

Article

Interpolation of Instantaneous Air Temperature Using Geographical and MODIS Derived Variables with Machine Learning Techniques

Marcos Ruiz-Álvarez ^{1,†} , Francisco Alonso-Sarria ^{1,*,†}  and Francisco Gomariz-Castillo ^{1,2,†} 

¹ Instituto Universitario de Agua y Medio Ambiente, Campus de Espinardo, Universidad de Murcia, 30001 Murcia, Spain

² Instituto Euromediterráneo del Agua, Campus de Espinardo, 30001 Murcia, Spain

* Correspondence: alonsarp@um.es; Tel.: +34-868-888-695

† These authors contributed equally to this work.

Academic Editor: Wolfgang Kainz

Received: 5 August 2019; Accepted: 29 August 2019; Published: 31 August 2019



Abstract: Several methods have been tried to estimate air temperature using satellite imagery. In this paper, the results of two machine learning algorithms, Support Vector Machines and Random Forest, are compared with Multiple Linear Regression and Ordinary kriging. Several geographic, remote sensing and time variables are used as predictors. The validation is carried out using two different approaches, a leave-one-out cross validation in the spatial domain and a spatio-temporal k-block cross-validation, and four different statistics on a daily basis, allowing the use of ANOVA to compare the results. The main conclusion is that Random Forest produces the best results ($R^2 = 0.888 \pm 0.026$, Root mean square error = 3.01 ± 0.325 using k-block cross-validation). Regression methods (Support Vector Machine, Random Forest and Multiple Linear Regression) are calibrated with MODIS data and several predictors easily calculated from a Digital Elevation Model. The most important variables in the Random Forest model were satellite temperature, potential irradiation and cdayt, a cosine transformation of the julian day.

Keywords: air temperature; MODIS; machine learning; interpolation

1. Introduction

Air temperature (T_a) is a very relevant climatic variable that controls several environmental processes, particularly evapotranspiration [1]; it is also a key feature in global change studies, and reflects the surface energy balance [2]. So, accurate estimations of T_a and its spatio-temporal variability are important in several Earth and environmental sciences and in land surface process modelling [1,3]. Air temperature is usually measured at weather stations at a standard height of 1.5–2 m with diverse temporal resolutions. However, because weather stations provide limited information about spatial patterns at regional or global scales [3], several methods have been used to estimate the spatial distribution of T_a [4,5]:

- Vertical lapse methods [4] use height as the main variable to explain temperature spatial distributions. The vertical lapse rate is evaluated from the sampling data and then applied to the whole study area. A more sophisticated approach uses daily atmospheric profiles provided by the Moderate Resolution Imaging Spectroradiometer (MODIS) product MOD07_L2 to locally estimate the adiabatic lapse rate [6]. The main drawback of this approach is the coarse spatial resolution (5 km) of those products.

- Simple linear regression using land surface temperature (LST), retrieved by remote sensing, as a predictor for T_a [7]. MODIS, for example, provides global coverage of several environmental variables with large temporal resolution and moderate spatial resolution [6].
- Multiple regression models using LST and other variables such as NDVI, solar zenith angle, solar radiation, altitude, julian day, distance to the coast, normalised difference water index (MNDWI) or albedo as predictors [8–15]. The algorithms used range from multiple linear regression [5,10,11,16] to more sophisticated machine learning algorithms, such as neural networks [9] or Random Forest (RF) [14].
- Geostatistical techniques (kriging) [17,18] estimate T_a as a weighting average of the sampling points with the weighting coefficients obtained after a statistical analysis of the spatial variability of the variable (semivariogram functions). The main drawbacks of such interpolation methods are that they do not use covariates and that they may be sensitive to a clustered distribution of weather stations [19].
- The Temperature-Vegetation Index (TVX), proposed by [20,21], is based on the correlation between NDVI and LST, assuming that T_a is approximately equal to LST in fully vegetated areas. Significant uncertainties appear in sparse vegetation areas [22].
- Methods based on the surface energy balance such as ADEBAT [4,13]. The objective is to estimate T_a using a more physical approach. It has two main drawbacks: several variables that can only be measured in weather stations are needed and, as the Bowen ratio (the ratio between the sensible heat flux and the latent heat flux) is one of them, it is necessary to know the latent heat flux (LE) to use ADEBAT. However, T_a is usually, as in this case, estimated in order to estimate LE from it, so the use of a surface energy balance is not suitable.

Remote sensing methods are constrained by the time of the day when images are taken. During the night, LST is a very accurate proxy for T_a as solar radiation has no effect, simplifying the ground surface energy balance. During the day, it is necessary to take into account several variables, such as cloud cover, wind speed, soil moisture and surface roughness, which remote sensors cannot retrieve [6]. However, as remote sensing provides LST spatially distributed estimations, they may be used as highly correlated predictors to estimate air temperature. Tao et al. [23] showed that correlation between LST at night and minimum T_a is higher ($R^2 = 0.93$) than between their daytime equivalents ($R^2 = 0.79$).

Several approaches have been recently used to estimate air temperature from satellite data; one of the objectives being to use just remote sensing data adding, at most, easy to obtain environmental variables. Jin and Dickinson [24] estimated the diurnal cycle of LST through temporal interpolation of NOAA-AVHRR. Sun et al. [25] used the same method for modeling Geostationary Operational Environmental Satellite (GOES) measurements. Gholamnia et al. [26] attempted to introduce methods that estimate air temperature based on LST data by direct approach and without auxiliary environmental information. Golkar et al. [27] estimate instantaneous air temperature using standard remote sensing methods with minimum data. Mira et al. [28] tried to introduce LST and elevation as predictors to a multiple linear regression model (MLRM). Recently, machine learning methods have been compared with traditional MLRM; for instance, Xu et al. [14] stated that linear regression cannot predict the air temperature based on LST in all conditions and that RF models obtain better accuracy than the MLRM.

As a result of these attempts, several global weather products have been produced that estimate variables using satellite imagery. For instance, Hooker et al., [29] estimated a monthly temperature dataset at a spatial resolution of 0.05° for the interval 2003–2016.

We think that more research is needed to investigate how the use of satellite information and different geographic predictors to calibrate machine learning algorithms might outperform traditional interpolation methods. Thus, the objective of this work is to use two well known machine learning algorithms, Support Vector Machines (SVM) and RF, to estimate T_a at the AQUA passage time (between 12:00 and 14:20) in a medium size basin in SE Spain. We used as predictors a set of environmental

variables included in some of the previous works that we think would influence the temperature spatial patterns. The results are compared to those obtained with more traditional approaches: MLRM and Ordinary Kriging (OK). Finally, regression-kriging will be tested using OK to interpolate the residuals of the three regression models, provided that such residuals show spatial autocorrelation. Two cross-validation approaches were carried out to deal with accuracy over-estimation induced by spatio-temporal autocorrelation: a simple leave-one-out cross-validation (LOO-CV) in the spatial domain and a spatio-temporal k-block cross-validation (k-block-CV).

The use of both approaches allows to evaluate accuracy in two different situations: spatio-temporal k-block-CV provides an accuracy estimation for models calibrated with a large observations set but avoiding the overfitting due to spatio-temporal dependence. On the other hand, LOO-CV in the spatial domain produces an accuracy estimation when only spatial but not temporal dimensions are considered (models are generated using observations for just one day, an observation is extracted in each iteration to estimate the model error on it).

2. Methodology

Figure 1 shows a flow chart summarising the methodology, which will be detailed below. All the research was carried out using free and open source software. For step 1, that includes downloading and preprocessing of information, GRASS [30], a multi-purpose Geographical Information System, was used. Step 2, that includes model calibration, validation and the generation of raster predictions, was carried out using the R-CRAN program [31].

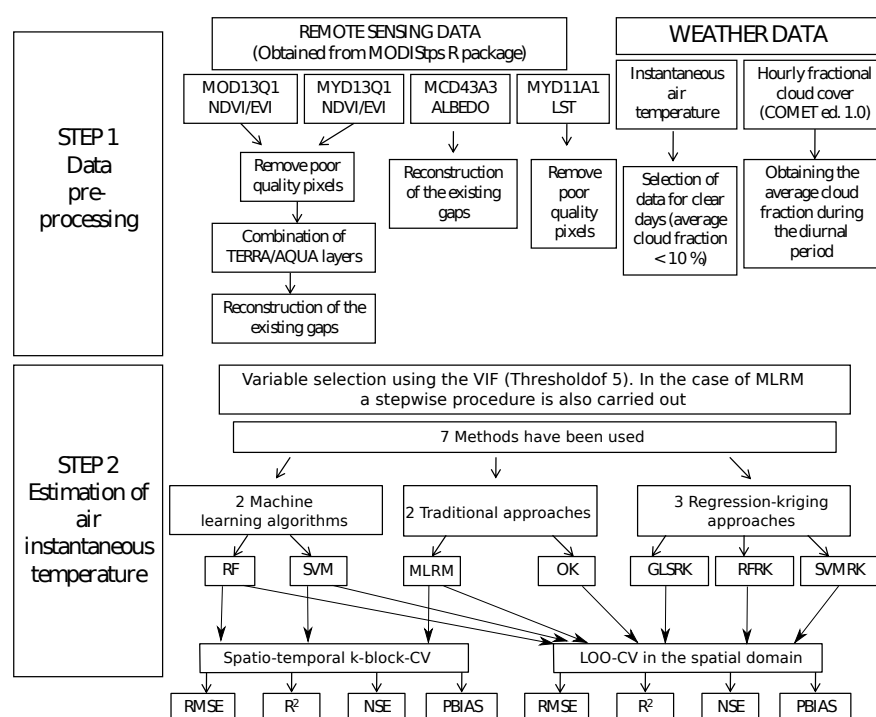


Figure 1. Flow diagram of the methodology.

2.1. Study Area

This research was carried-out in the area controlled by the River Segura Water Authority (DHS in its Spanish abbreviation) (Figure 2), which includes the Segura river basin (19,000 km²) and several minor coastal basins in SE Spain. It is a semiarid area with scarce and irregular precipitation, high temperatures, and a large number of hours of sun that cause high potential evapotranspiration. Despite the scarcity of water, agriculture is an important economic sector using both groundwater

(available because the predominance of carbonate rocks) and water transferred from other basins. Population density and intensive irrigated agriculture represent a significant water demand.

