autoregressive structure of Eq. 2, the state at time $t+h$, β_{t+h}, can be rewritten from β_t by

$$\beta_{t+h} - \mu = \phi^h(\beta_t - \mu) + \sum_{j=1}^{h-1} \phi^{h-j}\varepsilon_{t+j} + \varepsilon_{t+h}.$$

Thus, the predictor of β_{t+h} at time t is the expected value

$$\begin{aligned}\widehat{\beta}_{t+h|t} &= E\left(\beta_{t+h} \mid Y_1, \ldots, Y_t\right) \\ &= \mu + \phi^h(E(\beta_t \mid Y_1, \ldots, Y_t)) - \mu) \\ &= \mu + \phi^h\left(\widehat{\beta}_{t|t} - \mu\right).\end{aligned}$$

The mean square error of $\widehat{\beta}_{t+h|t}$ is given by

$$\begin{aligned}P_{t+h|t} &= E\left[\left(\beta_{t+h} - \widehat{\beta}_{t+h|t}\right)^2\right] \\ &= \phi^{2h}P_{t|t} + \sum_{j=0}^{h-1} \phi^{2j}\sigma_\varepsilon^2.\end{aligned}$$

Thus, the h-steps ahead forecast of Y_{t+h} is computed with the filtered prediction of the state at time t, $\widehat{\beta}_{t|t}$ and its mean square error $P_{t+h|t}$ with the mean square error of the filtered prediction of the state, $P_{t|t}$ obtained by the Kalman filter equations (Fig. 1).

The 95% asymptotic confidence interval of $Y_{t+h|t}$ is given by

$$\left]\widehat{Y}_{t+h|t} - 1.96\sqrt{\Sigma_{t+h|t}}, \widehat{Y}_{t+h|t} + 1.96\sqrt{\Sigma_{t+h|t}}\right[. \quad (3)$$

To evaluate and compare the performance of the applied model to $h = 1, 2, \ldots, 6$, some evaluation measures are considered, namely the root mean square error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE)

$$\text{RMSE}_{(h)} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{Y}_{t|t-h}\right)^2}, \quad (4)$$

$$\text{MAE}_{(h)} = \frac{1}{n}\sum_{t=1}^{n}|Y_t - \widehat{Y}_{t|t-h}|, \quad (5)$$

$$\text{MAPE}_{(h)} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \widehat{Y}_{t|t-h}}{Y_t}\right| \times 100, \quad (6)$$

where Y_t is the observed maximum temperature recorded by the portable weather station and $\widehat{Y}_{t|t-h}$ is its forecast at time $t - h$.

3 Dataset description

The dataset used in this study comes from two different databases. The first corresponds to daily maximum temperature (°C) obtained from a portable weather station installed in a farm called Senhora da Ribeira (Fig. 2), in Carrazeda de Ansiães, located in the district of Bragança, in the northern region of Portugal. These observations were collected in the period between February 20, 2019 and October 11, 2019, totaling 234 records. The second database refers to the forecasts obtained from the website https://weatherstack.com/, with a time horizon up to 6 days for the same meteorological variable and for the same location and time period.

3.1 Exploratory data analysis

The database corresponding to observations recorded at the farm by the portable station, Y_t, has missing values between April 24 and May 1, 2019, and on September 29 and 30, 2019. To deal with this problem, a complementary dataset was considered to perform a data imputation for missing values. This dataset is related to the same period with the maximum temperature measured by a fixed meteorological station in Vila Real, located about 50 km from the farm under study. The imputation procedure consists in establishing a linear interpolation model fitting a simple linear regression model given by $T_{t,\text{CA}} = \alpha + \beta T_{t,\text{VR}} + \zeta_t$, where $T_{t,\text{CA}}$ and $T_{t,\text{VR}}$ are the maximum air temperature recorded in Carrazeda de Ansiães and Vila Real, respectively. In fact, the Vila Real dataset is a good predictor of farm location, since these data have a strong linear correlation The model was fitted to data from January 1 to December 31, 2019, and its estimates are $\widehat{\alpha} = 0.1538$ and $\widehat{\beta} = 1.1241$, with a coefficient of determination equal to 0.9340.

Figure 3 shows (in black) the observed maximum temperature measured in the farm, the website's 1-step ahead forecasts (top) and 6-steps ahead forecasts (bottom) coming from the website (in blue), and the imputed values (in red) between February 20 and October 11, 2019. The analysis of the website's forecasts for $h = 1, 2, \ldots, 6$ and the

