


Article

Interpolation of Instantaneous Air Temperature using Geographical and MODIS derived variables with Machine Learning Techniques

Marcos Ruiz-Álvarez ^{1,‡} , Francisco Alonso-Sarria ^{2,‡,*}, Francisco Gomariz-Castillo ^{3,‡}

¹ Instituto Universitario de Agua y Medio Ambiente. Universidad de Murcia. Campus de Espinardo, 30001, Murcia, Spain; marcos.ruiz1@um.es

² Instituto Universitario de Agua y Medio Ambiente. Universidad de Murcia. Campus de Espinardo, 30001, Murcia, Spain; alonsarp@um.es

³ Instituto Euromediterráneo del Agua. Campus de Espinardo, 30001, Murcia, Spain; ffgomariz@um.es

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

‡ These authors contributed equally to this work.

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 Machine and Random Forest, are compared with Multivariate Linear Regression, TVX and Ordinary kriging. Several geographic, remote sensing and time variables are used as predictors. The validation is carried out using four different statistics on a daily basis allowing the use of ANOVA to compare the results. The main conclusion is that Random Forest with residual kriging produces the best results ($R^2=0.612 \pm 0.019$, $NSE=0.578 \pm 0.025$, $RMSE=1.068 \pm 0.027$, $PBIAS=-0.172 \pm 0.046$), whereas TVX produces the least accurate results. The environmental conditions in the study area are not really suited to TVX, moreover this method only takes into account satellite data. On the other hand, regression methods (Support Vector Machine, Random Forest and Multivariate Linear Regression) use several parameters that are easily calculated from a Digital Elevation Model, adding very little difficulty to the use of satellite data alone. 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 in 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 MODIS product MOD07_L2 to locally estimate the adiabatic lapse rate [6]. The main drawback of this approach is that spatial resolution is 5 km.
- 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].
- Multivariate regression models using LST and other variables such as NDVI, solar zenith angle, solar radiation, altitude, julian day, distance to the coast, normalized difference water index (MNDWI) or albedo as predictors [8–15]. The algorithms used range from multivariate 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 have uncertainties due to the clustered distribution of weather stations [19].
- The Temperature-Vegetation Index (TVX), proposed by [20] and [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 approach T_a estimation from a more physical point of view. It has two main drawbacks: several variables that can only be measured in weather stations are needed and, as the Bowen ratio is one of them, it is necessary to know LE to use ADEBAT. Frequently, T_a is estimated in order to estimate LE from it, as in our case, so the use of a surface energy balance is not suitable in this case for practical reasons.

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].

[23] showed that correlation between LST at night and minimum T_a is higher ($R^2=0.93$) than between their daytime equivalents ($R^2=0.79$).

The objective of this work was 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) using all the variables included as predictors in previous works. The results are compared to those obtained with more traditional approaches: Multivariate linear regression, Ordinary Kriging (OK) and TVX. Finally, regression-kriging will be tested using OK to interpolate the residuals of the three regression models, provided that such residuals show spatial autocorrelation.

2. 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 1), which includes the Segura river basin (19,000 km²) and several minor coastal basins. 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.

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 that represents unfavourable conditions for the use of TVX.

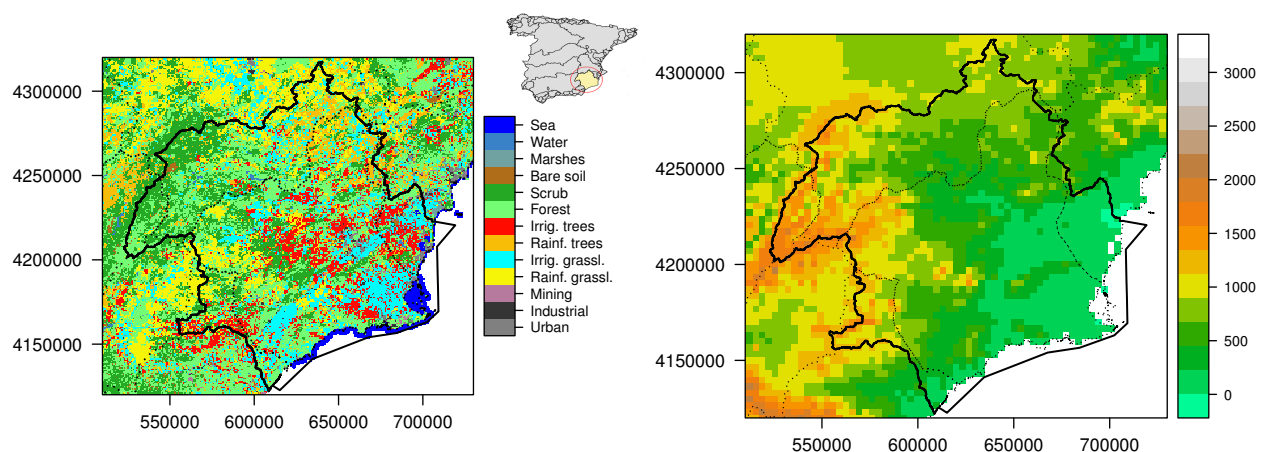


Figure 1. Land uses (left) and elevation (right) of the study area. Solid lines represent DHS boundaries and dotted lines province boundaries.

2.2. Data set

Different variables were used as predictors: 1. geographical variables: longitude, latitude, altitude, TWI index [24] (used to describe the potential accumulation of cold air), monthly potential irradiation ($Whm^{-2}month^{-1}$) obtained from heights using the method proposed by [25] and distance from the coast; 2. time variables: day duration, cdayt (a cosine transformation of julian day following [26], equation 3) 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 equation to calculate cdayt is:

$$cdayt = \cos\left([t_D - \phi] \cdot \frac{2\pi}{365}\right) \quad (1)$$

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 minutes 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). In this work, only data for 2012 were analysed to test which model produces the most accurate results, .

