

Article

Stochastic Spatio-Temporal Models for Analysing NDVI Distribution of GIMMS NDVI3g Images

Ana F. Militino ^{1,2,3,*}, Maria Dolores Ugarte ^{1,2,3} and Unai Pérez-Goya ¹

¹ Department of Statistics and Operations Research, Public University of Navarre, 31006 Pamplona, Spain; lola@unavarra.es (M.D.U.); unai.perez@unavarra.es (U.P.-G.)

² Institute for Advanced Materials (InaMat), Public University of Navarre, 31006 Pamplona, Spain

³ Department of Mathematics, Spanish Open University (UNED), 31006 Pamplona, Spain

* Correspondence: militino@unavarra.es; Tel.: +34-948-169-206

Academic Editors: Lenio Soares Galvao, Clement Atzberger and Prasad S. Thenkabail

Received: 29 June 2016; Accepted: 8 January 2017; Published: 15 January 2017

Abstract: The normalized difference vegetation index (NDVI) is an important indicator for evaluating vegetation change, monitoring land surface fluxes or predicting crop models. Due to the great availability of images provided by different satellites in recent years, much attention has been devoted to testing trend changes with a time series of NDVI individual pixels. However, the spatial dependence inherent in these data is usually lost unless global scales are analyzed. In this paper, we propose incorporating both the spatial and the temporal dependence among pixels using a stochastic spatio-temporal model for estimating the NDVI distribution thoroughly. The stochastic model is a state-space model that uses meteorological data of the Climatic Research Unit (CRU TS3.10) as auxiliary information. The model will be estimated with the Expectation-Maximization (EM) algorithm. The result is a set of smoothed images providing an overall analysis of the NDVI distribution across space and time, where fluctuations generated by atmospheric disturbances, fire events, land-use/cover changes or engineering problems from image capture are treated as random fluctuations. The illustration is carried out with the third generation of NDVI images, termed NDVI3g, of the Global Inventory Modeling and Mapping Studies (GIMMS) in continental Spain. This data are taken in bymonthly periods from January 2011 to December 2013, but the model can be applied to many other variables, countries or regions with different resolutions.

Keywords: kriging; spatial statistics; stochastic modelling

1. Introduction

The Normalized Difference Vegetation Index (NDVI) reflects vegetation growth and it is closely related to the amount of photosynthetically absorbed active radiation as indicated in [1,2]. It is calculated using the radiometric information obtained for the red (R) and near-infrared (NIR) wavelengths of the electromagnetic spectrum in the following way: $NDVI = ((NIR) - R) / ((NIR) + R)$ (see [3] for more details). As mentioned in [4], this parameter is sensitive to the blueness of the observed area, which is closely related to the presence of vegetation. Although numerical limits of NDVI can vary for the vegetation classification, it is widely accepted that negative NDVI values correspond to water or snow. NDVI values close to zero could correspond to bare soils, yet these soils can show a high variability. Values between 0.2 and 0.5 (approximately) to sparse vegetation, and values between 0.6 and 1.0 conform to dense vegetation such as that found in temperate and tropical forests or crops at their peak growth stage.

However, in remote sensing data, atmospheric conditions or cloud presence alter the correct estimation of NDVI. A large number of papers have been devoted to completing, reconstructing and predicting the spatial and temporal dynamics of the future NDVI distribution using a time

series of images (see, for example, [5–10]). These studies are mainly based on including temporal correlation of individual pixels at different resolutions but ignoring spatial dependence among them. Perhaps the most broadly used method for analysing NDVI temporal changes is the non-parametric Mann–Kendall test (see, for example, [11–14]). When plotting significant changes, a discrete pixel by pixel map of the NDVI trend changes is obtained. Figure 1 shows the coloured pixels where significant trend NDVI changes have been detected in continental Spain from October 2011 to December 2013. This discretization comes because the Mann–Kendall test only assumes a time dependence within the same pixel across years, but it does not encompass the spatial dependence among neighbour pixels. Therefore, unless random disturbances occur because of fire events, land-use/cover changes, crop rotation, land degradation or many other causes, we expect that close locations present similar trend changes. Some improvements of this test have been also provided. For example, Neeti and Eastman [15] introduced the contextual Mann–Kendall approach for assessing the trend significance of the NDVI time series, by removing serial correlation through a prewhitening process. It consists of evaluating the trend at a regional scale comprised of the 3×3 neighborhood around each pixel, providing a smoother picture of the trend changes but without completely avoiding the final discretization of images.