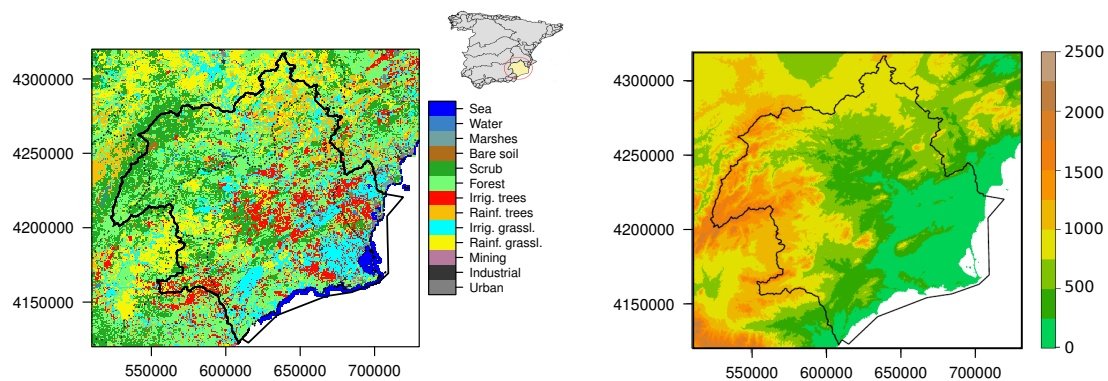


Figure 2. Land uses (left) and elevation (right) of the study area. Solid lines represent DHS boundaries and dotted lines province boundaries.

The study area is also characterised by substantial height differences over short distances, which together with the semiarid climate and the use (limited in space) of groundwater and transferred water, create a strong environmental variability.

The spatial distribution of temperature is strongly influenced by height distribution, the annual average temperature increases gradually from the Northwest mountain ranges (10 °C) to the coast (18 °C). The annual temperature regime has a winter low in December and January, while the highest values occur in July and August [32]. The proximity to the sea softens summer and winter temperatures in the coast, while an increase in continental features is observed towards the northern area, both because of the distance to the sea and because of the presence of successive mountain alignments [33].

The cloudiness in the DHS is more abundant in the equinoctial seasons, since in winter and summer long periods with anticyclonic situations are common. The northwestern sector has the highest cloud cover of the DHS, as it is exposed to the Atlantic fronts and suffers frequent convective storms in the May–September period, while coastal areas experience less cloudiness, especially the southwestern coast, although is not infrequent the presence of coastal mist. Thus, the average cloud fraction varied over the 2012–2014 period in much of the DHS between 20–30% in summer, between 30–40% in winter and between 40–50% in spring and autumn. In the north-west, such numbers are increased by 10 %. With respect to clear days (those with an average cloud fraction of less than 10%), in summer they range from 30 in the north-western extreme to 60 in coastal areas. In spring and autumn the number of clear days presents a greater homogeneity in the DHS, oscillating between 15 and 25 days. Finally, in winter, clear days range from 20 to 30 in the study area, with slightly more than 30 days in coastal areas.

2.2. Data Set

Different variables were used as predictors: (1) geographical variables: longitude, latitude, altitude, Topographic Wetness Index (TWI) [34] (used to describe the potential accumulation of cold air), monthly potential irradiation ($\text{Wh m}^{-2} \text{ month}^{-1}$) obtained from heights using the method proposed by [35] and distance from the coast; (2) time variables: day duration, cdayt (a cosine transformation of julian day following [36]), and the satellite passage time; and (3) variables provided by the MODIS sensor: EVI, albedo and the terrestrial surface temperature (LST) at the passage time of AQUA satellite.

The objective was to have a set of easy to obtain predictors that are, at the same time related with processes that explain air temperature. Latitude and altitude relation with temperature is well known; longitude was included to complete the 3 spatial axes. TWI estimates the accumulation of water in the

terrain contributing to a cooling of the air. Monthly potential irradiation accounts for the differences in the amount of solar radiation reaching land surface due to differences in slope and aspect. LST is surface temperature as estimated by the satellite. As LST will vary in relation with the time when the satellite takes the image, this time needs to be taken into account. Cdayt is a transformation of the julian day to guarantee smooth transition between December and January; it oscillates between -1 (summer) and 1 (winter). Albedo measures the proportion of incident radiation reflected to the atmosphere. Distance to the coast is a proxy for continentality. EVI is a vegetation index useful as vegetation moderates extreme temperatures. Finally, the duration of the day increase the daily heat accumulation contributing to increase air temperature.

The classical formulation of TWI [37] is:

$$TWI = \log\left(\frac{a}{\tan(b)}\right) \quad (1)$$

where a is the upslope contributing area per unit contour and b is the slope. A more detailed description of methods to calculate slope and contributing area from a DEM can be found in [38].

The equation to calculate cdayt is:

$$cdayt = \cos\left([t_D - \phi] \cdot \frac{2\pi}{365}\right) \quad (2)$$

where t_D is the julian day and ϕ is a time delay from the coldest day.

2.2.1. Weather Data

Air temperature data every 30 min is recorded by 53 weather stations belonging to the SIAM and SIAR networks (*Sistema de Información Agrometeorológico de la Región de Murcia*, Agro-meteorological information System in Murcia Region) and *Sistema de Información Agroclimática para el Regadío*, Agro-meteorological Information System for Irrigation) (Figure 3). In this work, only data for 2012 were analysed to test which model produces the most accurate results.

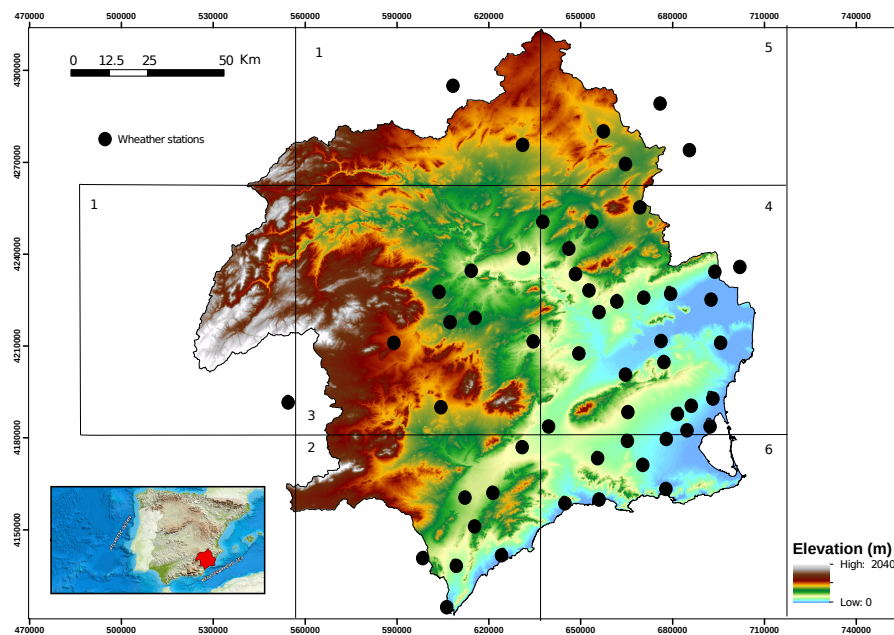


Figure 3. Location of weather observatories in the study area. Blocks for the k-block-CV are also shown.

We assumed that the mean air temperature data (obtained from the average of 3 measurements taken every 10 min) provided by the SIAR and SIAM networks are equal to the air temperature at the

time of satellite passage. Therefore, these data were used to calibrate and validate the different models used in this work for the prediction of the instantaneous air temperature.

Only data from clear days were used. Since there is no consensus in the scientific literature on the definition of a clear day, we considered as such those days with an average cloud cover lower than 10%.

2.2.2. Remote Sensing Data

The three environmental variables used as predictors in this work (albedo, vegetation index and surface temperature) were obtained from different MODIS products (Table 1). The R packages MODISTsp [39] package was used to download and process the images.

Table 1. Main characteristics of MODIS products used in this work.

Sensor	Platform	Layer	Name	Spatial Res. (km)	Temporal Res. (days)
MODIS	Aqua	LST_{day}, LST_{night}	MYD11A1	1	1
MODIS	Terra + Aqua	Albedo, α	MCD43A3	0.5	1 (16 days average)
MODIS	Terra	EVI	MOD13Q1	0.25	16
MODIS	Aqua	EVI	MYD13Q1	0.25	16

The 8 day delay between the MOD13Q1 and MYD13Q1 products allows to obtain layers of both EVI combining AQUA and TERRA satellites with a temporal resolution of 8 days. The process involves 3 steps: (1) Elimination of pixels with very low quality or with large observation errors. To this end, according to [40] all pixels with a value greater than 5 in the VI usefulness index (VIUI) [41,42] were removed. (2) Combination of the EVI layers obtained by the MOD13Q1 and MYD13Q1 in those pixels whose values of both products were available. (3) Filling the gaps after step 2. If EVI values one week before and one week after were available, the average of both layers was used. Otherwise, the values estimated one week earlier or later were used. If no EVI values were available for the prior or the posterior weeks, the same process were carried out with the estimated values for 2 weeks before and after.

With regard to the LST layers, we decided to use those provided by the product MYD11A1, since AQUA passage time is closer than TERRA's to the temperature recording time. The LST layers were corrected taking into account the quality of the different pixels. In this case, according to [43] the LST pixels with an error estimation of more than 3 degrees Kelvin were removed. The GRASS module `i.modis.qc` was used to obtain the error layers. We also used the MODIS White-Sky Albedo layer as albedo estimation.

The hourly cloud fraction layers available in the database CMSAF-COMET Edition 1 [44] were used to obtain layers of the mean cloud fraction during the daytime period. This raster database (resolution of 0.05°) collects information on the fraction of cloudiness obtained from METEOSAT satellites, with an hourly time resolution. With the objective of giving greater importance to the central hours of the day, when the irradiation is concentrated, the daily layers were obtained as a weighting average of the hourly layers. The cosine of the angle of the solar zenith (β), proportional to the radiation received at any given time, was used; for the nocturnal layers $w = 0$ and for the diurnal layers $w = \cos \beta / \sum(\cos \beta)$.

2.2.3. Geographical Variables

Some of the geographical variables (elevation, distance to the sea, potential insolation and TWI) used as predictors with the regression methods were calculated from the official 25 m resolution Digital Elevation Model (DEM) downloaded from the Spanish National Geographical Institute (*Instituto Geográfico Nacional, IGN*) website (<http://centrodedescargas.cnig.es/CentroDescargas/index.jsp>).

2.3. Estimation Methods

Four methods were used to estimate air temperature at the satellite passage time: MLRM, SVM and RF are regression (global interpolation) methods that rely on predictor variables to estimate the dependent variable; in this paper, only satellite data and other predictors that can be easily obtained were used, the first based on classical statistics and the last two on machine learning; finally OK, a local interpolation method that does not take into account predictors, was used for comparison. Finally, the residuals of the three regression models were interpolated with OK to obtain regression-kriging models (GLSRK, SVMRK and RFRK).