We assumed that the mean air temperature data (obtained from the average of 3 measurements taken every 10 minutes) 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 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 [27] package was used to download and process the images.

The 8 day delay between the MOD13Q1 and MYD13Q1 products allowed combined layers (AQUA and TERRA satellites) of NDVI and EVI to be obtained, with a temporal resolution of 8 days. The process involves 3 steps: 1) Elimination of pixels of very low quality or with large observation errors. To this end, according to [28] all pixels with a value greater than 5 in the VI usefulness index

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
MODIS	Terra	NDVI	MOD13Q1	0.25	16
MODIS	Aqua	NDVI	MYD13Q1	0.25	16

(VIUI) [29,30] were considered to be of very low quality or to have large observation errors, and were therefore removed. 2) Combining the NDVI/EVI layers. Each layer of NDVI/EVI with an 8-day time resolution was obtained by combining the layers obtained by the MOD13Q1 and MYD13Q1 products representing each 8-day period. In this step, therefore, NDVI/EVI values were only obtained from those pixels in which the values of both products were available. 3) Reconstruction of the existing gaps after step 2. For this purpose, the average of both layers was used preferably, if the NDVI/EVI values obtained one week before and one week after were available. Otherwise, the values estimated one week earlier or later were used. When there were no NDVI/EVI values a week before or after, the same process were carried out with the estimated values for 2 weeks before and after.

With regard to the LST layers, it was 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 [31] 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 assumed, according to [32], that the albedo layers are equal to the 20th layer (White-Sky Albedo) of the product MCD43A3. In this case, the reconstruction of the existing gaps was carried out using estimations for the layers 8 days before or after the layer being reconstructed. Within this interval, the values closest to the existing gap were used.

2.2.3. Geographical variables

The geographical variables used as predictors with the regression methods were calculated from the official 25 m resolution Digital Elevation Model downloaded from the Spanish National Geographical Institute (*Instituto Geográfico Nacional, IGN*) website (<http://centrodedescargas.cnig.es/CentroDescargas/index.jsp>). This MDE was also used to obtain distance to the sea, monthly potential irradiation and TWI.

2.3. Estimation methods

Five methods were used to estimate air temperature at the satellite passage time: TVX is a well-known method based solely on satellite data; Multiple linear regression, SVM and RF are regression (global interpolation) methods that rely on covariates to estimate the dependent variable; in this paper, only covariates that can be easily obtained from a DEM were used, the first based on classical statistics and the last two on machine learning; finally Ordinary kriging, a local interpolation method that does not take into account covariates, was used for comparison. Finally, the residuals of the three regression models were interpolated with OK to obtain regression-kriging models.

2.3.1. The temperature-vegetation index (TVX) method

TVX is based on two hypotheses: 1) linear relation between NDVI and LST (equation 2) and 2) $T_a = LST$ under full vegetation cover [2,22] while bare soils are significantly hotter in the same meteorological conditions [22]. It also assumes uniform soil moisture and atmospheric conditions [33]. The first assumption is achieved by using close pixels in a 7x7 window to perform the regression in equation 3 [34], and the second by only using TVX on clear days [5].

$$T_a = a_i + b_i \cdot NDVI_{max} \quad (2)$$

There are, however, several other limitations: the LST/NDVI ratio is reduced when there are differences in sun illumination caused by topography [22], and all cells in the 7x7 window should be of similar height to prevent temperature variations due to differences in height from adding noise to the LST/NDVI relation [22]. The method works better when a large range of NDVI is present in the 7x7 window; however, soil moisture should be fairly constant. This can be problematic in agricultural areas where irrigated and non-irrigated areas are spatially mixed. Moreover, it is also assumed that there are no differences in aerodynamic resistance within the moving window due to different surface types [33]. Also, in accordance with [22], sparse vegetation regions show a great uncertainty when using TVX.

According to [19], TVX does not take into account other factors, such as latitude, longitude, elevation or julian day, among others, that may affect the relation between LST and T_a . Interestingly, these are the factors usually taken into account in multivariate regressions. Higher residuals have been reported in winter than in other months [21], which have been attributed to the presence of snow, although this is not usually a problem in our study area.

The coefficients a_i and b_i in equation 2 are obtained cell by cell calibrating a linear regression (equation 3) to predict LST using NDVI as predictor (both available as MODIS data). This linear model

assumes uniform atmospheric forcing and similar soil moisture conditions in all the cells used to calibrate. A 7x7 window around each cell is used to approach this assumptions [5,34].

$$LST = a_i + b_i \cdot NDVI \quad (3)$$

As a result of this first step, layers of a and b are obtained. To obtain the value of $NDVI_{max}$ (a single value for the whole study area) we used the technique proposed by [3] and also used by [5]. A new linear model is calibrated (equation 4) with data from weather observations of T_a . In this case, $NDVI_{max}$ is the parameter to calibrate.

$$T_a - a_i = b_i \cdot NDVI_{max} \quad (4)$$

The parameter $NDVI_{max}$ varies among different study areas and sensors [3,34]. Prihodko & Goward [2] derived a $NDVI_{max}$ of 0.86 for AVHRR images, while [22] reduced it to 0.7 and [33] to 0.65 using SEVIRI images in the Senegal river basin.[2] even stated that $NDVI_{max}$ can vary among different vegetation covers.