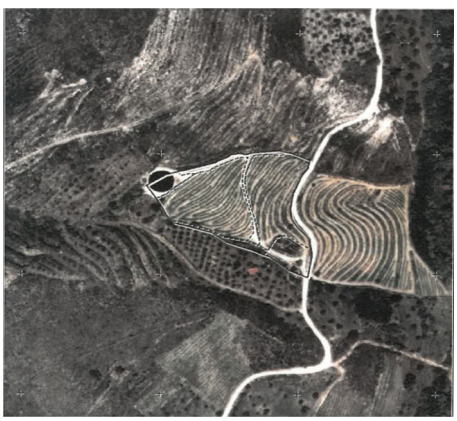

Fig. 2 Senhora da Ribeira farm in Bragança, Portugal

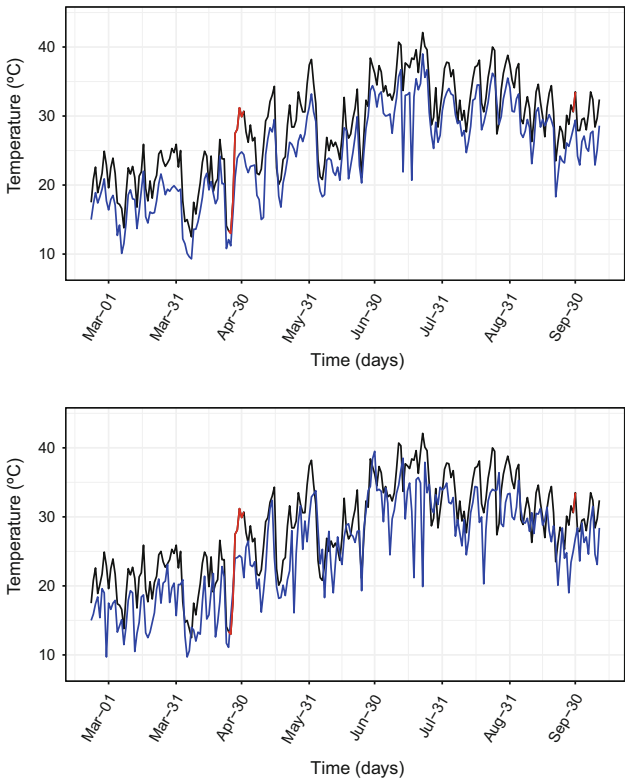

Fig. 3 Observed maximum temperature (in black), website's 1-step ahead (top) and 6-steps ahead (bottom) forecasts (in blue), and imputed values (in red) between February 20 and October 11, 2019

observed values indicate that, as a rule, the website's predictions tend to underestimate the observed values at the farm for all values of forecasting time horizons h. On the one hand, some forecasts have large forecasting errors, the greatest being equal to 22.8 degrees on July 21, 2019 for $h = 5$.

Despite these discrepancies, there is a global linear relationship between forecasts and observations. Table 1 shows sample Pearson's linear correlation coefficients between the observations of maximum temperature recorded at the farm and the website's h-steps ahead forecasts, $\widehat{\rho}(Y_t, W_{t,(h)})$, whose correlations are greater than 0.86 and statistically significant (p values < 0.01), i.e., the forecasts show a strong linear correlation with their respective observations. Nevertheless, the linear correlation decreases as the forecasts' time horizon increases.

Table 1 Sample Pearson linear correlation coefficients, $\widehat{\rho}(Y_t, W_{t,(h)})$, between the observed maximum temperature and the website's h-steps ahead forecasts

h	1	2	3	4	5	6
$\widehat{\rho}(Y_t, W_{t,(h)})$	0.96	0.95	0.93	0.92	0.89	0.86

3.2 Modeling and assessment strategy

The adopted strategy consists of dividing the dataset into two subsets: the training and test sets. The training set is composed of data from February 20 to September 10, 2019 (203 days); the test set is composed of data from September 11 to October 11, 2019 (31 days), which corresponds to 1 month for the quality evaluation of the forecast correction procedure in a *online* framework. The training set is used to fit the model and the test set is used to evaluate the fitted models in the training dataset.

Figure 4 shows the boxplots of the ratios between the observed maximum temperature and the respective website forecasts in the training period. The increase of the dispersion is visible as the time horizon of the forecasts increases, where 1-step ahead forecasts tend to be more accurate given the lower variability. It is also identified that medians have values close to 1.2, indicating that the website's forecasts underestimate the observations by about 20%, in the median perspective, and the presence of moderate and severe outliers is evidenced by the boxplots.

The presence of outliers can have an impact on parameter estimates, forecasting and on inference results (You et al. 2020). It is common practice to remove them completely from the series, but sometimes it may not be the most correct choice. Such decision depends on the origin of the outliers, which may be an error or a genuine extreme value (Chatfield and Xing 2019). In this context, we remove these outliers in the series of maximum temperature forecasts from the website, $W_{t,(h)}$, and replace them with estimates through linear interpolation, denoted by $W^*_{t,(h)}$, based on the observation at time $t-1$ and the observation at time $t+1$, since these data are non-seasonal (Hyndman and Athanasopoulos 2018). Moritz et al. (2015) compared several interpolation methods in time series, but the linear interpolation was one of the methods that showed the best results.

Thus, we replaced the outliers with their linear interpolation estimates only in the training subset, keeping the real observations in the test series for a more adequate evaluation of the performance of the proposed methodology in a online framework.

4 Results and discussion

Table 2 shows the parameters estimates and respective standard errors for each of the six models, $h = 1, \ldots, 6$. First, the estimates of ϕ are between -1 and 1, and therefore state processes are stationary.

The estimates of the state mean, μ, for each of the models are greater than 1 for all time horizons. This

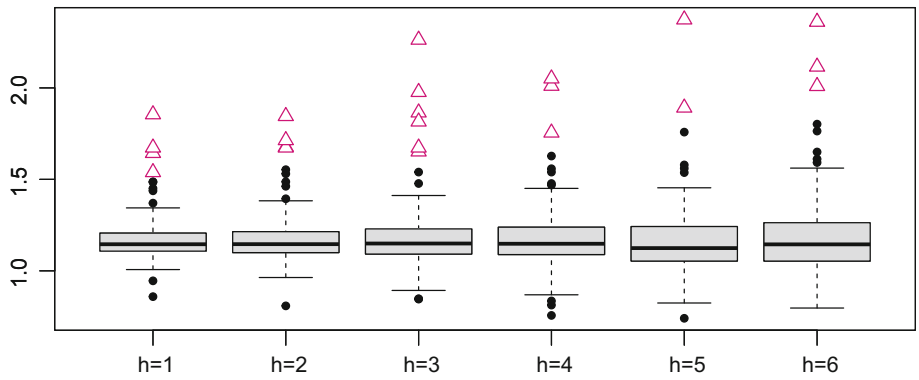

Fig. 4 Boxplots of the ratio between the observed maximum temperature and the respective forecasts at h-steps ahead, $h = 1, \ldots, 6$ the in training period