2.3.1. Multiple Linear Regression

Simple and multiple linear regression models (MLRM) are the most popular models for estimating air temperature. MLRM is a global interpolation method in which a functional relationship is defined between the dependent variable (in this case maximum and minimum temperature estimated in different observatories in the territory), and a set of spatially distributed environmental and geographical variables. The parameters may be estimated using ordinary least squares (OLS) or generalised least squares (GLS), a modification of OLS that takes into account the heterocedasticity and the spatial correlation in the observations. In this work, we used the implementation of GLS in the R package nlme [45].

The assumptions of the model were assessed by hypothesis contrasts: the Kolmogorov-Smirnov test to assess the normality of residuals (usually met) and the Breush-Pagan test to assess homocedasticity (usually not met).

With all the regression methods, a first step of variable selection was carried out using the Variance Inflation Factor (VIF) methodology proposed by [46]. In the case of MLRM a subsequent stepwise procedure was carried out to minimise the number of variables to provide a more parsimonious model.

2.3.2. Support Vector Machines

SVM [47] was originally developed for classification but it was adapted to regression as a robust method that tries to minimise the effect of outliers. SVM for regression is described in [48]. Instead of trying to minimise the sum of squared errors, data points whose residual absolute values are lower than a user defined threshold (ϵ) do not contribute to the fit, whereas points with $|e| > \epsilon$, the so called support vectors, contribute linearly rather than quadratically to the error objective function to be minimised. This somewhat counterintuitive approach (the more accurately predicted points are not used to fit the line) has proven effective. A detailed description of this method can be found in [48] or [49].

SVM has been reported to obtain similar accuracy as RF and better accuracy than other machine learning methods such as neural networks [50–52]; however, its main drawback is that there is no way to know in advance which kernel and parameter values will give the best results. In this case, we reduced the problem by using a radial basis function (RBF) kernel. We optimised parameters C and σ simultaneously using a grid search method, 10-fold cross validation with the training data and mean squared error as performance measure. The implementation used was that of the R package e1071 [53].

2.3.3. Random Forest

RF [54] is an ensemble of decision trees. Decision trees are characterised by a small bias but a high variance. RF tries to solve this issue by training several (500 as default) decision trees. Each tree is trained with a bootstrapped subsample of the available training cases. In addition, every time a variable has to be selected to split a node, only a subset of the independent variables is considered (by default the integer part of \sqrt{p} where p is the number of variables). In this way, although each tree provides a high variance estimation, averaging the results of all of them will result in a low bias and low variance estimation.

RF has 4 main advantages over other machine learning methods: (1) It provides an internal cross-validation procedure, (2) the default values for the parameters provide optimal estimations most of the times, (3) the decrease in heterogeneity provided by each variable along the calibration process of each tree, when aggregated, provides a measure of the importance of each variable, (4) it is possible to obtain an estimation of the effects of the different predictors on the model, allowing the operator to decide if such effects are physically sound or not. Points 3 and 4 mean that RF is not really a black box model, as other ML techniques; rather it might be considered a grey box model.

In this work, we used the version in the R ranger package [55], a very fast and memory efficient implementation of RF. We used the default values for the *ntree* (500) and *mtry* ($\text{floor}(\sqrt{p})$) where p is the number of predictors. Previous research [56,57] has shown that the accuracy achieved with such parameters is usually near the optimum.

2.3.4. Ordinary Kriging

OK is a local interpolation method based on the regionalized variable theory [58]. It uses solely the values measured in the observation points and their location. Its main advantage over other local interpolation methods (such as Inverse Distance Weighting) is that a statistical analysis of the spatial variability of the values is previously performed and summarised in the semivariogram function. Finally, OK performs, at each pixel, a weighted average of the values in the surrounding observation points with the weights calculated as a function of the semivariogram. The assumptions of OK are normality and first and second order stationarity, that is, the mean and the variance are constant in the area. These assumptions are rarely met, so several variations have been proposed to deal with trends in the data: e.g. Universal Kriging (taking into account spatial trends in the values) and regression-kriging (taking into account other predictors).

In this work, we used the R package *automap* [59], its main advantage being the automatization of a weighted least squares optimal estimation of semivariogram parameters using Gauss-Newton [60]. The variable was log-transformed to obtain a normal variable.

2.3.5. Regression-Kriging

A different approach is to perform a global (regression) interpolation, and interpolate its residuals with a local (e.g. kriging) method. In this way the local variability not explained by the global model and accounted in their residuals is dealt with by the local method.

Universal kriging is a well known method that firstly fits with a linear model a global surface using coordinates as predictors and secondly interpolates the residuals with OK [61]. Regression-kriging [62,63] is the same approach but using any environmental variables as predictors. Obviously any global method (GLS, RF, SVM, etc.) can be used in the regression part and any local interpolation method can be used instead of kriging to deal with the residuals. In this work we used OK to interpolate the residuals of MLRM, SVM and RF.

2.3.6. Validation

As the aim of this work is to evaluate the predictive performance of different models, cross validation was carried out for the seven estimation methods. Two types of problem have been raised about the use of cross-validation with spatio-temporal data when the objective is to extrapolate outside the spatio-temporal structure. Firstly, error estimates in a random cross validation may be too optimistic [64]. Secondly, time-static spatially distributed predictors (such as elevation) are prone to over-fit models [65].

In order to solve the spatial autocorrelation problem, it is advisable to validate with cases that are far enough from the training data to decrease spatial dependence and avoid too optimistic error estimates [64,66]. Something similar occurs with time-correlated data [64].

To carry out cross-validation, we have considered two approaches. Firstly a simple leave-one-out CV (LOO-CV) in the spatial domain, in which each case is used to validate a model calibrated with

temperature data from the other observatories in the same day. The next pseudocode shows how it works:

```

for each location i {
  for each time t {
    calibrate a model with D[-i,t]
    estimate Tair in D[i,t]
    compute error[i,t] as Estimation - Real
  }
}

```

Secondly, a block cross-validation strategy [67]. Blocks are similar to folds in k-fold cross-validation, but, instead of the cases in each block being randomly selected (as in folds), all cases in a block correspond to a spatio-temporal subset of the 3D space formed by the 2D space and time. Spatial blocks forces testing on more spatially distant records, thus decreasing spatial dependence and reducing optimism in error estimates [66]; temporal autocorrelation is functionally similar, and blocks may be drawn in the same manner to better ensure independence between cross-validation folds [64]. The size of such blocks is determined after an analysis of the spatio-temporal autocorrelation identified in the data studied.

In order to implement k-block-CV, a prior study of the temporal and spatial autocorrelation of the daily instantaneous temperature was carried out. The temporal autocorrelation was analysed using autocorrelation functions (ACF) and partial autocorrelation functions (PACF), and the spatial autocorrelation was analysed using semivariograms. In both cases the objective was to identify the distance (meters) and time lag (days) beyond which autocorrelation ceases to affect. The lag values of the ACF in each station and the ranges of the semivariograms in each day were obtained, allowing to estimate the adequate size for the k-spatial and k-temporal blocks inside each k-spatial block. The number of resulting blocks was the number of spatial multiplied by the number of temporal blocks.

In order to solve the second problem, concerning the predictor selection, a forward feature selection algorithm (FFS) [65] was applied to the variables selected in the VIF analysis. The algorithm is implemented in the R CAST package [65]. It, firstly, trains models of all possible 2-variable combinations, keeping the best model. Then, the number of predictors is iteratively increased and the improvement of the model tested for each additional predictor. The process stops when none of the remaining variables decreases the error of the currently best model [65]. The k-block-CV framework is used in FFS to obtain performance statistics.

Bennett et al., [68] state that Goodness of fit statistics measure different performance aspects, so several statistics should be used to decide which is the most accurate model. Four statistics, whose detailed description and interpretation criteria can be consulted in [69] or [68], were used in this research:

Root mean square deviation (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - E_i)^2}{n}} \quad (3)$$

Coefficient of determination (R-Squared):

$$R^2 = \left(\frac{COV(O, E)}{s_O \cdot s_E} \right)^2 \quad (4)$$

The modified Nash-Sutcliffe efficiency (NSE) measures the relative magnitude of the residual variance compared to the observed data variance, it is less sensitive than R^2 to outliers [69]:

$$NSE = \frac{\sum_{i=1}^n (O_i - E_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (5)$$

NSE ranges from -Inf to 1, this last value indicating a perfect match. If $NSE = 0$, the model is as accurate as using the mean of the observed data as predictor. If $NSE < 0$, the observed mean is a better predictor than the model.

Percent bias (PBIAS) measures the relative tendency of the estimated values to differ from the observed values. It is less sensitive than RMSE to outliers and to the magnitude of the data:

$$PBIAS = \frac{\sum_{i=1}^n (O_i - E_i)}{\sum_{i=1}^n O_i} \quad (6)$$

The optimal value of PBIAS is 0, with positive values indicating overestimation and negative values indicating underestimation.

In Equations (3) to (6), O_i are observed values and E_i estimated values. The four statistics were calculated for both validation methodologies. For LOO-CV in the spatial domain, they were calculated each day, so statistical distributions of their values along the year can be obtained and compared. An alternative option would have been to calculate statistics for each observatory; however, the temperature variations along the year would produce high R^2 and NSE values even if the accuracy of the different observatories were small (Simpson paradox).

As normality and homocedasticity could not be assumed, a robust version of one-way ANOVA using a heteroscedasticity-consistent covariance matrix of the parameters (HC3) method [70] to test whether differences among the methods were significant. If that was the case, a post-hoc contrast between pairs of models, based on Tukey–Kramer contrast and using HC3 method to correct p-values among classes, was performed to discover groups of non-significantly different methods.

3. Results

Two different validation approaches were used in this work, but the main objective was to compare different interpolation methods. For this reason we will present the results comparing the accuracy of the different methods, first with the LOO-CV in the spatial domain approach, and, later, with k-block Cross-Validation.

3.1. Leave-One-Out Cross-Validation in the Spatial Domain

For the regression models (GLS, RF, SVM) a VIF analysis was previously performed to recursively eliminate those variables with a high linear correlation with the rest. Recommended threshold values for VIF range from 5 [71] to 10 [72]. In this work it was firstly established in 5, but it was relaxed to allow the inclusion of LST (VIF = 6.43) as we considered LST, a priori, the most significant predictor, it is near to the lowest recommended value than to the higher recommended value, and the finally eliminated predictors have VIF values way larger: day duration (VIF = 39.04) and latitude (VIF = 24.43).

Both latitude and day duration were filtered out for their high correlation with other variables. Day duration is highly correlated with cdayt, and, due to the NW-SE altitude descent, altitude and longitude make latitude redundant in the study area (see Figure 3).