A new GRASS module is programmed to calibrate a linear model from two input layers (independent and dependent) in windows of NxN cells. This produces 3 output layers: containing the coefficients of the regression and the R^2 value.

```
r.neighbors.lm independent=NDVI dependent=LST size=7 a=TVXa b=TVXb r2=TVXr2
```

Once a_i and b_i are available, both layers are read in R as well as the weather station data to calibrate equation 4 in order to obtain $NDVI_{max}$. Finally r.mapcalc is used to apply equation 2 in order to obtain T_a :

```
r.mapcalc "Ta=TVXa + TVXb * NDVImax"
```

Our study area has some of the problems that, according to the literature, may reduce the effectiveness of the model: large variability of land uses in small areas, especially taking into account the MODIS resolution, sparse vegetation due to semiarid conditions, and frequent cloudiness due to proximity to the sea.

2.3.2. Multiple Linear regression

Simple and multiple linear regression models (MRLM) are the most popular models for estimating air temperature. MRLM 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

generalized 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 [35].

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 [36]. In the case of multiple linear regression a subsequent stepwise procedure was carried out to minimize the number of variables to provide a more parsimonious model.

2.3.3. Support Vector Machine

SVM was originally developed for classification but it was adapted to regression as a robust regression method that tries to minimise the effect of outliers. SVM for regression is described in [37]. 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 $|e| > \epsilon$ contribute linearly rather than quadratically to the error objective function to be minimized. This somewhat counterintuitive approach (the more accurately predicted points are not used to fit the line) has proven effective. SVM is calibrated by minimizing equation 5 [38].

$$J = C \sum_{i=1}^N (\xi_i^+ + \xi_i^-) + \frac{\|\mathbf{w}\|^2}{2} \quad (5)$$

where C is a cost parameter to penalise large residuals. The solution of equation 5 is also subject to the following linear constraints:

$$y_i \leq f(x_i) + \epsilon + \xi_i^+ \quad (6)$$

$$y_i \geq f(x_i) - \epsilon - \xi_i^- \quad (7)$$

$$\xi_i^+ \geq 0 \quad (8)$$

$$\xi_i^- \geq 0 \quad (9)$$

$$(10)$$

and the optimal solution has the form:

$$\mathbf{w} \sum_i \alpha_i \cdot \mathbf{x}_i \quad (11)$$

The estimation of new cases (\mathbf{u}) is made with the equation:

$$\hat{y} = \beta_0 + \sum_{i=1}^n \alpha_i \cdot K(\mathbf{x}_i, \mathbf{u}) \quad (12)$$

where $K(x_i, \mathbf{u})$ is a kernel function, \mathbf{x}_i represent each case in the training data and \mathbf{u} represents predicted values in the new point. There are several possible kernel functions:

- Linear kernel: $K(x_i, \mathbf{u}) = \mathbf{x}_i' \cdot \mathbf{u}$
- Radial kernel: $K(x_i, \mathbf{u}) = \exp(-\sigma \cdot (\mathbf{x}_i \cdot \mathbf{u})^2)$
- Polynomial kernel: $K(x_i, \mathbf{u}) = (\phi(\mathbf{x}_i' \cdot \mathbf{u}) + 1)^d$
- Hyperbolic tangent kernel: $K(x_i, \mathbf{u}) = \tanh(\phi(\mathbf{x}_i' \cdot \mathbf{u}) + 1)$

The last three kernels allow a non-linear generalization of SVM.

In accordance with equation 12, there is an α value for each data point. This over-parametrization is only apparent because $\alpha_i = 0$ for all points where $|e_i| < \epsilon$; the others are the so-called Support Vectors.

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. In this case, we reduced the problem by using a radial basis function (RBF) kernel using the default values for the parameters. The implementation used was that of the R package e1071 [39].

2.3.4. Random Forest

RF is an ensemble of decision trees. In a regression decision tree, heterogeneity of a sample of training data is measured as the variance of the dependent variable. For a set of samples, the weighted average of the variances is used. Decision tree algorithm begins with the whole sample and selects the independent variable and its threshold value, which minimize heterogeneity in the resulting partition. The process continues recursively until a minimum number of cases is reached in all partitions. Each partition is called a node, and the final partitions are called final nodes. The estimated value of the dependent variable in each of the final nodes is the average of all cases inside it.

Once the tree is calibrated, the dependent variable for a new case can be estimated by driving the case through the regression tree until a final node is reached. The previously calculated average for that final node is the estimated value.

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 (the so-called in-bag 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 [40], 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 [41,42] has shown that the accuracy achieved with such parameters is usually near the optimum.

2.3.5. Ordinary kriging

Ordinary kriging is a local interpolation method based on the regionalized variable theory [43]. It uses solely the values measured in the observation points and their location. Its main advantage over other local interpolation methods (such as IDW) is that a statistical analysis of the spatial variability of the values is previously performed and summarised in the semivariogram function. Finally, ordinary kriging 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 ordinary kriging 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 covariates).

In this work, we used the R package `automap` [44], its main advantage being the automatization of a weighted least squares optimal estimation of semivariogram parameters using Gauss-Newton [45].

2.3.6. Validation

As the aim of this work is to evaluate the predictive performance of the different models, a leave-one-out cross validation (LOOCV) was carried out for the 8 estimation methods. Bennett et al. [46] 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 [47] or [46], were used in this research:

Root mean square deviation (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - E_i)^2}{n}} \quad (13)$$

Correlation coefficient:

$$R = \frac{COV(O, E)}{s_O \cdot s_E} \quad (14)$$

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 [47]:

$$nse = \frac{\sum_{i=1}^n (O_i - E_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (15)$$

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:

$$PBIAS = \frac{\sum_{i=1}^n (O_i - E_i)}{\sum_{i=1}^n O_i} \quad (16)$$

The four statistics 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 accuracies of the different observatories were small (Simpson paradox).

As normality and homocedasticity could not be assumed, a Kruskal Wallis contrast was used to test whether differences among the methods were significant. If that was the case, a post-hoc contrast between pairs of models, based on Mann-Whitney and using Holm method to correct p-values among classes, was performed to discover groups of non-significantly different methods.