Table 2 Parameters estimates and respective standard errors of the applicated models to h-days ahead ($h = 1, \ldots, 6$) of the maximum temperature

h	1	2	3	4	5	6
ϕ						
Estimate	0.9757	0.9445	0.9795	0.9677	0.8886	0.6557
Std. error	0.0360	0.0386	0.0218	0.0291	0.2016	0.2223
μ						
Estimate	1.1421	1.1432	1.1335	1.1315	1.1263	1.1256
Std. error	0.0367	0.0284	0.0370	0.0281	0.0201	0.0179
σ_ε						
Estimate	0.0102	0.0217	0.0100	0.0122	0.0282	0.0738
Std. error	0.0071	0.0069	0.0037	0.0049	0.0310	0.0331
σ_e						
Estimate	1.3491	1.3333	1.9962	2.2864	2.6444	2.5710
Std. error	0.0966	0.1101	0.1067	0.1240	0.3007	0.4954
$\log L$	− 180.3209	− 196.4321	− 252.6198	− 280.1320	− 318.1104	− 346.8648
BIC	381.8946	414.1169	526.4925	581.5169	657.4737	714.9823

indicates that the website's forecasts tend to underestimate the observed temperature in its forecasts. Estimates of μ for maximum temperature (Table 2) are between 1.13 and 1.14, which means that, on average, the observed maximum temperature is about 13% and 14% higher compared to the forecasts. This bias can be justified by taking into consideration that Carrazeda de Ansiães lies on an interfluvial plain, demarcated to the South and West by the deep valleys of the Douro and Tua rivers, and to the North and Northeast by a plateau. These characteristics of the farm's location justify the poor quality of the website's predictions.

The state equation errors show low variability compared to the observation equation error, whose variability tends to increase with the increase of the forecasts' time horizon. This happens because the error of the state equation is dimensionless or has no units, since it is a calibration factor between Y_t and $W_{t,(h)}$, while the error of the observation equation is in °C.

The Schwarz information criterion (BIC) is a measure of the goodness of fit that is smaller for the model with $h = 1$ and is therefore considered the best model.

Figure 5 shows the observed maximum temperature, h-steps ahead forecasts, and the corrected h-steps ahead forecasts with the 95% empirical confidence intervals in the training period for $h = 1, 6$, where the range of confidence intervals are larger for $h = 6$ compared to $h = 1$. Although uncertainty is higher for the corrected 6-steps ahead forecasts, the point estimates are accurate.

Table 3 shows sample Pearson's linear correlation coefficients between the observations of maximum temperature recorded at the farm and the corrected h-steps ahead forecasts, $\widehat{\rho}(Y_t, \widehat{Y}_t)$, whose in-sample correlations are greater than 0.90. In the out-of-sample correlations there is a clear decrease from $h = 5$, thus corroborating the impact of the time horizon on the forecasts.

To validate the models, the residuals analysis was performed to verify if they have a similar behavior to white noise. The normality assumption was verified for all 6 models using the Kolmogorov–Smirnov test (all p values > 0.05). To verify the assumption of independence, the Ljung–Box portmanteau test was performed taking $k = 10$ as the maximum lag to be considered (Hyndman and Athanasopoulos 2018), whose hypothesis of independence

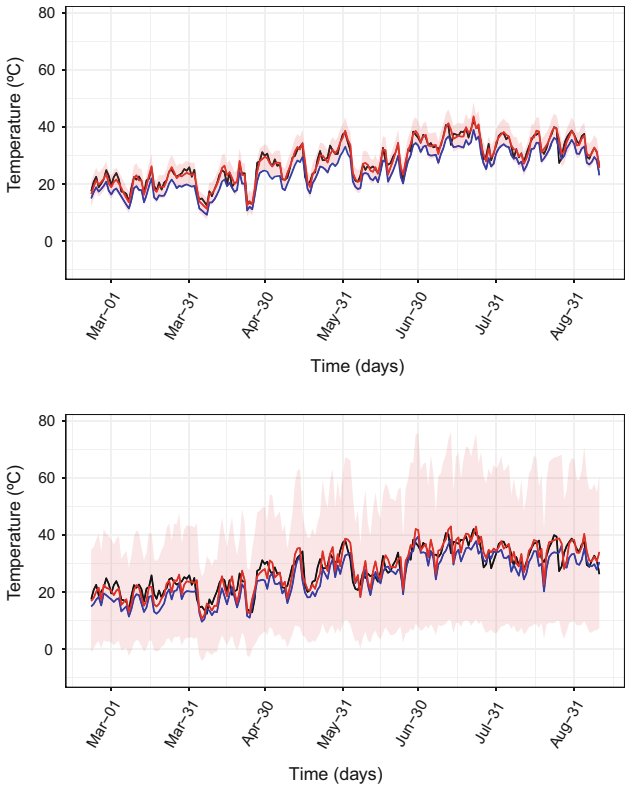

Fig. 5 Observed maximum temperature (in black), h-steps ahead forecasts (in blue) and corrected h-steps ahead forecasts (in red) between February 20 and September 10, 2019 (training period); top: $h = 1$, bottom: $h = 6$. Shadow: 95% empirical confidence intervals

Table 3 Sample Pearson linear correlation coefficients, $\widehat{\rho}(Y_t, \widehat{Y}_t)$, between the observed maximum temperature and the corrected h-steps ahead forecasts

h	1	2	3	4	5	6
$\widehat{\rho}(Y_t, \widehat{Y}_t)$ in-sample	0.98	0.98	0.96	0.95	0.92	0.90
$\widehat{\rho}(Y_t, \widehat{Y}_t)$ out-of-sample	0.86	0.83	0.81	0.83	0.69	0.35

was rejected for the maximum temperature models with $h = 4$ (p value $= 0.017$) and $h = 6$ (p value $= 0.004$), and checked by autocorrelation function (ACF) and partial autocorrelation function (see Fig. 6). For example, the significant correlations in the partial ACF for h $= 6$ are lags 10 and 11, and the rejection hypothesis depends on the maximum lag being considered. In these cases, only a few lags have a significant correlation and do not represent any kind of underlying correlation structure. Therefore, we can consider that globally the residuals have a white noise behavior for all $h = 1, 2, \ldots, 6$.