Table 2 and Figure 4 show the results of the cross-validation of the eight methods. The number of days were 121 with RF, RFRK, SVM, SVMRK and OK, and 118 with GLS and GLSRK. The cross validation was limited to those days in which the number of observations (weather stations without clouds and with data available for that day) was greater than the number of predictors (10). That means 121 days including 3995 predictions. However, in the case of GLS and GLSRK, the final set of predictions was 3928 and 118 days due to errors derived from the estimation of the variance-covariance matrix.

On average, validation statistics indicate that RFRK obtains the most accurate results, followed by RF. In addition, GLS and GLSRK show very high prediction errors, although predictions with absolute residuals larger than 20 °C were filtered out. When this filtering was carried out, the final number was 3886 in the GLS and GLSRK models. In the case of GLS and GLSRK, errors occurred in areas outside the domain and/or range of the data and in areas with local effects that implied situations not considered in the linear relationship estimated in the general model. In these cases, as can be seen

in Figure 4, machine learning methods are more accurate. This difference is due to the fact that the linear models predict poorly outside the boundaries of the training data set, whereas the machine learning methods have a better generalisation capability and predict well outside the boundaries of the training data set [73].

Table 2. Mean and confidence intervals of the four validation statistics calculated for the seven temperature estimation methods and LOO-CV in spatial domain approach. The contrast for PBIAS is done with absolute value.

	GLS	RF	SVM	GLSRK	RFRK	SVMRK	OK
R ²	0.493 ± 0.024	0.558 ± 0.019	0.460 ± 0.018	0.506 ± 0.024	0.612 ± 0.019	0.482 ± 0.018	0.515 ± 0.021
NSE	−0.679 ± 0.43	0.504 ± 0.022	0.405 ± 0.019	−0.620 ± 0.423	0.578 ± 0.025	0.452 ± 0.021	0.468 ± 0.029
RMSE	1.587 ± 0.097	1.181 ± 0.025	1.301 ± 0.025	1.550 ± 0.096	1.068 ± 0.027	1.236 ± 0.023	1.214 ± 0.036
PBIAS	0.009 ± 0.144	0.036 ± 0.038	0.184 ± 0.037	0.060 ± 0.151	−0.172 ± 0.046	0.125 ± 0.038	0.340 ± 0.033

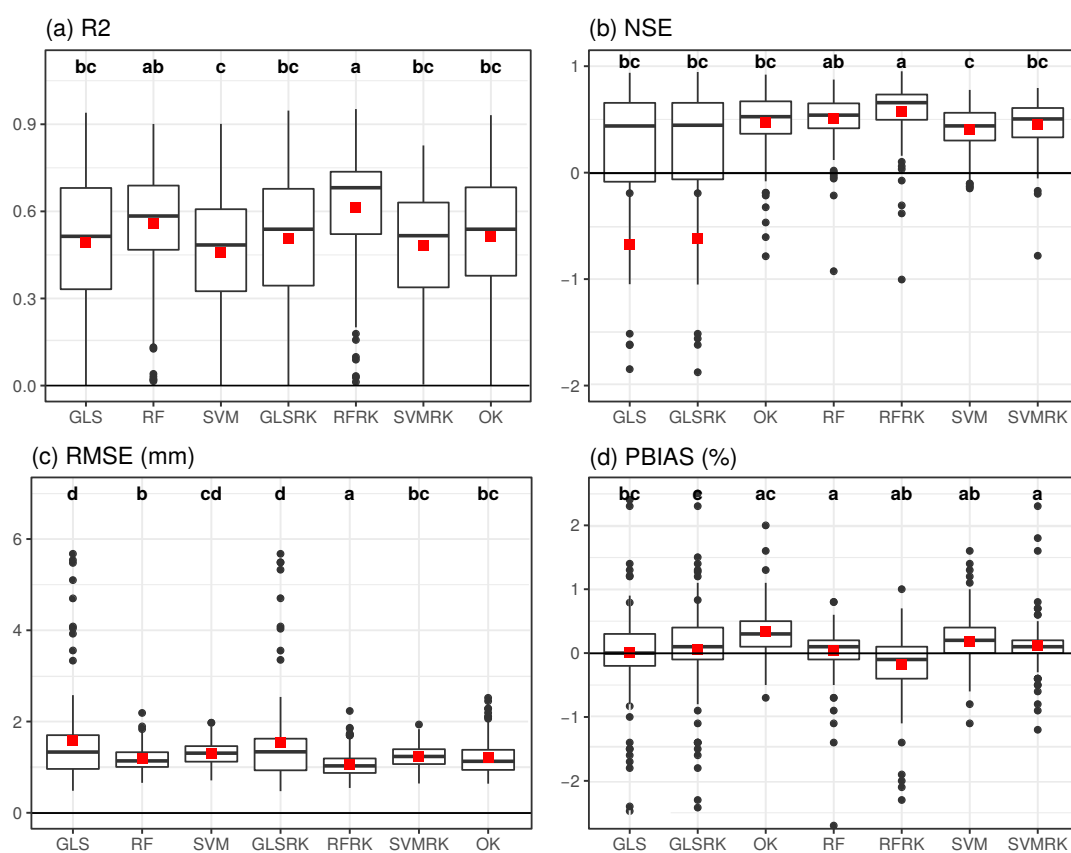


Figure 4. Boxplot showing Anova results of the 4 validation statistics calculated for the seven temperature estimation methods and LOO-CV in spatial domain approach. Significantly different groups (Tukey–Kramer contrast using heteroscedasticity-consistent covariance matrix of the parameters (HC3), alpha = 0.05) are represented by different letters.

The ANOVA shows significant differences among the algorithms for all the statistics with large F values (R²: 18.907, NSE: 17.053, RMSE: 67.698, PBIAS: 26.847) and p-values lower than 0.001 in all cases. Table 2 and Figure 4 show detailed results of such an analysis. The letters above the plots in Figure 4 indicate to which groups of non-significantly different values belong each method. According to NSE and RMSE, RF with kriging of the residuals (RFRK) is significantly better than the other methods (group a in Figure 4a,c). The next more accurate method is RF ($NSE = 0.504 \pm 0.022$ and $RMSE = 1.181 \pm 0.025$ °C), increasing NSE and worsening the RMSE results of the previous method ($NSE = 0.578 \pm 0.025$ and $RMSE = 1.068 \pm 0.027$ °C). According to R² RFRK is the most accurate method with the exception of RF without residual kriging. Several methods (all except GLS and

GLSRK) obtain PBIAS values near to zero and do not show significant differences, so those methods do not overestimate or underestimate temperature values. Figure 4d represents the percentage of bias obtained, but the groups in the a posteriori contrast are obtained using bias in absolute terms. RF and RFRK (group a and ab respectively) are the most accurate cases. RF underestimates the temperature by 0.04%, while RFRK overestimates it by 0.17%. It is noteworthy that GLS and GLSRK, whose mean PBIAS are practically 0, obtain biases in absolute terms greater than the rest. The R^2 values obtained by the regression models were very high when measured in calibration (GLS: 0.9165, RF: 0.9939 and SVM: 0.9627).

From these results we conclude that, in general, RF-based methods are more accurate and their performance increases when spatial component are included (RFRK). It is also noteworthy the greater variance observed in the GLS method versus machine learning and OK based methods (Figure 4). This results show that the accuracy is very different among observatories and, therefore the robustness of the methods is lower.

Figure 5 shows the importance of the predictors and their effects on the RF model. Effects in RF are computed using partial plots. In such plots, the prediction for a variable X , evaluated at $X = x$, is:

$$f(x) = \frac{\sum_{i=1}^n \hat{f}(x, x_i, o)}{N} \quad (7)$$

where x_i, o represents the value for predictors other than X for individual i and \hat{f} is the predicted value. For continuous variables, red points indicate partial values and dashed red lines indicate a smoothed error bar of \pm two standard errors. Black dashed line are the partial values. We used the function `partial.plot` in the R package `randomForestSRC` [74].

The most important predictors are LST, cdayt and radiation. The effect of all of them are clearly as expected, except the slight reduction in estimated temperature for the largest radiation values. We think the real behaviour might be a stabilisation of temperature for high radiation, although this was altered because of the three points with larger temperature estimation, which may represent a local effect. The least important predictors also have sound effects, although in two of them (albedo and TWI) confidence intervals are too large to draw any clear conclusion. On the other hand, the effects of longitude and TWI are very small, less than 1 °C.

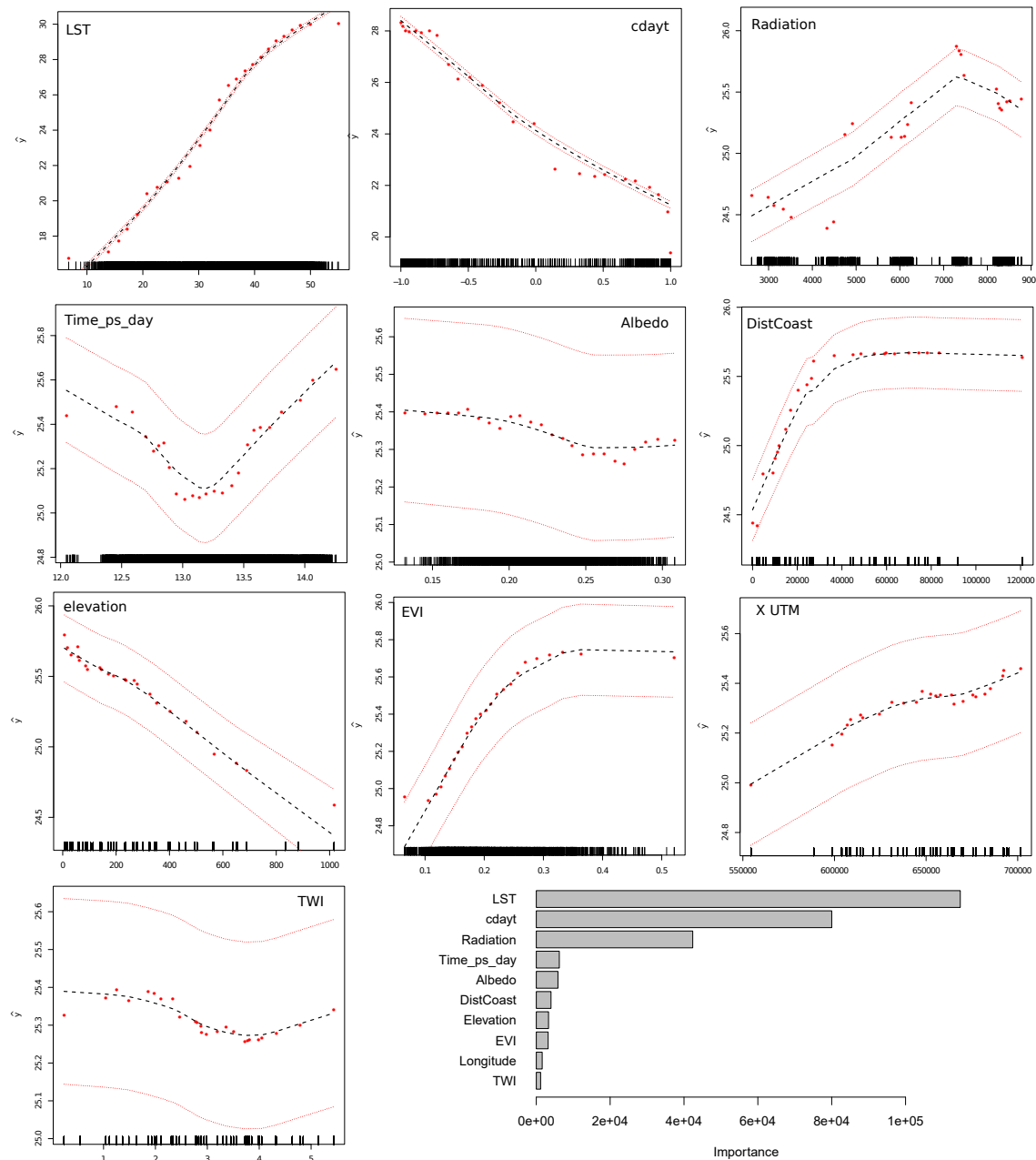


Figure 5. Effects and importance of predictors in the Random Forest Model and LOO-CV in spatial domain approach.