3. Results and discussion

When TVX was calculated for the 2012 data, the results were disappointing. Figure 2 shows the regression parameters of TVX on May 25, 2012. The map in the upper left shows the correlation

coefficients of the regression model; the map in the upper right shows the b parameter (slope) of such model; and the map in the lower left shows the intercept of the model (T_a when $NDVI=0$). Although most slopes were negative, there was a large group of positive ones that were not expected. Although there is a considerable amount of speckle in the maps, there is also a pattern showing negative slopes in the higher, more forested areas, and positive in the cultivated areas. The agricultural pattern in the study area is characterised by small patches due both to the complex topographical patterns and to the predominance of small properties. This patchiness might be affecting the regression models calibrated in the 7×7 windows, producing abnormal results.

Finally, the map in the lower right shows the resulting T_a map for May 25, 2012. The map shows a very convincing pattern, with lower temperature in the highest areas and higher temperature values in the plains. However, there is a wide variability with completely unrealistic maximum and minimum values, although the frequency of such values was quite low.

Figure 3 shows an analysis of the TVX errors through space and time. The first thing to notice is the high mean error (Figure 3 a) with a large overestimation of temperature, although it is reduced on winter days. Mean square errors (Figure 3 b) are quite high and also tend to be higher in spring. The correlation between observation and estimation (Figure 3 c) also shows very abnormal values, especially on spring days. $NDVI_{max}$ (Figure 3 d) is too high, with most of the values above 1, which is clearly an artifact produced by the problems involved in calibrating the focal linear regression in 7×7 windows. However, the temporal pattern is consistent, with the largest values in spring. We tried to restrict $NDVI_{max}$ beyond 0 and 1, but the errors increased considerably. Despite the high values of $NDVI_{max}$, the r^2 of equation 3 (Figure 3 e) seems rather high and with no temporal pattern. Finally, plot f in Figure 3 shows the spatial pattern of errors. Positive errors (overestimations) are represented in red and negative errors (underestimations) are represented in blue. The size of the dots represents the absolute value of the error. The black circles represent the standard deviation of the errors. Most of the underestimations appear concentrated in the same areas, but it is difficult to deduce a pattern in relation with other covariates.

We attribute these poor results to the problems associated with the TVX methods. These have been highlighted in the literature and are indeed present in the study area: sparse vegetation, strong land use variability at short distances, and also strong topographical variability at short distances.

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. A threshold of $VIF=5$ was established in principle, but it was relaxed to allow the inclusion of LST ($VIF=6.43$). The variables finally eliminated were day duration ($VIF=39.04$) and latitude ($VIF=24.43$).

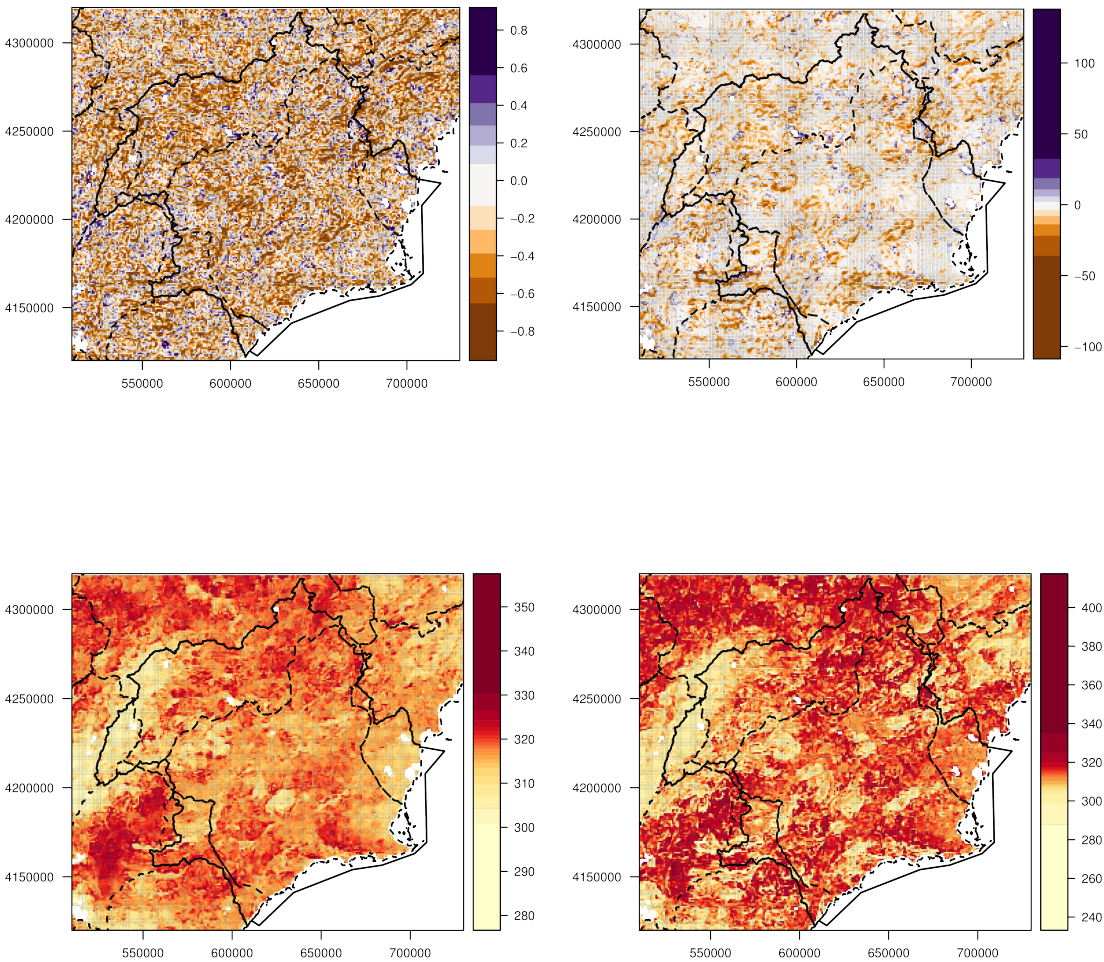


Figure 2. TVX regression parameters on day 145 (May 25) in 2012. Solid lines represent DHS boundaries and dotted lines province boundaries.