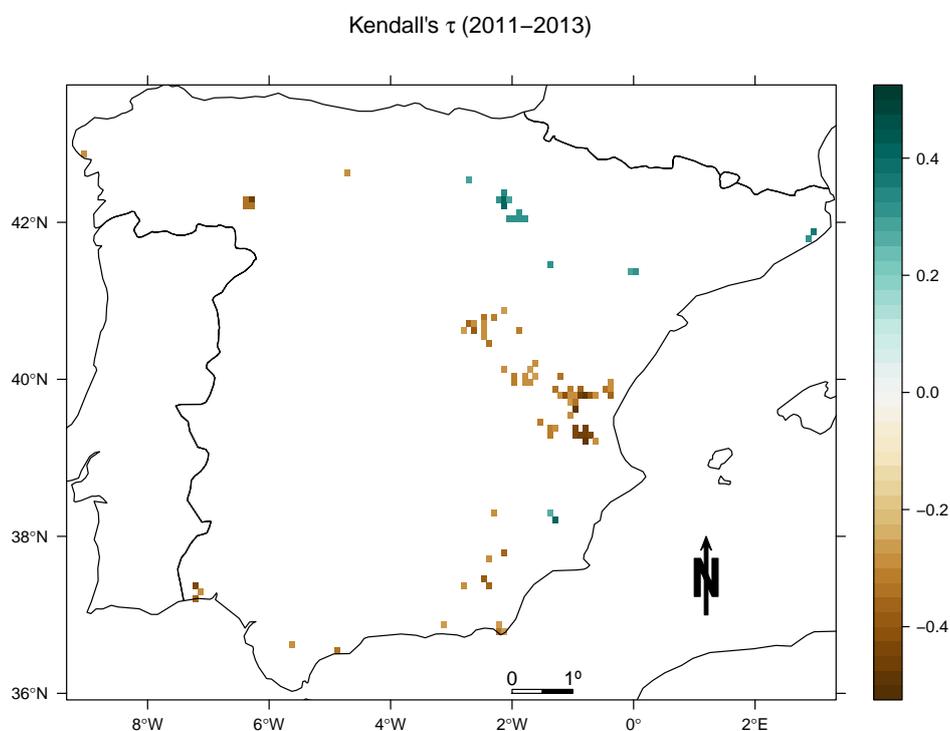


Figure 1. Coloured pixels correspond to significant trend changes of the third generation of the normalized difference vegetation index (NDVI3g) of the Global Inventory Modeling and Mapping Studies (GIMMS), with a Mann–Kendall test in Spain from October 2011 to December 2013.

To obtain a smooth map of the NDVI trend changes, only a few alternatives are found in the spatio-temporal literature. For example, Xu et al [16] proposed a spatio-temporal iteration method to reconstruct contaminated pixels of the MODIS13Q1 NDVI time series dataset. The method is not stochastic but based on numerical approximations. The authors first compute contaminated pixels of NDVI through linear interpolation of adjacent high-quality pixels, and, later, the NDVIs of the remaining contaminated pixels are determined based on the NDVI of a high-quality pixel located in the same ecological zone, showing the most similar NDVI change trajectories. They iterate the process using the estimated NDVIs as high-quality pixels to predict undetermined NDVIs of contaminated pixels until the NDVIs of all contaminated pixels are estimated. The well-known proposal by Eklundh and Jönsson [17] provides the TIMESAT [18] free program, designed primarily for analyzing the time

series of satellite data. It uses an adaptive Savitzky–Golay filtering and methods based on upper envelope weighted asymmetric Gaussian and double logistic model functions. This program can be downloaded from [19] and it has been used in this paper for comparison purposes.

The use of stochastic spatio-temporal models (see [20]) is scarce with satellite data. Hengl et al. [21] use a spatio-temporal regression kriging for smoothing land surface temperature data of MODIS MOD11A2. The time series data consists of 46 daytime and nighttime eight-day composite land surface temperature (LST) images in 2008 and the ground data of 159 Croatia meteorological stations. The difficulty of this method lies in fitting the variogram necessary for modelling the spatio-temporal dependence that increases depending on the number of periods and stations. As an alternative, we propose a stochastic state-space model that simultaneously exploits dependencies across space and time. Figure 2 shows the graphical summary followed in the paper.