To evaluate the quality of the forecasts obtained by the models, the last part of the series was reserved without any outliers replaced by imputation, that is, the original out-of-sample (September 11 to October 11, 2019). For each of the models, which are identified by the time horizons of the forecasts from 1 to 6 days, forecasts were obtained according to the data available in real time respecting their time horizons, i.e., on day t the observation of the portable station at the farm, Y_t, is registered and the website's forecasts, $W_{t+h,(h)}$, $h = 1, \ldots, 6$, become available.

So, for example, for the model with $h = 1$ fitted to the training sample, the forecasts for September 11 were obtained; on September 11, the observation of this day was included in the model and the website's 1-step ahead forecast was used to predict the maximum temperature for September 12, and so on. Considering the time horizon of 6 days and its model fitted to the training sample, the website's forecast for September 16 was obtained on September 10. So, a new observation on September 11 by the portable station located in the farm is available and allows obtaining the filtered state prediction on September 11, which, associated to the website's 6-steps ahead forecast, produces the calibrated maximum temperature for September 17, and so on (Fig. 7).

To evaluate the predictive quality of the models, two absolute measures, RMSE and MAE and a relative measure, MAPE, were compared (Table 4). These results are also represented in Figs. 8 and 9 for a better comparative analysis.

As overall evaluation measures and to characterize the initial errors between the website's forecasts and the observed temperatures at the farm, we find that the website's forecasts are significantly inaccurate, as the RMSE varies from 4.1504 for $h = 1$ to 5.0082 for $h = 6$, increasing when the temporal horizon of the forecast increases. For $h = 4$, we see that the website's predictions have a slightly lower RMSE than the RMSE for $h = 3$, i.e., the original forecasts for $h = 4$ are globally and slightly more accurate than for h $= 3$. For the original data, MAE does not show an overall trend. In fact, the MAE varies from 3.7231 for $h = 2$ to 4.0919 for $h = 6$, becoming clear that it is smaller for $h = 1, 2$. The relative MAPE indicator ranges from a minimum of 16.9248% for the 1-step ahead forecasts to a maximum of 19.3711% for the 6-steps ahead forecasts.

It can be seen that the evaluation measures in the training period (in-sample uncorrected) are very similar to the evaluation measures for the overall available period. Naturally, this fact was already expected, given that the training period encompasses a large part of the available observations (203 out of 234 available). It is noted that all measures tend to increase with the increase of the time horizon, considering the entire observation period. In the training period, there was a reduction greater than 64% in the RMSE, MAE, and MAPE of the corrected 1-step ahead

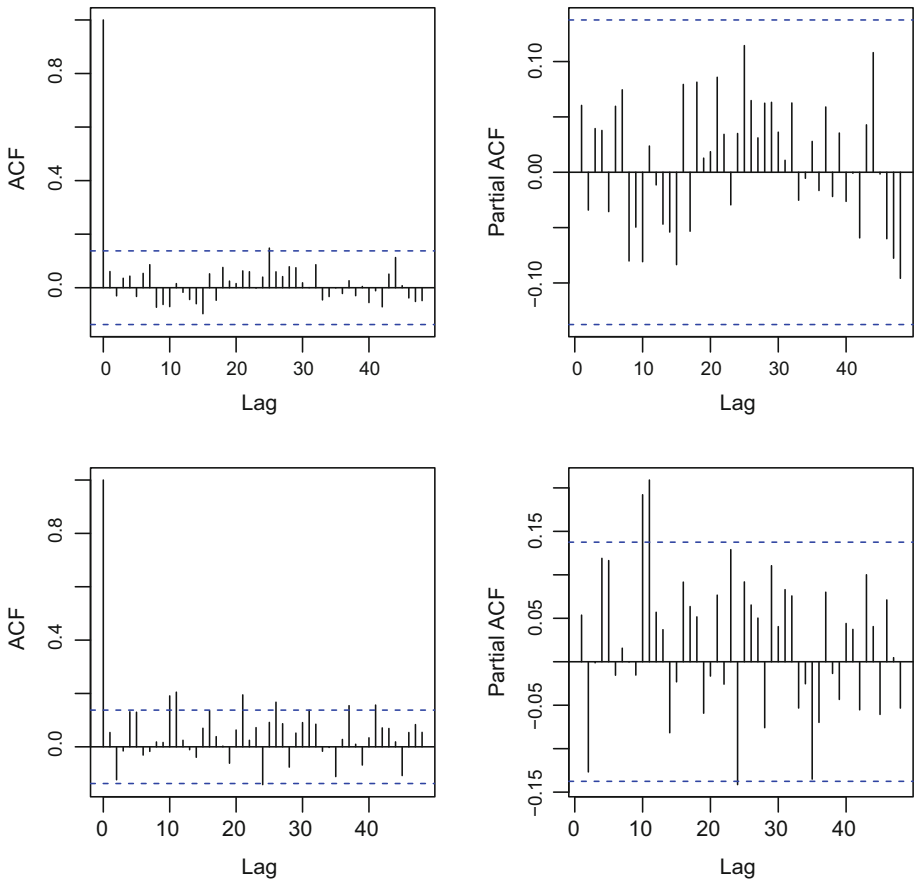

Fig. 6 ACF and partial ACF of standardized residuals; top: $h = 1$, bottom: $h = 6$. Dashed: 95% confidence intervals