Table 2. Mean, standard error and standard deviation of the four validation statistics calculated for the eight temperature estimation methods

	GLS	RF	SVM	GLSRK	RFRK	SVMRK	OK	TVX
R2	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	0.276 ± 0.025
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	-22.967 ± 2.748
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	5.582 ± 0.216
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	10.680 ± 0.821

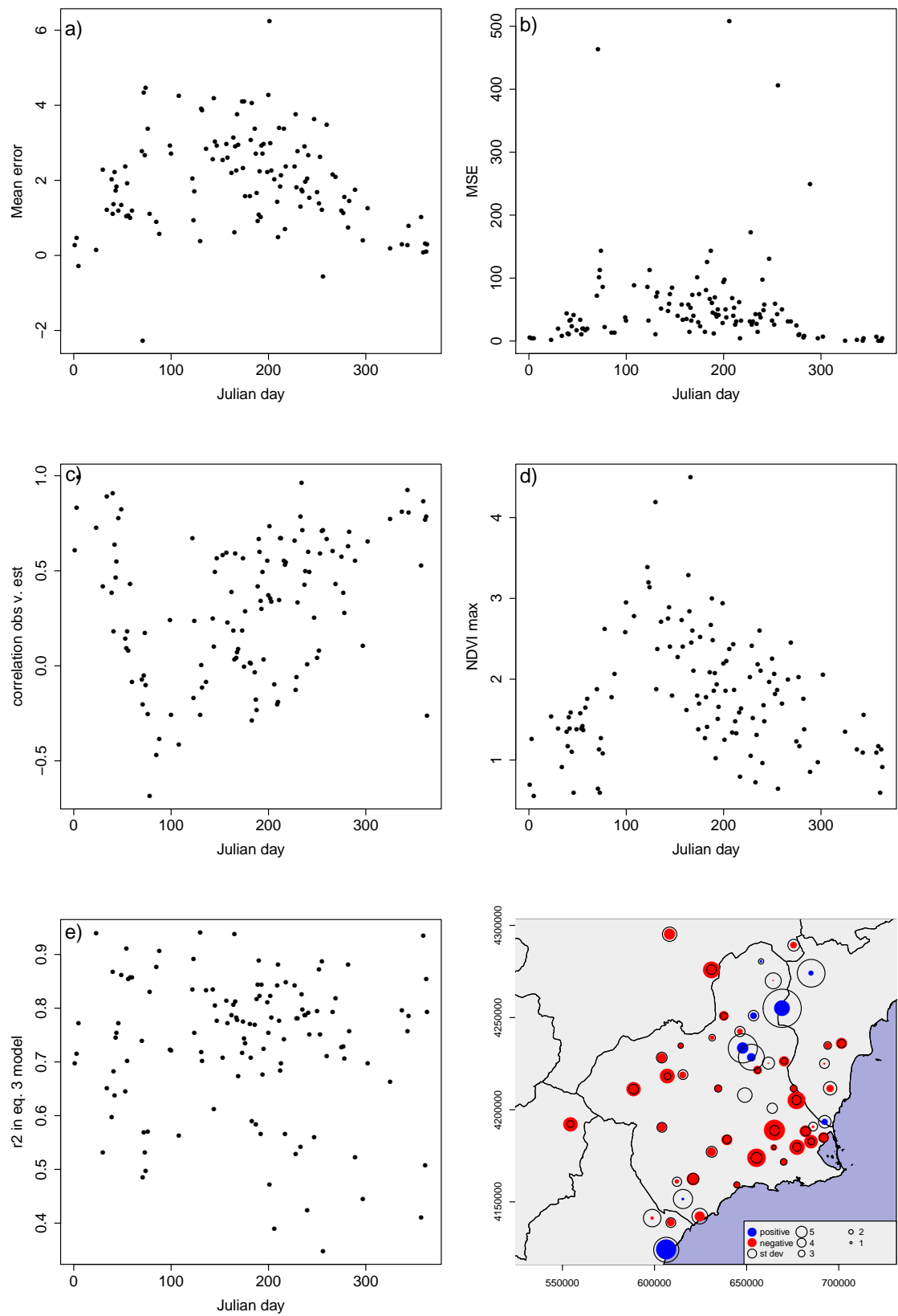


Figure 3. Spatial and temporal distribution of the TVX errors

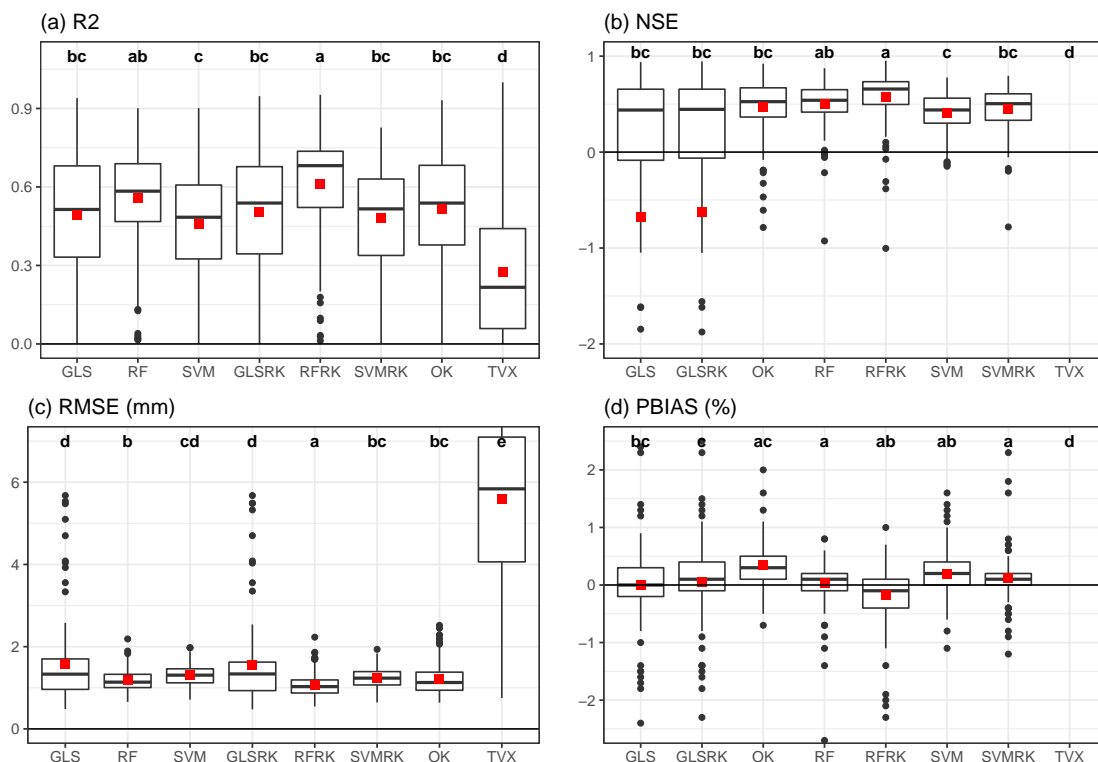


Figure 4. Boxplot showing Anova results of the 4 validation statistics calculated for the eight temperature estimation methods

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, 118 with GLS and GLSRK and 109 for TVX. 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 covariates (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. Finally, in the case of TVX, the regression in 7x7 windows extends the need for cloudless cells to the surroundings of the weather stations, diminishing the number of complete cases to 1175 predictions on 109 days.

In addition, TVX, 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 and 1159 with TVX.

The ANOVA shows significant differences among the algorithms for all the statistics with large F values (r^2 : 18.907, NSE: 17.053, RMSE: 67.698, PBIAS: 26.847) and p-values lower than 0.0001 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, random forest with kriging of the residuals (RFRK) is significantly better than

the other methods. According to r^2 RFRK is still the best method, but its results are not significantly different from those of RF without residual kriging. Several methods (all except GLS, GLSRK and TVX) obtain PBIAS values near to zero without being significantly different. On the other hand, for all statistics TVX produces significantly less accurate results than the other methods. 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).

The bad results of TVX can be explained by the fact that it only uses satellite data and also because the conditions in the study area are not adequate for the use of this method. The main advantage of regression methods, as used in this study, is that only predictors that are easily calculated from MODIS or conventional MDEs were used. It is also easy to implement in other areas.

Figure 5 shows the importance of the predictors and their effects on the model. 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.

Figure 6 shows the resulting temperature maps at the satellite passage time for the eight methods analysed. Due to the strong overestimation of T_a by TVX, a different colour palette is used in the TVX map. The spatial patterns produced by OK and TVX are clearly poorer as the latter do not use any ancillary data and the former uses only NDVI and LST. 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 extrapolation 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.