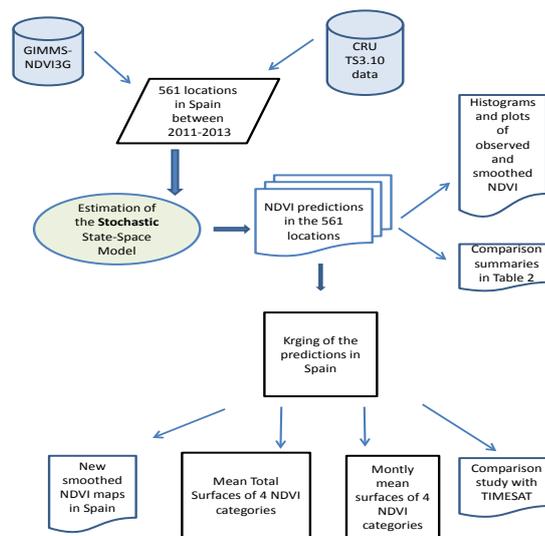


Figure 2. Graphical summary of the computational processes followed in this paper.

2. Data

Global Inventory Modeling and Mapping Studies Normalized Difference Vegetation Index of third generation (GIMMS NDVI3g), between 2011 and 2013 are used in this paper for analysing the spatio-temporal NDVI distribution in continental Spain. GIMMS NDVI3g data are bi-weekly composite NDVI data. The composite images are obtained by the Maximum Value Compositing (MVC). It has been shown to be more accurate than the GIMMS NDVI data for monitoring vegetation activity and phenological change [22]. More details on GIMMS NDVI3g can be found in [23]. Figure 3 shows the GIMMS NDVI3g image over the Earth in the first fifteen days of October 2011. The GIMMS NDVI3g time series is an improved normalized difference vegetation index (NDVI) data set produced from Advanced Very High Resolution Radiometer (AVHRR) instruments that extends from 1981 to the present onboard NOAA satellite. It has been largely used along recent years, for example in [24,25]. GIMMS NDVI3g data can be downloaded from [26]. The data have flags accounting for additional information about the pixel quality. These flags can vary between 1 and 7, where 1 or 2 indicates good quality, numbers between 3 and 6 indicate different kinds of processing, and 7 indicates missing data.

The spatial resolution of these data is 8 km at the equator, but it has been corrected for calibration, view geometry, volcanic aerosols, and other effects not related to vegetation changes, providing sometimes unrealistic values of the NDVI when downscaling the NDVI index to smaller regions [27]. Figure 4 shows the 72 scenes of original GIMMS NDVI3g data cropped to continental Spain from January 2011 to December 2013 and plotted in the free statistical software R [28]. In particular, library

`gimms` [29] has been used for reading the images in R, yet it can also be done with library `raster` [30]. According to this figure, western and northern Spanish regions have the maximum limit of NDVI, even in the summer, which is usually the driest season, which is an unlikely case in this country, particularly in the central western regions.

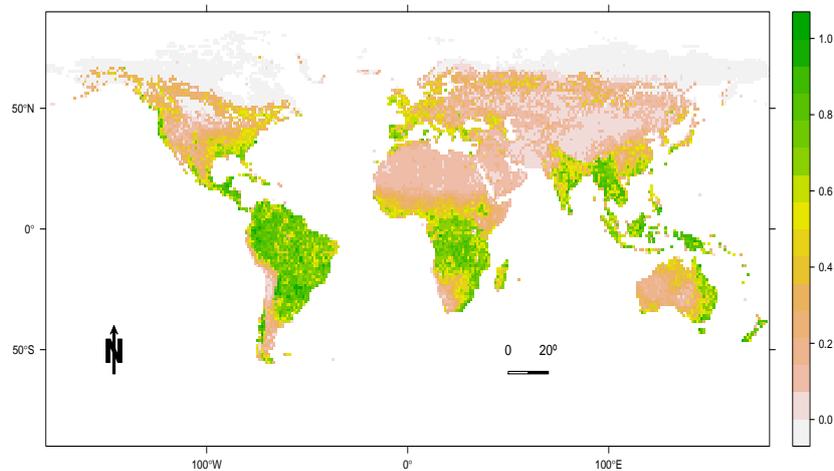


Figure 3. Image of the third generation of the normalized difference vegetation index (NDVI3g) corresponding to the Global Inventory Modeling and Mapping Studies (GIMMS) in October 2011.