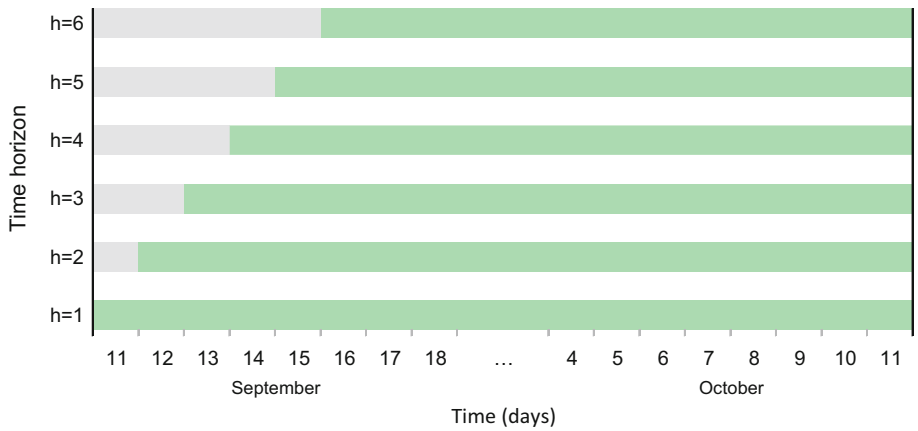

Fig. 7 Forecast period considering the forecasts' time horizon

forecasts, and a reduction greater than 33% in the corrected 6-steps ahead forecasts compared to the website's initial forecasts.

However, by analyzing the quality measures regarding the evaluation period (out-of-sample uncorrected) we find that for the various time horizons the website's forecasts present values of RMSE, MAE and MAPE globally lower than the same measures in the overall available period. That is, regarding the last 31 days, the website's forecasts present lower forecast errors in relation to the global series for the 234 days. In particular, for $h = 4$, in the evaluation period, the website's forecasts present an RMSE that corresponds to 89% of the global RMSE, while for $h = 3$ this value is 94%. And even in nominal terms, the RMSE for $h = 4$ is lower than for $h = 3$, both in the overall period and particularly in the evaluation period (out-of-sample). In the test period there was a reduction greater than 57% in the RMSE, MAE, and MAPE of the corrected 1-step ahead forecasts, and a reduction greater than 17% in the RMSE, MAE, and MAPE of the corrected 6-steps ahead forecasts

Table 4 Evaluation measures between the observed maximum temperature and the respective h-steps ahead forecasts before and after in-sample, out-of-sample, and global correction

	$h = 1$			$h = 2$			$h = 3$		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Global uncorrected Y_t versus $W_{t,(h)}$	4.1504	3.7795	16.9248	4.1962	3.7231	16.9718	4.5165	3.8893	18.0913
In-sample uncorrected Y_t versus $W_{t,(h)}$	4.1882	3.7951	17.3181	4.2158	3.7118	17.2955	4.5540	3.8640	18.4159
In-sample corrected Y_t versus $W^*_{t,(h)}$	3.7905	3.5759	15.7690	3.7442	3.4847	15.6868	3.9419	3.5719	16.0571
In-sample corrected Y_t versus $\widehat{Y}_t$	1.4776	1.0856	4.1737	1.5994	1.1728	4.5296	2.1104	1.6152	6.2547
In-sample reduction (%) $W_{t,(h)}$ versus $\widehat{Y}_t$	64.72	71.39	75.90	62.06	68.40	73.81	53.66	58.20	66.04
Out-of-sample uncorrected Y_t versus $W_{t,(h)}$	3.8936	3.6774	14.3497	4.0657	3.7968	14.8525	4.2626	4.0548	15.9657
Out-of-sample corrected Y_t versus $\widehat{Y}_t$	1.6469	1.2724	4.4941	1.7203	1.4058	4.8399	1.9701	1.4948	5.3969
Out-of-sample reduction (%) $W_{t,(h)}$ versus $\widehat{Y}_t$	57.70	65.40	68.68	57.69	62.97	67.41	53.78	63.14	66.20
	$h = 4$			$h = 5$			$h = 6$		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Global uncorrected Y_t versus $W_{t,(h)}$	4.4462	3.8457	17.6187	4.6234	3.8077	17.3027	5.0082	4.0919	19.3711
In-sample uncorrected Y_t versus $W_{t,(h)}$	4.5157	3.8512	18.0687	4.7099	3.8113	17.7174	5.0849	4.1557	20.0726
In-sample corrected Y_t versus $W^*_{t,(h)}$	3.9775	3.5448	15.8816	4.1698	3.5828	16.0467	4.4181	3.8291	17.4869
In-sample corrected Y_t versus $\widehat{Y}_t$	2.4181	1.8587	7.1440	2.9098	2.2855	8.8806	3.3755	2.6602	10.2684
In-sample reduction (%) $W_{t,(h)}$ versus $\widehat{Y}_t$	46.45	51.74	60.46	38.22	40.03	49.88	33.62	35.99	48.84
Out-of-sample uncorrected Y_t versus $W_{t,(h)}$	3.9611	3.8097	14.6723	4.0108	3.7839	14.5875	4.4731	3.6742	14.7775
Out-of-sample corrected Y_t versus $\widehat{Y}_t$	1.6812	1.1953	4.2094	2.1675	1.7608	6.1404	3.5127	3.0179	10.6064
Out-of-sample reduction (%) $W_{t,(h)}$ versus $\widehat{Y}_t$	57.56	68.62	71.31	45.96	53.47	57.91	21.47	17.86	28.23