3.2. K-block Cross-Validation

Figure 6a shows the semivariogram of the mean of the daily instantaneous temperature values; its range is about 49 km. Figure 6b shows a boxplot of the range values estimated in the different daily semivariograms. In every case the range is higher than 80 km, so this value was used as the size of the spatial blocks. The blocks were generated with the R blockCV package [67], obtaining finally 7 cells distributed in $k = 6$ blocks as the number of cases in the two cells forming block 1 is quite low. The cells, and block identification, are included in Figure 3.

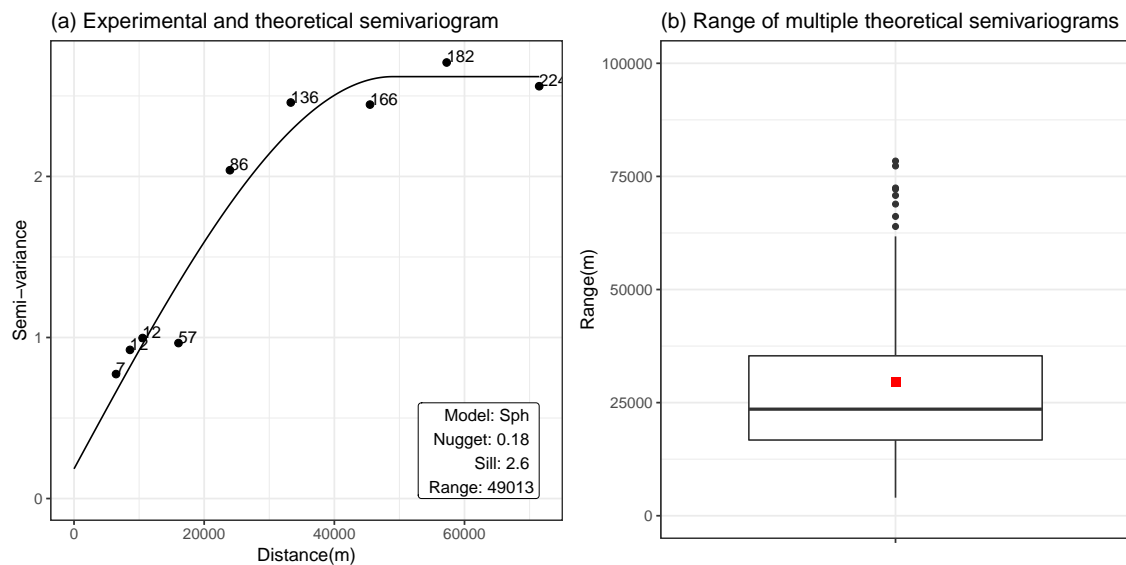


Figure 6. (a) semivariogram of the mean of the daily instantaneous temperature; (b) boxplot of the semivariogram estimated ranges of daily instantaneous temperatures.

Figure 7a represents the autocorrelation function (ACF) of the daily average of instantaneous temperature of all weather stations. As expected, this variable shows a clear intra-annual cyclic behaviour. Figure 7b shows the distribution of the first non-significant lag time for all weather stations. According to the figure, 25 days may be considered a safe threshold beyond which no autocorrelation occurs. So, blocks with temporal size larger than 25 days guarantee that temporal autocorrelation is not affecting cross-validation. However, to reduce computation time and to avoid problems due to some missing data, just two temporal blocks were used.

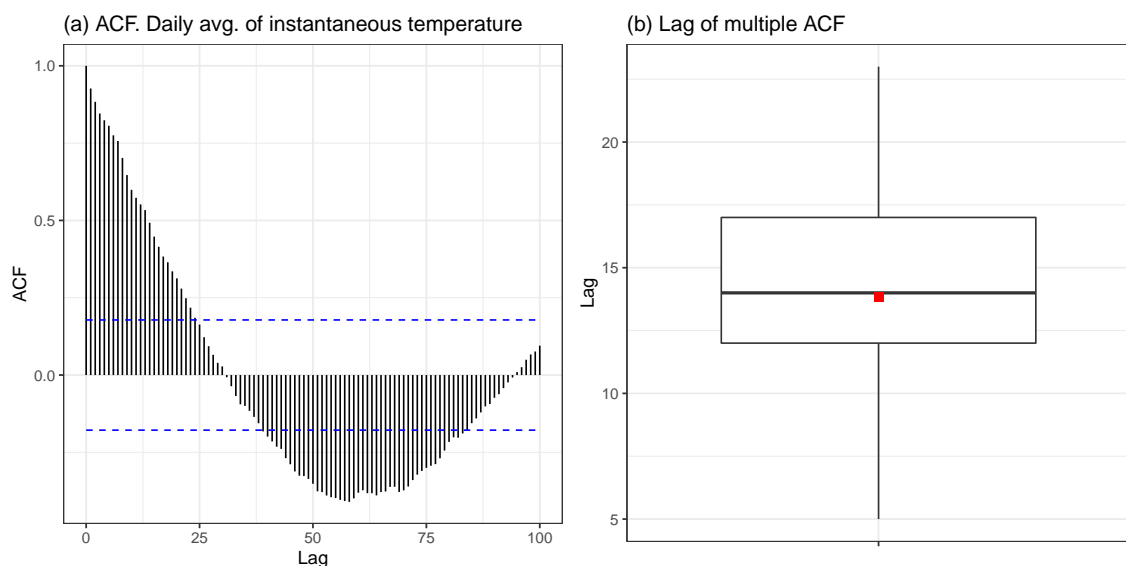


Figure 7. (a) ACF of the aggregated time series; blue dashed lines represent the 95% confidence intervals. (b) Distribution of the first non-significant lag time (in days).

An additional problem of this k-block-CV approach is that cases in the border of a spatio-temporal block are nearer to cases in the neighbour blocks than the advisable distance, so it is safer to reduce the number of blocks to reduce the percentage of such cases. To avoid this issue, to reduce computation time and to avoid problems due to some missing data, the number of temporal blocks was reduced

to 2. Therefore, in this study a spatio-temporal k-block-CV with $K = 36$ (6 spatial blocks and 2 temporal blocks) was implemented.

Table 3 shows the accuracy statistics for the three main predictive models. Neither local nor regression-kriging were validated with this approach because the computing time was excessive and because the semivariograms computed in a blocks are not necessarily representative for other blocks. Figure 8 shows these values and also the results of the comparisons of the values of the statistics. RF is the most accurate method according with all statistics, only PBIAS gives not very different results for all methods.

Table 3. Mean and confidence intervals of the four validation statistics calculated for the three temperature estimation regression methods and spatio-temporal k-block-CV approach. The contrast for PBIAS is done with absolute values.

	GLS	RF	SVM
R^2	0.805 ± 0.039	0.888 ± 0.026	0.827 ± 0.036
NSE	0.444 ± 0.135	0.805 ± 0.064	0.571 ± 0.135
RMSE	5.251 ± 0.693	3.009 ± 0.325	4.606 ± 0.968
PBIAS	8.142 ± 12.249	6.258 ± 4.013	10.092 ± 9.528

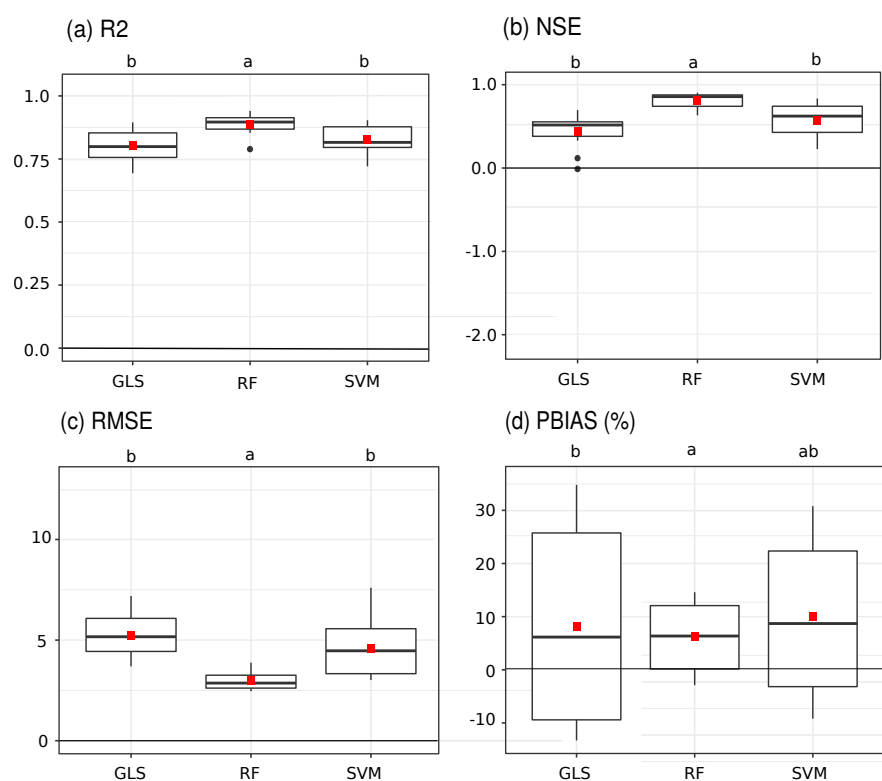


Figure 8. Boxplot showing Anova results of the four validation statistics calculated for the four temperature estimation regression methods and spatio-temporal k-block-CV approach. The contrast for PBIAS is done with absolute values. Significantly different groups (Tukey–Kramer contrast using heteroscedasticity-consistent covariance matrix of the parameters (HC3), $\alpha = 0.05$) are represented by different letters.