In this paper, climate data from the Climatic Research Unit (CRU) are additionally used as auxiliary information in the stochastic space-time model to calibrate satellite data. CRU data are the result of processed meteorological data that can be downloaded from [31]. This is a gridded climate data set of monthly observations taken at meteorological stations across the world land areas and referred to as CRU TS3.10. Station anomalies were interpolated into 0.5 degrees latitude/longitude grid cells covering the global land surface (excluding Antarctica) and combined with an existing climatology database to obtain absolute monthly values. Detailed information can be found in [32]. From Figure 5 left we can see the grid locations of CRU TS3.10 data where auxiliary meteorological information is drawn. This database contains the following auxiliary variables:

```

cld cloud cover percentage (%) x 10
dtr diurnal temperature range degrees Celsius x 10
frs frost day frequency days x 100
pet potential evapotranspiration millimetres per day x 10
pre precipitation millimetres per month x 10
tmp daily mean temperature degrees Celsius x 10
tmn monthly average daily minimum temperature degrees Celsius x 10
tmx monthly average daily maximum temperature degrees Celsius x 10
vap vapour pressure hectopascals (hPa) x 10
wet wet day frequency (rain days per month) days x 100

```

In this list, only *cld*, *frs*, *pre*, *tmx*, *vap* and *wet* variables are used because *dtr*, *pet*, *tmp* and *tmn* can be derived from the rest, and the stochastic spatio-temporal models require independent auxiliary variables for avoiding multicollinearity [33]. The six chosen variables will be called covariates hereafter.