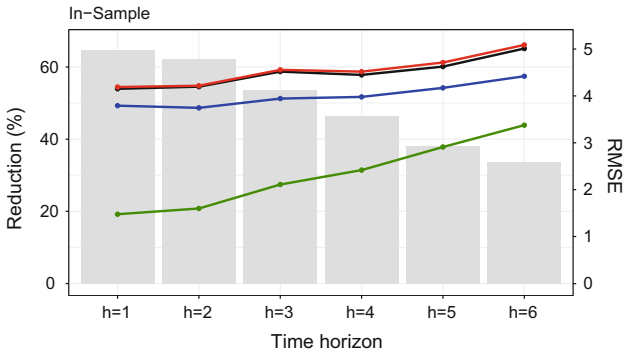

Fig. 8 Evaluation measures in the whole period (baseline) and in the training period: black: observations versus initial forecasts in the whole period (Y_t vs. $W_{t,(h)}$); red: observations versus initial forecasts in the training period (Y_t vs. $W_{t,(h)}$); blue: observations versus initial forecasts without outliers in the training period (Y_t vs. $W^*_{t,(h)}$); green: observations versus calibrated forecasts in the training period (Y_t vs. $\widehat{Y}_t$) and bar chart: reduction (%) of the RMSE in the training period $W_{t,(h)}$ versus $\widehat{Y}_t$

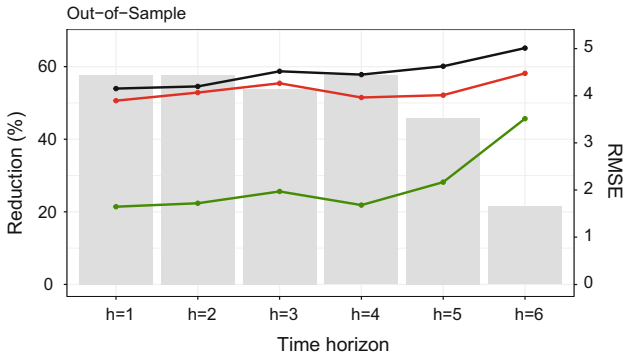

Fig. 9 Evaluation measures in the whole period (baseline) and in the test period: black: observations versus initial forecasts in the whole period (Y_t vs. $W_{t,(h)}$); red: observations versus initial forecasts in the test period (Y_t vs. $W_{t,(h)}$); green: observations versus calibrated forecasts in the test period (Y_t vs. $\widehat{Y}_t$) and bar chart: reduction (%) of the RMSE in the test period $W_{t,(h)}$ versus $\widehat{Y}_t$

in comparison with the website's initial forecasts. However, a slight improvement is noted for the corrected 4-steps forecasts compared to the corrected forecasts with $h = 1, 2, 3$ in the test period. Furthermore, as the time horizon of the predictions increases, the reduction in evaluation measures tends to be greater in out-of-sample than in-sample, except for $h = 6$. The good performance of the correction model for the 4-day time horizon

compared to $h = 2, 3$ is largely due to the improved quality of the original website's forecasts for this time horizon. All the RMSE, MAE and MAPE of the website's forecasts are lower for $h = 4$ relative to the values for $h = 2, 3$.

Figure 10 shows the observed maximum temperature (in black), the website's h-steps ahead forecasts (in blue), and corrected h-steps ahead forecasts (in red) in the test period (out-of-sample) for $h = 1, 6$. Through graphical analysis, it is noted that the corrected forecasts at $h = 6$ show greater variability than the corrected forecasts at $h = 1$, and therefore are more accurate.

5 Conclusions

The present study aims to study and analyze the observed maximum temperature and the respective forecasts for different time horizons from 1 to 6 days obtained from the https://weatherstack.com/ website. Improving the website's forecasts by combining accurate data from a portable station to minimize the forecasts' quality measures, allows obtaining more accurate data that will serve as inputs to other mathematical models. In particular, corrected forecasts will be considered in the optimization models to better manage the availability of water for irrigation within the paradigm of sustainability, as advocated in the TO CHAIR project.

For this purpose, a model with a state-space representation was introduced to model and forecast the daily maximum temperature to improve the forecasts for the