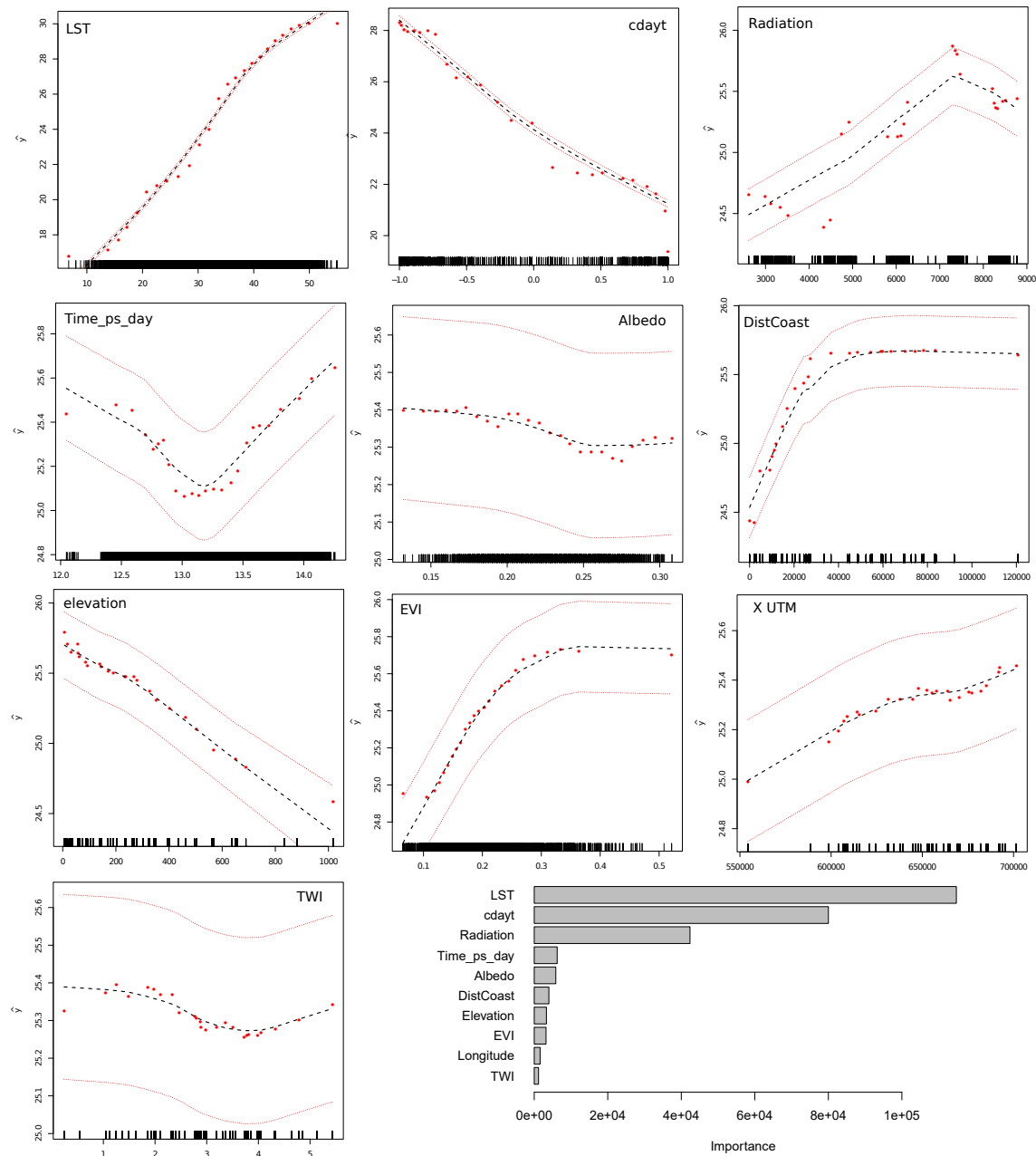


Figure 5. Effects and importance of predictors in the Random Forest Model

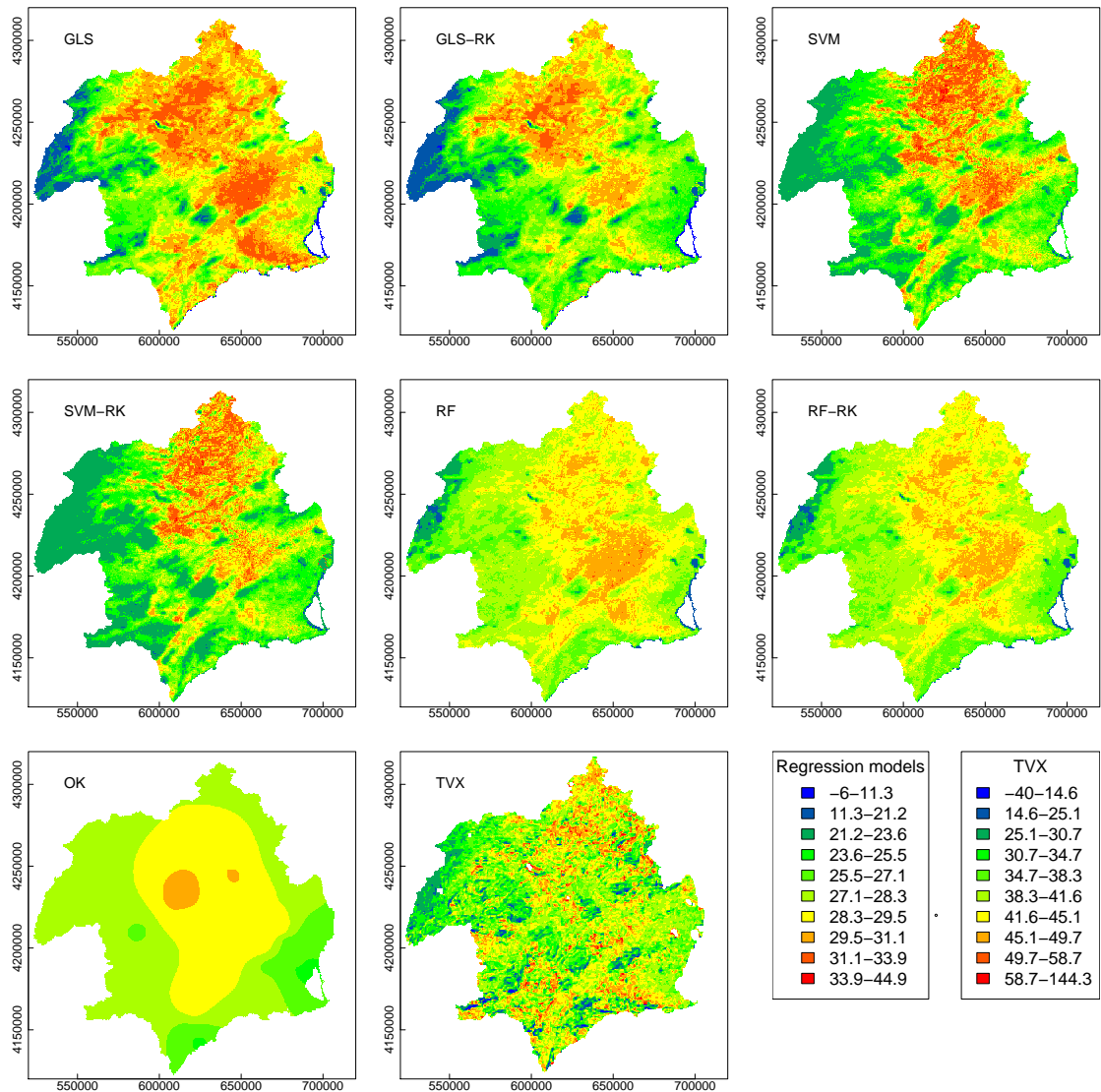


Figure 6. Predictions for day 145 in 2012

4. Conclusions

Random Forest with ordinary kriging of the residuals obtained the most accurate results with $r^2=0.612$, $SE=0.578$, $RMSE=1.068$ and $PBIAS=-0.17$. 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. It is interesting to note that RF obtained more accurate results than SVM, even when the parameters of the latter were optimized but not those of the former. Finally, TVX was the method that produced the worst results, probably because the environmental conditions in the study area are not suited to this method and also because the predictors used by the regression methods explain an important part of temperature variability.