Figure 9 shows the RF importance and the effects of the variables that overcome the fast forward feature selection method. The most important variables are satellite land surface temperature, cdayt and radiation. Considerably less important variables, but still in the model, are time passage of the satellite, distance to the coast, and elevation.

All the effects are quite significant as their confidence interval are quite narrow, both the confidence intervals and the shape of the effects plots are quite similar to those shown in Figure 5 for the same variables.

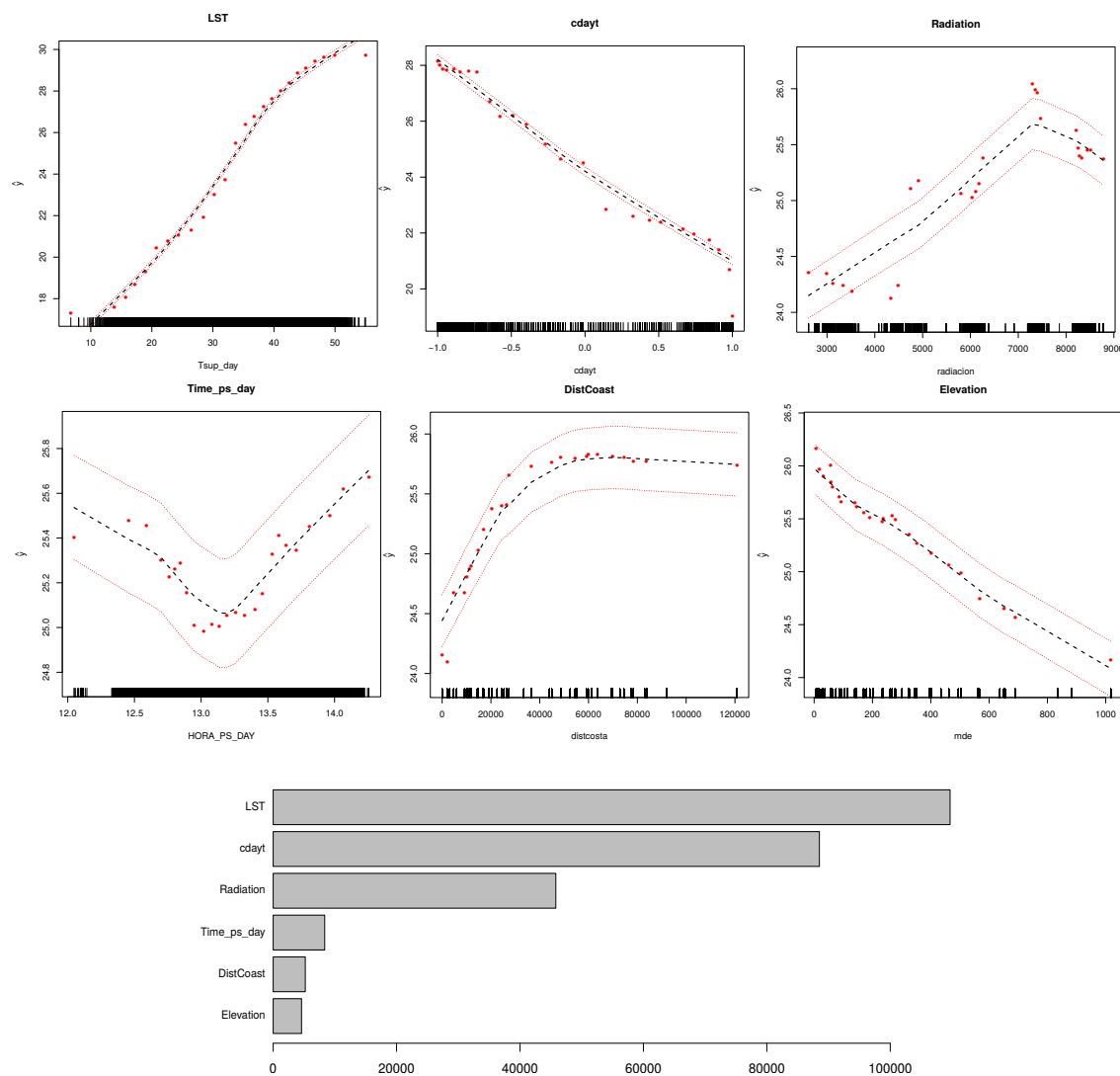


Figure 9. Effects and importance of predictors in the RF Model with the variables selected by FFS algorithm and k-block-CV to evaluate performance.

3.3. Final Prediction

Figure 10 shows the resulting temperature maps at the satellite passage time for the seven methods analysed. The spatial pattern produced by OK is clearly poorer as it does not use any ancillary data. The patterns produced by the regression methods are quite similar, reproducing the influence of the topographical variables that are relevant for the modelling of the spatial variability of temperature. The main differences among the regression methods involves the prediction of maxima and minima. The regression trees on which RF is based prevent the prediction of abnormally high or low values when the values of the predictors exceed the values used in calibration. This does not happen with GLS or SVM, whose maps show higher extremes, especially in the case of the GLS model. These extreme values are, however, smoothed by the kriging of the residuals. Although the lack of anomalous values when predicting beyond the range of values used to calibrate the model is probably a point in favour of RF, it is difficult to ascertain to what extent it may result in an overestimation of minima and underestimation of maxima. Finally, all regression methods produce an artifact in the cells nearest the coast, which show temperature values probably lower than the real values.

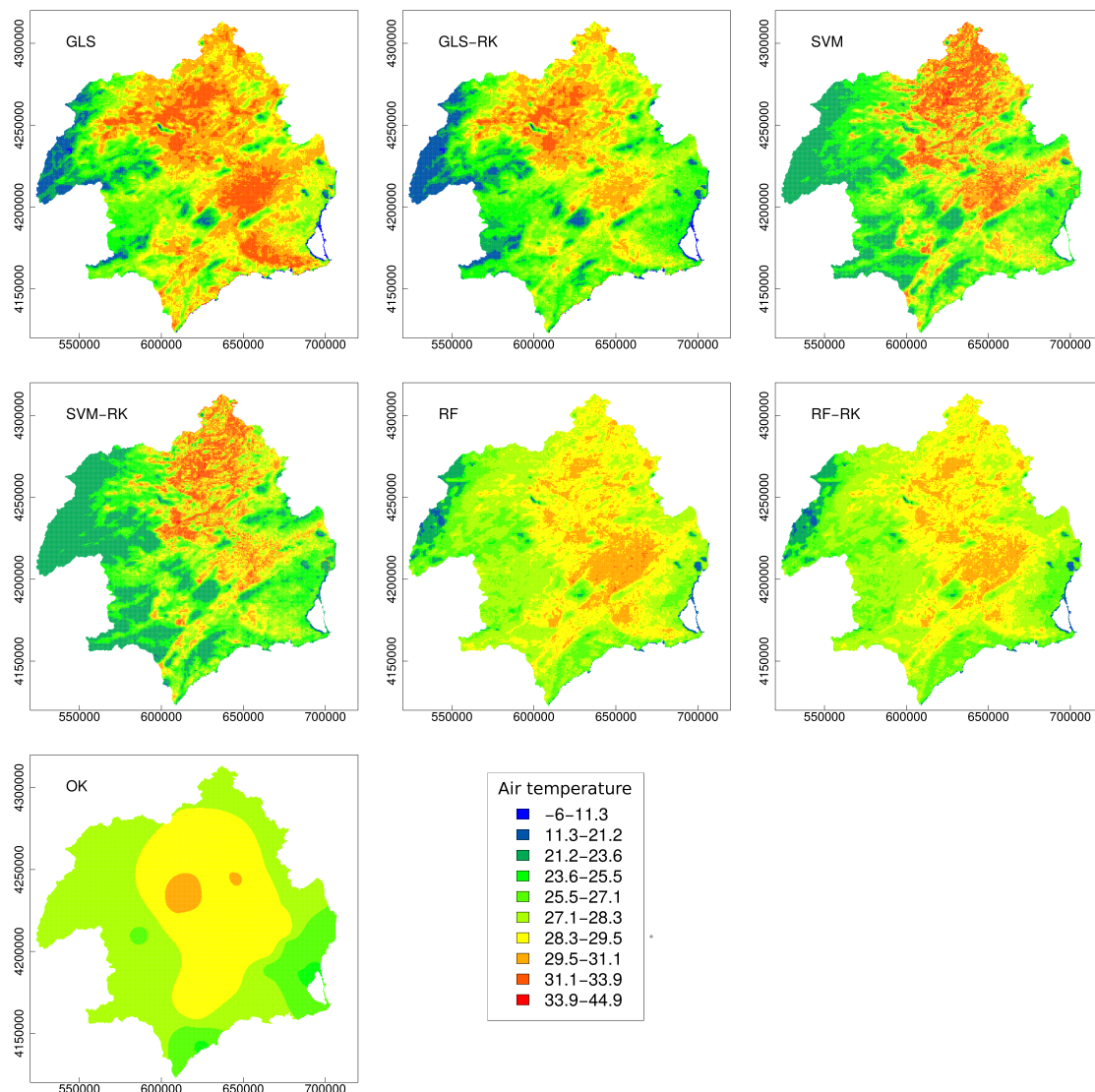


Figure 10. Predictions for day 145 in 2012.

As was mentioned in the introduction, SVM has been reported to obtain similar accuracy than RF and better accuracy than other machine learning methods such as neural networks; however, its main drawback is that there is no way to know in advance which kernel and parameter values will give the best results. It is noteworthy that, in this study, RF obtained more accurate results than SVM, even when the parameters of the latter were optimised ($C = 8$, $\sigma = 0.072$) but not those of the former.

4. Discussion

Some of the research in which instantaneous temperature values are obtained at the time of passage of the TERRA and AQUA satellites on clear days, are based on the adiabatic lapse rate [6,27]. In these research, RMSE values of 3.3°C and 2.6°C , respectively, were obtained. Other studies are based on the assumption of a hydrostatic equilibrium in the atmosphere [27,75], obtaining an RMSE of 2.9°C in both works. Most of the work related to the estimation of the air temperature from remote sensing data and using regression models are focused on obtaining the maximum and minimum temperatures, but not on the estimation of T_a at the time of satellite passage. In this sense, ref. [15] propose a multiple linear regression using as predictors some of the variables used in the present work and in which they obtain an RMSE of 1.9°C , although in this case the results do not correspond only to clear days, so their results are not extrapolable to those obtained in this work. The results obtained in

the present work, especially for RF and RFRK, represent an important improvement with respect to the models based on the adiabatic lapse rate or on the hydrostatic equilibrium of the atmosphere and can be of great help for the improvement of the estimation of the real evapotranspiration, on a regional scale. Traditionally, the temperature at the time of the passage of the satellites, required by most methods, has been obtained using mechanical interpolation techniques, as the Inverse Distance Weighting [76] or models based on the hydrostatic equilibrium of the atmosphere from the values provided by the MODIS MOD07 product [77], which in addition to presenting a lower accuracy than that obtained for RFRK in this work, have the disadvantage of having a spatial resolution of 5 km.