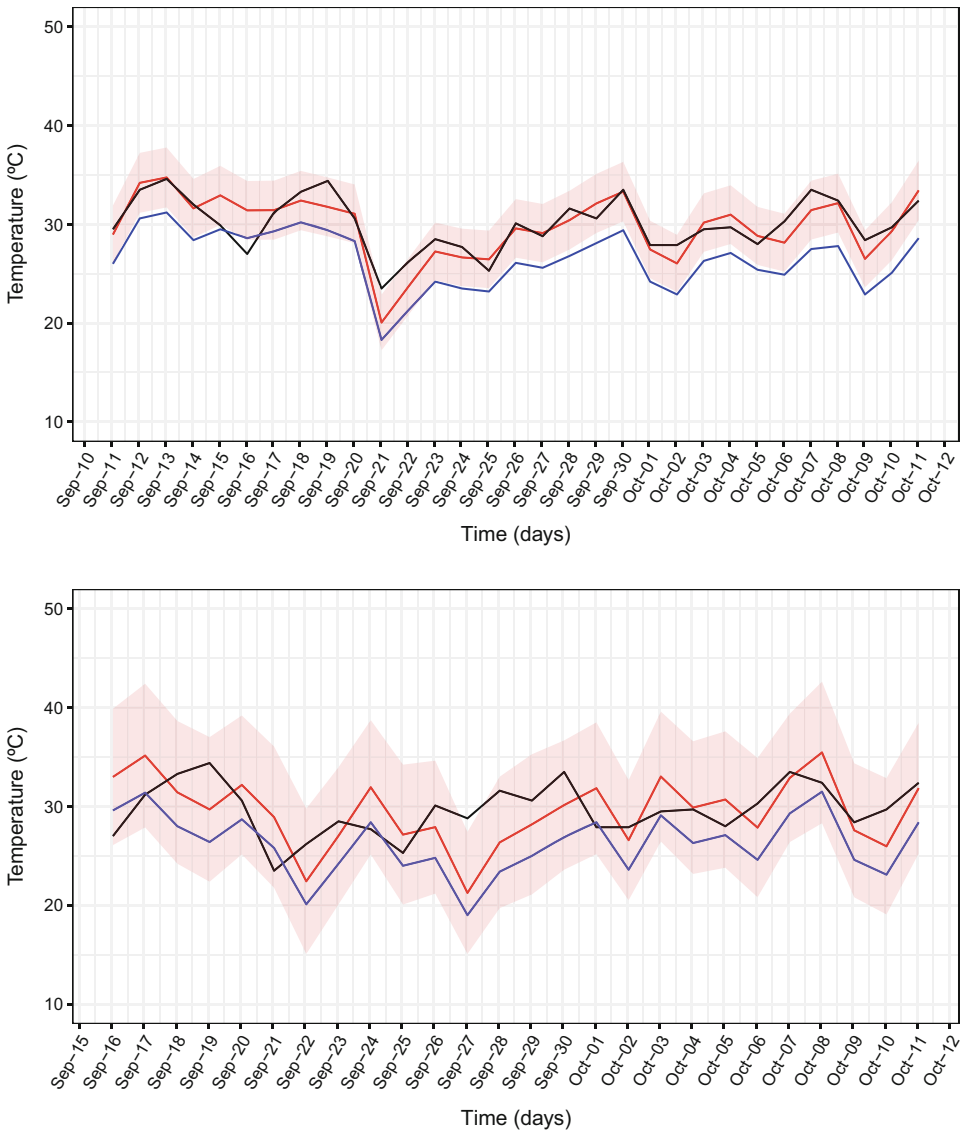

Fig. 10 Observed maximum temperature (in black), h-steps ahead forecasts (in blue), and corrected h-steps ahead forecasts (in red) in the test period (out-of-sample); top: $h = 1$, bottom: $h = 6$. Shadow: 95% confidence intervals

farm, for each $h = 1, \ldots, 6$ days, and was evaluated using evaluation measures such as RMSE, MAE, and MAPE. The model's parameters were obtained using the maximum likelihood method. The normality assumption was verified for all models; the independence assumption was rejected for two models ($h = 4, 6$).

Overall, the proposed model significantly reduced the RMSE, MAE, and the MAPE both in-sample and out-of-sample compared to the website's initial forecasts, where it was observed that these forecasts underestimated, on average, the observed maximum temperature by about 13–14%.

Although the confidence intervals of the corrected 6-steps ahead forecasts had higher ranges, the point forecasts proved accurate. It was also found that the 4-steps ahead corrected forecasts were more accurate than the 1, 2 and 3-steps ahead corrected forecasts in the test period (out-of-sample), although the models with $h = 1, 2, 3$ showed better fit, as evidenced by the BIC values. But sometimes a model with a better goodness of fit does not entail it will give the best forecasts. In addition, the out-of-sample reduction was greater than the in-sample reduction for the models with $h = 3, 4, 5$, thus indicating that these models perform better in terms of forecasting.

Overall, improved forecasts will lead to improved use of water resources, namely by planning irrigation more efficiently.

Acknowledgements This work has received funding from FEDER/COMPETE/NORTE 2020/POCI/FCT funds through Grants UID/EEA/-00147/20 13/UID/IEEA/00147/006933-SYSTEC project and To CHAIR - POCI-01-0145-FEDER-028247. A. Manuela Gonçalves was partially financed by Portuguese Funds through FCT (Fundação para a Ciência e a Tecnologia) within the Projects UIDB/00013/2020 and UIDP/00013/2020 of CMAT-UM. Marco Costa was partially supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT—Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020. F. Catarina Pereira was financed by national funds through FCT (Fundação para a Ciência e a Tecnologia) through the individual PhD research Grant UI/BD/150967/2021 of CMAT-UM.

Funding Funding was provided by POCI-01-0145-FEDER-028247.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

References

Adedotun AF, Latunde T, Odusanya OA (2020) Modelling and forecasting climate time series with state-space model. J Niger Soc Phys Sci 2:149–159. https://doi.org/10.46481/jnsps.2020.94

Aghelpour P, Mohammadi B, Mehdizadeh S, Bahrami-Pichaghchi H, Duan Z (2021) A novel hybrid dragonfly optimization algorithm for agricultural drought prediction. Stoch Environ Res Risk Assess 35:2459–2477. https://doi.org/10.1007/s00477-021-02011-2

Allan R, Pereira L, Smith M (1998) Crop evapotranspiration-guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, vol 56

Alzghool R, Lin Y-X (2008) Parameters estimation for SSMs: QL and AQL approaches. IAENG Int J Appl Math 38:34–43

Berndt C, Haberlandt U (2018) Spatial interpolation of climate variables in northern Germany-influence of temporal resolution and network density. J Hydrol Reg Stud 15:184–202. https://doi.org/10.1016/j.ejrh.2018.02.002