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: This research received no external funding.

Acknowledgments: Pensad si quereis meter alguno.

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

Abbreviations

The following abbreviations are used in this manuscript:

LST	Land Surface Temperature
MLRM	Multiple Linear Regression Model
OLS	Ordinary Least Squares
GLS	Generalised Least Squares
RF	Random Forest
SVM	Support Vector Machine
OK	Ordinary Kriging
RK	Regression Kriging
VIF	Variance Inflation Factor
TWI	Topographic Wetness Index

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. *Advances in Meteorology* **2016**, p. 11.
2. Prihodko, L.; Goward, S.N. Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment* **1997**, *60*, 335–346.
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.
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 Sensing* **2015**, *7*, 6005–6025.
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. *Journal of Applied Meteorology and Climatology* **2017**, *56*, 803–814.
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.
7. Vogt, J.; Viau, A.; Paquet, F. Mapping regional air temperature fields using satellite-derived surface skin temperatures. *International Journal of Climatology* **1997**, *17*, 1559–1579.
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.
9. Jang, J.; Viau, A.; Anctil, F. Neural network estimation of air temperatures from AVHRR data. *Int. J. Remote Sens.* **2004**, *25*, 4541–4554.
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.
11. Benali, A.; Carvalho, A.; Nunes, J.; Carvalhais, N.; Santos, A. Estimating Air Surface Temperature in Portugal Using MODIS LST Data. *Remote Sensing of Environment* **2012**, *124*, 108–121.
12. Kim, D.; Han, K. Remotely Sensed Retrieval of Midday Air Temperature Considering Atmospheric and Surface Moisture Conditions. *International Journal of Remote Sensing* **2013**, *34*, 247–263.
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.
14. Xu, Y.; Knudby, A.; Ho, H.C. Estimating daily maximum air temperature from MODIS in British Columbia, Canada. *International Journal of Remote Sensing* **2014**, *35*, 8108–8121.
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.
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.
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.
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.
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.
20. Nemani, R.R.; Running, S.W. Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. *Journal of Applied Meteorology* **1989**, *28*, 276–284.
21. Goward, S.N.; Waring, R.; Dye, D.; Yang, J. Ecological remote sensing at OTTER: Satellite macroscale observations. *Ecological Applications* **1994**, *4*, 322–343.
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.
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.

24. Sørensen, R.; Zinko, U.; Seibert, J. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences* **2006**, *10*, 101–112.
25. Hofierka, J.; Suri, M. The solar radiation model for Open source GIS: implementation and applications. International GRASS users conference, 2002.
26. 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. *Spatial Statistics* **2015**, *14*, 70–90.
27. Busetto, L.; Ranghetti, L. MODISr: An R package for automatic preprocessing of MODIS Land Products time series. *Computers & Geosciences* **2016**, *97*, 40–48.
28. Priyadarshi, N.; Chowdary, V.; Srivastava, Y.; Das, I.C.; Jha, C.S. Reconstruction of time series MODIS EVI data using de-noising algorithms. *Geocarto International* **2017**, pp. 1–19.
29. Huete, A.; Justice, C.; Van Leeuwen, W. MODIS vegetation index (MOD13). *Algorithm theoretical basis document* **1999**, *3*, 213.
30. Gu, J.; Li, X.; Huang, C.; Okin, G.S. A simplified data assimilation method for reconstructing time-series MODIS NDVI data. *Advances in Space Research* **2009**, *44*, 501–509.
31. Metz, M.; Andreo, V.; Neteler, M. A New Fully Gap-Free Time Series of Land Surface Temperature from MODIS LST Data. *Remote Sensing* **2017**, *9*, 1333.
32. Mu, Q.; Zhao, M.; Running, S.W. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment* **2011**, *115*, 1781–1800.
33. Stisen, S.; Sandholt, I.; Nørgaard, A.; Fensholt, R.; Eklundh, L. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. *Remote Sens. Environ.* **2007**, *110*, 262–274.
34. Zhu, W.; Lu, A.; Jia, S. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sens. Environ.* **2013**, *130*, 62–73.
35. Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D.; Team, R.C. nlme: Linear and Nonlinear Mixed Effects Models. Technical report, 2018.
36. Zuur, A.; Ieno, E.; Walker, N.; Saveliev, A.; Smith, G. *Mixed Effects Models and Extensions in Ecology with R*; Springer, 2009; p. 549.
37. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer, 2013.
38. Murphy, K. *Machine Learning. A probabilistic approach*; The MIT Press, 2012.
39. 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, 2018.
40. Wright, M.N.; Ziegler, A. ranger. A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* **2017**, *77*(1), 1–17.
41. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*(3), 18–22.
42. 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 Sensing* **2015**, *74*(4), 4651–4677.
43. Burrough, P.; McDonnell, R.; Lloyd, C. *Principles of Geographical Information Systems*; Oxford, 2015.
44. Hiemstra, P.; Pebesma, E.; Twenhöfel, C.; Heuvelink, G. Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. *Computers & Geosciences* **2008**, *35*(8), 1711–1721.
45. Cressie, N. *Statistics for Spatial Data (Revised Ed)*; Wiley, 1993.
46. 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.; Pierce, S.A.; Robson, B.; Seppelt, R.; Voinov, A.A.; Fath, B.D.; Andreassian, V. Characterising performance of environmental models. *Environmental Modelling & Software* **2013**, *40*, 1 – 20.
47. Legates, D.R.; McCabe, G.J. Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **1999**, *35*, 233–241. doi:10.1029/1998WR900018.