In this paper, two different approaches to cross validation have been used: LOO-CV in the spatial domain and spatio-temporal k-block-CV. The latter approach arises from concerns about the problems of doing random cross-validation with spatio-temporal data due to a lack of independence among samples. In our case, quite larger R^2 and NSE values when doing k-block-CV, whereas RMSE and PBIAS values were also larger. The reason of the difference in RMSE and PBIAS when comparing the two validation methods is that the range of air temperature values in LOO-CV in the spatial domain models (using just one day) is lower than the corresponding range in the k-block-CV models (using 182 days). For this reason, comparison of RMSE and PBIAS among different validation approaches could be misleading; however, comparisons of R^2 and NSE are quite safer. The fact that LOO-CV in the spatial domain produces lower accuracy statistics is probably due to the use of a quite lower number of data when calibrating the models. On the other hand the results on which method is more accurate are also similar, with k-block-CV highlighting RF as the most accurate method.

It is difficult to establish the generalisation capability of this approach. The fact that in this case RFRK were the best model does not mean that it is going to be the best model in other areas. In general, it is not easy to pointing out a machine learning method as the best, in any research field. We think that similar investigations are needed in different areas, so that a meta-analysis of their results might be carried out to obtain some insight on which methods work best and under what circumstances.

5. Conclusions

The results obtained in the present work, especially for RF and RFRK, represent an important improvement with respect to the most widely used models. The accuracy of different machine learning algorithms was tested and compared with those of more traditional methods, such as MLRM and OK. The framework built for that goal is based on free and open source software and would be easily reproducible in other study areas. Easy to obtain environmental variables were used as predictors. Finally, this methodology produces daily temperature distributed surfaces with an accurate fit to the observations. It is noteworthy that the spatial variability of both temperature and environmental predictors in Mediterranean semiarid environments and the heterogeneous landscape due to intense human pressure difficult the estimation of environmental variables.

Both validation approaches show that machine learning algorithms, specially RF, reach significantly higher accuracy values. The use of both approaches allows a better knowledge of model accuracy in two different situations: Spatio-temporal k-block-CV allows to evaluate in a realistic way the accuracy of models that include all observations in time and space, considering the effect of spatio-temporal dependence. On the other hand, LOO-CV in spatial domain is an accuracy estimation considering a very small number of observations: $n - 1$ where n is the number of observatories, and all the data corresponding to the same day.

RF with OK of the residuals obtained the most accurate results with $R^2 = 0.888$, $NSE = 0.805$, $RMSE = 3.009$ and $PBIAS = 6.258$. The last result indicates that it is the only method that produces a slight temperature underestimation, whereas the rest of the methods overestimate temperature to a greater or lesser degree. The maps obtained with RF do not show the extreme values usually present in other regression methods, and that also appear in this study.

The introduction of environmental information related with temperature and easy to obtain from a DEM improves the prediction capability of the model. In particular, potential radiation, distance to the coast and elevation turned to be selected variables in both feature selection strategies.

K-block-CV produced a higher accuracy estimation than LOO-CV in the spatial domain. Probably because in k-block-CV calibrated models used quite more data than LOO-CV in the spatial domain, so the space of variables are more thoroughly sampled. The purpose of LOO-CV in the spatial domain is not really to obtain a global accuracy measurement, but to obtain the accuracy in specific days. However, the ranking of methods was similar.

The variable selection based on k-block-CV produced a smaller set of predictors (6) than the variable selection based on LOO-CV in the spatial domain (10). All the predictors not included by the first one turned to be of scarce importance in the RF model. The most important predictors are LST, cdayt and potential radiation. This three predictors are physically quite correlated with air temperature. The satellite passage time, distance to the coast and elevation are less important. The last two predictors have a quite clear and physically sound effect on air temperature time. It is evident in the case of elevation, and in the case of distance to the coast, due to the high temperatures in the study area, the sea acts as a cooling factor more than a warmer factor. It is less clear the effect of the satellite passage time that should be analysed in further research.

Author Contributions: Conceptualization, Marcos Ruiz-Álvarez, Francisco Alonso-Sarria and Francisco Gomariz-Castillo; Investigation, Marcos Ruiz-Álvarez; Methodology, Marcos Ruiz-Álvarez, Francisco Alonso-Sarria and Francisco Gomariz-Castillo; Project administration, Francisco Alonso-Sarria; Software, Francisco Alonso-Sarria and Francisco Gomariz-Castillo; Supervision, Francisco Alonso-Sarria Writing—original draft, Marcos Ruiz-Álvarez and Francisco Alonso-Sarria; Writing—review & editing, Francisco Alonso-Sarria.

Funding: Project CGL2017-84625-C2-2-R financed by (MINECO/AEI/FEDER, UE).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DEM	Digital Elevation Model
EVI	Enhanced Vegetation Index
GLS	Generalised Least Squares
LST	Land Surface Temperature
MLRM	Multiple Linear Regression Model
MODIS	Moderate Resolution Imaging Spectroradiometer
OK	Ordinary Kriging
OLS	Ordinary Least Squares
RF	Random Forest
RK	Regression-kriging
SVM	Support Vector Machine
TVX	Temperature-Vegetation Index
TWI	Topographic Wetness Index
VIF	Variance Inflation Factor

References

1. Liu, S.; Su, H.; Zhang, R.; Tian, J.; Wang, W. Estimating the Surface Air Temperature by Remote Sensing in Northwest China Using an Improved Advection-Energy Balance for Air Temperature Model. *Adv. Meteorol.* **2016**, *11*. [[CrossRef](#)]
2. Prihodko, L.; Goward, S.N. Estimation of air temperature from remotely sensed surface observations. *Remote Sens. Environ.* **1997**, *60*, 335–346. [[CrossRef](#)]

3. Nieto, H.; Sandholt, I.; Aguado, I.; Chuvieco, E.; Stisen, E. Air temperature estimation with MSG-SEVIRI data: Calibration and validation of the TVX algorithm for the Iberian Peninsula. *Remote Sens. Environ.* **2011**, *115*, 107–116. [[CrossRef](#)]
4. Zhang, R.; Rong, Y.; Tian, J.; Su, H.; Li, Z.L.; Liu, S. A Remote Sensing Method for Estimating Surface Air Temperature and Surface Vapor Pressure on a Regional Scale. *Remote Sens.* **2015**, *7*, 6005–6025. [[CrossRef](#)]
5. Liu, S.; Su, H.; TIAN, J.; Zhang, R.; Wang, W.; Wu, Y. Evaluating Four Remote Sensing Methods for Estimating Surface Air Temperature on a Regional Scale. *J. Appl. Meteorol. Climatol.* **2017**, *56*, 803–814. [[CrossRef](#)]
6. Zhu, W.; Lu, A.; Jia, S.; Yan, J.; Mahmood, R. Retrievals of all-weather daytime air temperature from MODIS products. *Remote Sens. Environ.* **2017**, *189*, 152–163. [[CrossRef](#)]
7. Vogt, J.; Viau, A.; Paquet, F. Mapping regional air temperature fields using satellite-derived surface skin temperatures. *Int. J. Climatol.* **1997**, *17*, 1559–1579. [[CrossRef](#)]
8. Cresswell, M.; Morse, A.; Thomson, M.; Connor, S. Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model. *Int. J. Remote Sens.* **1999**, *20*, 1125–1132. [[CrossRef](#)]
9. Jang, J.; Viau, A.; Anctil, F. Neural network estimation of air temperatures from AVHRR data. *Int. J. Remote Sens.* **2004**, *25*, 4541–4554. [[CrossRef](#)]
10. Lin, S.; Moore, N.; Messina, J.; DeVisser, M.; Wu, J. Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 128–140. [[CrossRef](#)]
11. Benali, A.; Carvalho, A.; Nunes, J.; Carvahais, N.; Santos, A. Estimating Air Surface Temperature in Portugal Using MODIS LST Data. *Remote Sens. Environ.* **2012**, *124*, 108–121. [[CrossRef](#)]
12. Kim, D.; Han, K. Remotely Sensed Retrieval of Midday Air Temperature Considering Atmospheric and Surface Moisture Conditions. *Int. J. Remote Sens.* **2013**, *34*, 247–263. [[CrossRef](#)]
13. Zakšek, K.; Schroedter-Homscheidt, M. Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 414–421. [[CrossRef](#)]
14. Xu, Y.; Knudby, A.; Ho, H.C. Estimating daily maximum air temperature from MODIS in British Columbia, Canada. *Int. J. Remote Sens.* **2014**, *35*, 8108–8121. [[CrossRef](#)]
15. Cristóbal, J.; Ninyerola, M.; Pons, X. Modeling air temperature through a combination of remote sensing and GIS data. *J. Geophys. Res.* **2008**, *113*, D13106. [[CrossRef](#)]
16. Fu, G.; Shen, Z.; Zhang, X.; Shi, P.; Zhang, Y.; Wu, J. Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. *Acta Ecol. Sin.* **2011**, *21*, 8–13. [[CrossRef](#)]
17. Vancutsem, C.; Ceccato, P.; Dinku, T.; Connor, S. Evaluation of MODIS Land Surface Temperature Data to Estimate Air Temperature in Different Ecosystems over Africa. *Remote Sens. Environ.* **2010**, *114*, 449–465. [[CrossRef](#)]
18. Kloog, I.; Nordio, F.; Coull, B.; Schwartz, J. Predicting Spatiotemporal Mean Air Temperature Using MODIS Satellite Surface Temperature Measurements across the Northeastern USA. *Remote Sens. Environ.* **2014**, *150*, 132–139. [[CrossRef](#)]
19. Yang, Y.Z.; Cai, W.H.; Yang, J. Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China. *Remote Sens.* **2017**, *9*, 410. [[CrossRef](#)]
20. Nemani, R.R.; Running, S.W. Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. *J. Appl. Meteorol.* **1989**, *28*, 276–284. [[CrossRef](#)]
21. Goward, S.N.; Waring, R.; Dye, D.; Yang, J. Ecological remote sensing at OTTER: Satellite macroscale observations. *Ecol. Appl.* **1994**, *4*, 322–343. [[CrossRef](#)]
22. Czajkowski, K.; Mulhern, T.; Goward, S.; Cihlar, J.; Dubayah, R.; Prince, S. Biospheric environmental monitoring at BOREAS with AVHRR observations. *J. Geophys. Res.* **1997**, *102*, 651–662. [[CrossRef](#)]
23. Tao, J.; Zhang, Y.; Zhu, J.; Jiang, Y.; Zhang, X.; Zhang, T.; Xi, Y. Elevation-Dependent Temperature Change in the Qinghai-Xizang Plateau Grassland During the Past Decade. *Theor. Appl. Climatol.* **2013**, *117*, 61–71. [[CrossRef](#)]
24. Jin, M.; Dickinson, R. Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle: 1. without clouds. *J. Geophys. Res. Atmos.* **1999**, *104*, 2105–2116. [[CrossRef](#)]
25. Sun, D.; Pinker, R.; Kafatos, M. Diurnal temperature range over the United States: A satellite view. *Geophys. Res. Lett.* **2006**, *33*. [[CrossRef](#)]