Chatfield C, Xing H (2019) The analysis of time series: an introduction with R, 7th edn. Chapman and Hall/CRC, Boca Raton

Citakoglu H (2021) Comparison of multiple learning artificial intelligence models for estimation of long-term monthly temperatures in Turkey. Arab J Geosci 14(2131):1–16. https://doi.org/10.1007/s12517-021-08484-3

Costa M, Alpuim T (2010) Parameter estimation of state space models for univariate observations. J Stat Plan Inference 140(7):1889–1902. https://doi.org/10.1016/j.jspi.2010.01.036

Costa M, Alpuim T (2011) Adjustment of state space models in view of area rainfall estimation. Environmetrics 22:530–540. https://doi.org/10.1002/env.1064

Costa M, Monteiro M (2016) Bias-correction of Kalman filter estimators associated to a linear state space model with estimated parameters. J Stat Plan Inference 176:22–32. https://doi.org/10.1016/j.jspi.2016.04.002

Costa M, Monteiro M (2019) A periodic mixed linear state-space model to monthly long-term temperature data. Environmetrics 30(5):1–15. https://doi.org/10.1002/env.2550

Durbin J, Koopman S (2001) Time series analysis by state space methods. Oxford University Press, Oxford

Glynis KG, Iliopoulou T, Dimitriadis P, Koutsoyiannis D (2021) Stochastic investigation of daily air temperature extremes from a global ground station network. Stoch Environ Res Risk Assess 35:1585–1603. https://doi.org/10.1007/s00477-021-02002-3

Gonçalves AM, Costa M (2013) Predicting seasonal and hydro-meteorological impact in environmental variables modelling via Kalman filtering. Stoch Environ Res Risk Assess 27:1021–1038. https://doi.org/10.1007/s00477-012-0640-7

Gonçalves AM, Costa M, Costa C, Lopes S, Pereira R (2021) Temperature time series forecasting in the optimal challenges in irrigation (TO CHAIR). https://doi.org/10.1007/978-3-030-57422-2_27

Harvey AC (2009) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, New York

Hyndman RJ, Athanasopoulos G (2018) Forecasting: principles and practice, 2nd edn. OTexts, Melbourne

Kalman R (1960) A new approach to linear filtering and prediction problems. ASME J Basic Eng 82:35–45. https://doi.org/10.1115/1.3662552

Kitagawa G (2021) Introduction to time series modeling with applications in R, 2nd edn. Chapman and Hall/CRC, Boca Raton. https://doi.org/10.1201/9780429197963

Kokic PN, Crimp S, Howden M (2011) Forecasting climate variables using a mixed-effect state-space model. Environmetrics 22:409–419. https://doi.org/10.1002/env.1074

Ksiksi T, Al Blooshi L (2019) Climate change in the UAE: modeling air temperature using ARIMA and STI across four bio-climatic zones. F1000Research 8:973. https://doi.org/10.12688/f1000research.19557.1

Leadbetter MR, Lindgren G, Rootzén H (1983) Maxima and minima and extremal theory for dependent processes. Springer, New

York, pp 205–215. https://doi.org/10.1007/978-1-4612-5449-2_11

Lindgren N, Olsson H, Nyström K, Nyström M, Ståhl G (2021) Data assimilation of growing stock volume using a sequence of remote sensing data from different sensors. Can J Remote Sens. https://doi.org/10.1080/07038992.2021.1988542

Moritz S, Sardá A, Bartz-Beielstein T, Zaefferer M, Stork J (2015) Comparison of different methods for univariate time series imputation in R. https://doi.org/10.48550/ARXIV.1510.03924

Navon IM (2009) Data assimilation for numerical weather prediction: a review. Springer, Berlin, Heidelberg, pp 21–65. https://doi.org/10.1007/978-3-540-71056-1_2

Petris G, Petrone S, Campagnoli P (2009) Dynamic linear models with R.useR! Springer, New York

Prokosch J, Bernitz Z, Bernitz H, Erni B, Altwegg R (2019) Are animals shrinking due to climate change? Temperature-mediated selection on body mass in mountain wagtails. Oecologia 189:841–849. https://doi.org/10.1007/s00442-019-04368-2

Rangapuram SS, Seeger M, Gasthaus J, Stella L, Wang Y, Januschowski T (2018) Deep state space models for time series forecasting. In: Advances in neural information processing systems, vol 31. NIPS'18. Curran Associates Inc., Red Hook, pp 7796–7805

Rodrigues GC, Braga RP (2021) A simple procedure to estimate reference evapotranspiration during the irrigation season in a hot-summer Mediterranean climate. Sustainability. https://doi.org/10.3390/su13010349

Rodríguez A, Ruiz E (2012) Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters. Comput Stat Data Anal 56(1):62–74. https://doi.org/10.1016/j.csda.2011.07.010

Seo D, Saifuddin MM, Lee H (2018) Conditional bias-penalized Kalman filter for improved estimation and prediction of extremes. Stoch Environ Res Risk Assess 32:183–201. https://doi.org/10.1007/s00477-017-1442-8

United Nations (2021) The United Nations world water development report 2021: valuing water. UNESCO, Paris

Woody J, Lu Q, Livsey J (2020) Statistical methods for forecasting daily snow depths and assessing trends in inter-annual snow depth dynamics. Environ Ecol Stat 27:609–628. https://doi.org/10.1007/s10651-020-00461-5

You D, Hunter M, Chen M, Chow SM (2020) A diagnostic procedure for detecting outliers in linear state-space models. Multivar Behav Res 55(2):231–255. https://doi.org/10.1080/00273171.2019.1627659