26. Gholamnia, M.; Alavipanah, S.K.; Boloorani, A.D.; Hamzeh, S.; Kiavarz, M. Diurnal Air Temperature Modeling Based on the Land Surface Temperature. *Remote Sens.* **2017**, *9*, 915. [\[CrossRef\]](#)
27. Golkar, F.; Sabziparvar, A.A.; Khanbilvardi, R.; Nazemosadat, M.J.; Zand-Parsa, S.; Rezaei, Y. Estimation of instantaneous air temperature using remote sensing data. *Int. J. Remote Sens.* **2018**, *39*, 258–275. [\[CrossRef\]](#)
28. Mira, M.; Ninyerola, M.; Batalla, M.; Pesquer, L.; Pons, X. Improving Mean Minimum and Maximum Month-to-Month Air Temperature Surfaces Using Satellite-Derived Land Surface Temperature. *Remote Sens.* **2017**, *9*, 1313. [\[CrossRef\]](#)
29. Hooker, J.; Duveiller, G.; Cescatti, A. A global dataset of air temperature derived from satellite remote sensing and weather stations. *Sci. Data* **2018**, *5*, 1–11. [\[CrossRef\]](#)
30. Neteler, M.; Mitasova, H. *Open Source GIS: A GRASS GIS Approach*; Springer Science & Business Media: Boston, MA, USA, 2013; Volume 689.
31. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
32. Confederación Hidrográfica del Segura. *Plan Hidrológico de la demarcación hidrográfica del Segura. Ciclo de planificación hidrológica 2015–2021*; Technical Report; Ministerio para la Transición Ecológica: Madrid, Spain, 2015.
33. Conesa García, C.; Alonso Sarria, F. *El clima de la Región de Murcia*; Editum: Murcia, Spain, 2006; pp. 95–128.
34. Sørensen, R.; Zinko, U.; Seibert, J. On the calculation of the topographic wetness index: Evaluation of different methods based on field observations. *Hydrol. Earth Syst. Sci.* **2006**, *10*, 101–112. [\[CrossRef\]](#)
35. Hofierka, J.; Suri, M. The solar radiation model for Open source GIS: implementation and applications. In Proceedings of the International GRASS Users Conference, Trento, Italy, 11–13 September 2002.
36. Gasch, C.; Hengl, T.; Gräler, B.; Meyer, H.; Magney, T.; Brown, D. Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. *Spat. Stat.* **2015**, *14*, 70–90. [\[CrossRef\]](#)
37. Beven, K. Runoff production and flood frequency in catchments of order n: An alternative approach. In *Scale Problems in Hydrology*; Gupta, V.K., Rodriguez-Iturbe, I., Wood, E., Eds.; Reidel: Dordrecht, The Netherlands, 1986; pp. 107–131.
38. Hengl, T.; Reuter, H.I. *Geomorphometry: Concepts, Software, Applications*; Elsevier: Amsterdam, The Netherlands, 2009; p. 796.
39. Busetto, L.; Ranghetti, L. MODISstp: An R package for automatic preprocessing of MODIS Land Products time series. *Comput. Geosci.* **2016**, *97*, 40–48. [\[CrossRef\]](#)
40. Priyadarshi, N.; Chowdary, V.; Srivastava, Y.; Das, I.C.; Jha, C.S. Reconstruction of time series MODIS EVI data using de-noising algorithms. *Geocarto Int.* **2017**, *33*, 1095–1113. [\[CrossRef\]](#)
41. Huete, A.; Justice, C.; Van Leeuwen, W. MODIS vegetation index (MOD13). *Algorithm Theor. Basis Doc.* **1999**, *3*, 213.
42. Gu, J.; Li, X.; Huang, C.; Okin, G.S. A simplified data assimilation method for reconstructing time-series MODIS NDVI data. *Adv. Space Res.* **2009**, *44*, 501–509. [\[CrossRef\]](#)
43. Metz, M.; Andreo, V.; Neteler, M. A New Fully Gap-Free Time Series of Land Surface Temperature from MODIS LST Data. *Remote Sens.* **2017**, *9*, 1333. [\[CrossRef\]](#)
44. Stöckli, R.; Duguay-Tetzlaff, A.; Bojanowski, J.; Hollmann, R.; Fuchs, P.; Werscheck, M. *CM SAF CLOUD Fractional Cover Dataset from METeosat First and Second Generation—Edition 1 (COMET Ed. 1)*; Satellite Application Facility on Climate Monitoring: Darmstadt, Germany, 2017. [\[CrossRef\]](#)
45. Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D.; R Core Team. *Nlme: Linear and Nonlinear Mixed Effects Models*; Technical Report; R Foundation for Statistical Computing: Vienna, Austria, 2018.
46. Zuur, A.; Ieno, E.; Walker, N.; Saveliev, A.; Smith, G. *Mixed Effects Models and Extensions in Ecology with R*; Springer: Berlin/Heidelberg, Germany, 2009; p. 549.
47. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
48. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: Berlin/Heidelberg, Germany, 2013.
49. Murphy, K. *Machine Learning. A Probabilistic Approach*; The MIT Press: Cambridge, MA, USA, 2012.
50. Radhika, Y.; Shashi, M. Atmospheric Temperature Prediction using Support Vector Machines. *Int. J. Comput. Theory Eng.* **2009**, *1*, 1793–8201. [\[CrossRef\]](#)

51. Balabin, R.; Lomakina, E. Support vector machine regression (LS-SVM)—An alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? *Phys. Chem. Chem. Phys.* **2011**, *13*, 11710–11718. [[CrossRef](#)]
52. Nagendra, K.V.; Jahnavi, Y.; Haritha, N. A Survey on Support Vector Machines and Artificial Neural Network in Rainfall Forecasting. *Int. J. Future Revolut. Comput. Sci. Commun. Eng.* **2017**, *3*, 20–24.
53. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*; Technical Report; R Foundation for Statistical Computing: Vienna, Austria, 2018.
54. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
55. Wright, M.N.; Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [[CrossRef](#)]
56. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
57. Cánovas-García, F.; Alonso-Sarria, F. Optimal Combination of Classification Algorithms and Feature Ranking Methods for Object-Based Classification of Submeter Resolution Z/I-Imaging DMC Imagery. *Remote Sens.* **2015**, *74*, 4651–4677. [[CrossRef](#)]
58. Burrough, P.; McDonnell, R.; Lloyd, C. *Principles of Geographical Information Systems*; Oxford University Press: Oxford, UK, 2015.
59. Hiemstra, P.; Pebesma, E.; Twenhöfel, C.; Heuvelink, G. Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. *Comput. Geosci.* **2008**, *35*, 1711–1721. [[CrossRef](#)]
60. Cressie, N. *Statistics for Spatial Data (Revised Ed)*; Wiley: Hoboken, NJ, USA, 1993.
61. Hengl, T.; Heuvelink, G.B.; Rossiter, D.G. About regression-kriging: From equations to case studies. *Comput. Geosci.* **2007**, *33*, 1301–1315. [[CrossRef](#)]
62. Odeh, I.; McBratney, A.; Chittleborough, D. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. *Geoderma* **1995**, *67*, 215–226. [[CrossRef](#)]
63. Hengl, T.; Heuvelink, G.B.; Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **2004**, *120*, 75–93. [[CrossRef](#)]
64. Roberts, D.; Bahn, V.; Ciuti, S.; Boyce, M.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
65. Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ. Model. Softw.* **2018**, *101*, 1–9. [[CrossRef](#)]
66. Trachsel, M.; Telford, R. Technical note: Estimating unbiased transfer-function performances in spatially structured environments. *Clim. Past* **2016**, *12*, 1215–1223. [[CrossRef](#)]
67. Valavi, R.; Elith, J.; Lahoz-Monfort, J.; Guillera-Arroita, G. blockcv: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **2019**, *10*, 225–232. [[CrossRef](#)]
68. Bennett, N.D.; Croke, B.F.; Guariso, G.; Guillaume, J.H.; Hamilton, S.H.; Jakeman, A.J.; Marsili-Libelli, S.; Newham, L.T.; Norton, J.P.; Perrin, C.; et al. Characterising performance of environmental models. *Environ. Model. Softw.* **2013**, *40*, 1–20. [[CrossRef](#)]
69. Legates, D.R.; McCabe, G.J. Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [[CrossRef](#)]
70. Long, J.; Ervin, L. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *Am. Stat.* **2000**, *54*, 217–224.
71. O’Brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* **2007**, *45*, 673–690. [[CrossRef](#)]
72. Kutner, M.H.; Nachtsheim, C.; Neter, J. *Applied Linear Regression Models*; McGraw-Hill/Irwin: New York, NY, USA, 2004.
73. Juutilainen, I.; Röning, J.; Laurinen, P. A study on the differences in the interpolation capabilities of models. In Proceedings of the 2005 IEEE Midnight-Summer Workshop on Soft Computing in Industrial Applications, 2005 SMCia/05, Espoo, Finland, 28–30 June 2005; pp. 202–207. [[CrossRef](#)]

74. Ishwaran, H.; Kogalur, U. *Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC)*; R package version 2.9.0; R Foundation for Statistical Computing: Vienna, Austria, 2019.
75. Bisht, G.; Bras, R.L. Estimation of net radiation from the MODIS data under all sky conditions: Southern Great Plains case study. *Remote. Sens. Environ.* **2010**, *114*, 1522–1534. [[CrossRef](#)]
76. Xiong, Y.J.; Zhao, S.H.; Tian, F.; Qiu, G.Y. An evapotranspiration product for arid regions based on the three-temperature model and thermal remote sensing. *J. Hydrol.* **2015**, *530*, 392–404. [[CrossRef](#)]
77. Zhu, W.; Jia, S.; Lv, A. A Universal Ts-VI Triangle Method for the Continuous Retrieval of Evaporative Fraction From MODIS Products. *J. Geophys. Res. Atmos.* **2017**, *122*, 10–206. [[CrossRef](#)]

Sample Availability: Samples of the compounds are available from the authors.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).