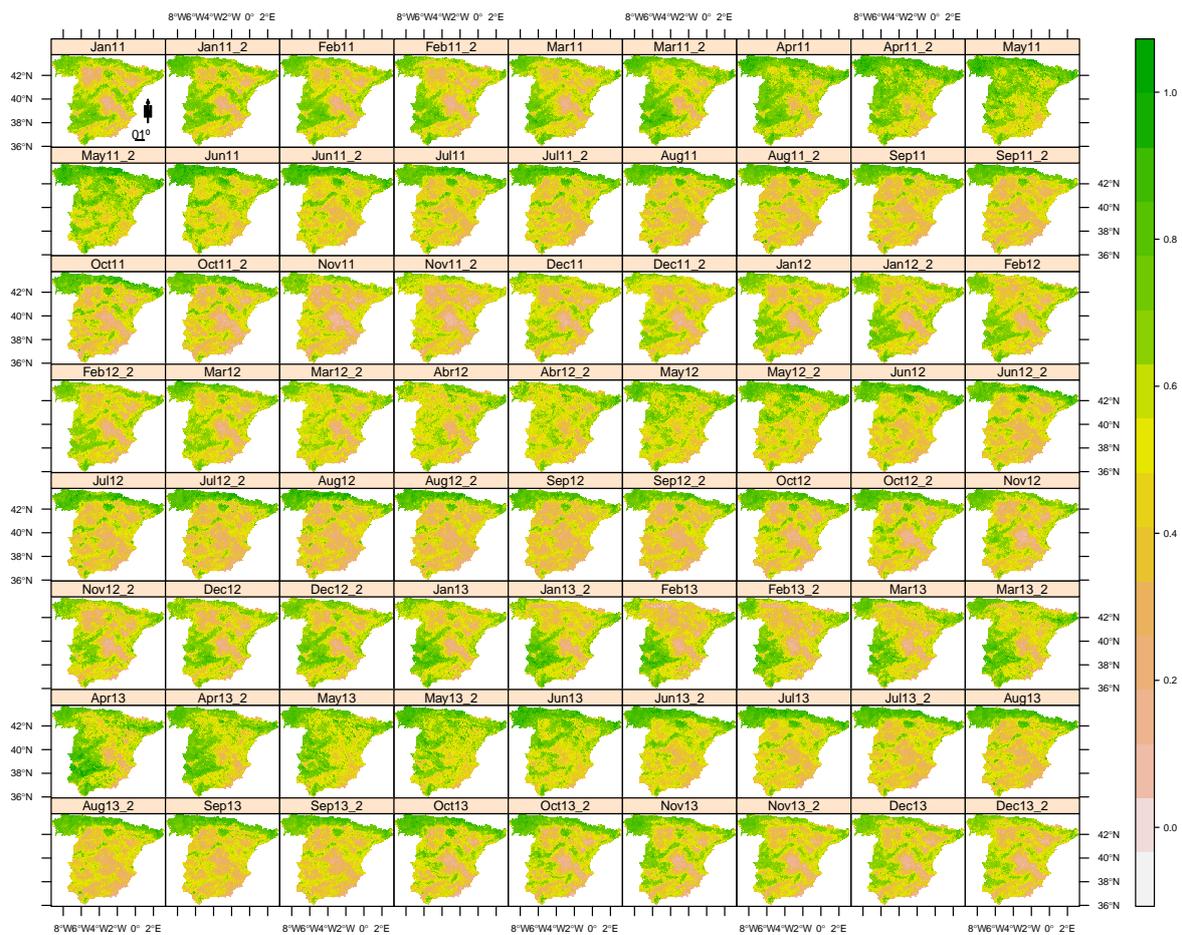


Figure 4. GIMMS NDVI3g images in continental Spain from January 2011 to December 2013.

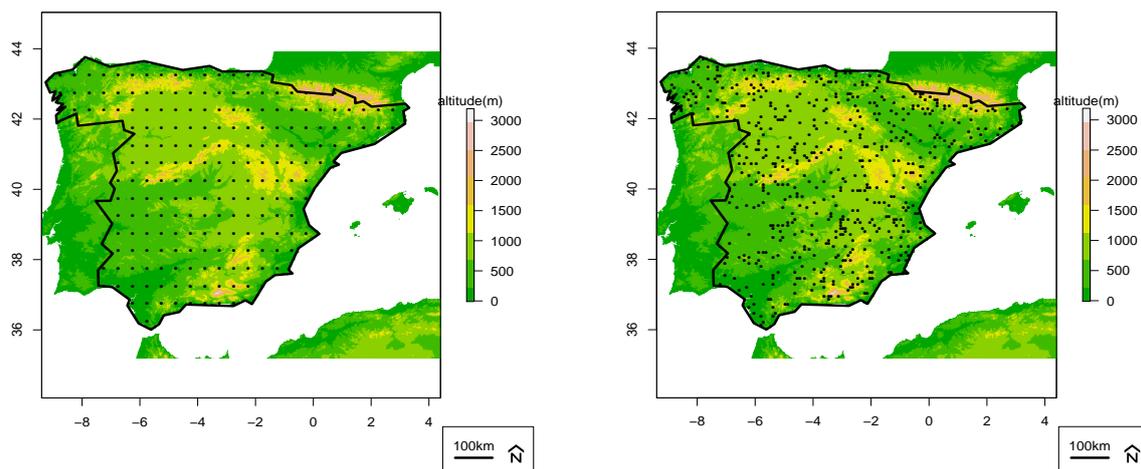


Figure 5. (Left) Grid locations of the Climatological Research Unit (CRU) TS3.10 meteorological data where auxiliary information is drawn for calibrating satellite data and (right) sampled locations used for estimating the state-space model.

From the GIMMS NDVI3g data, we randomly choose $n = 561$ locations among those with good flag attributes (indicating high quality). These locations are plotted on Figure 5right. In these locations, we extract the meteorological information of the six covariates. As the temporal resolution of CRU

data differs from GIMMS NDVI3g data (monthly versus bi-monthly data), we decided to transform CRU monthly data in bi-monthly data. In particular, *cld*, *frs*, *tmax* and *vap* remain invariant in the corresponding fifteen days, but *pre*, *frs* and *wet* are divided by two. Next, the CRU covariates and the altitude of the sampled locations are organized in a 561×433 matrix. The first column corresponds to the height values of the n sampled observations, and the rest are blocks of 72 periods by six covariates. The number of sampled locations have been chosen after checking different sizes between 300 and 1000 locations. From 300 locations, similar results have been obtained. This number is closely related to the meteorological data resolution because meteorological data must be drawn at these sampled locations, yet only a limited number of 211 pixels of CRU TS3.10 data are inside continental Spain. It means that only 211 different sets of covariates are available for being used in the model, and, then, negligible differences in model coefficient estimates are found when increasing the number of sampled locations.

3. Material and Methods

The State-Space Model

The state-space model is a very well-known mathematical tool used in dynamical systems. It became very used in econometrics since the publication of [34], and, more recently, Fassò and Cameletti [35] have developed it in a spatio-temporal context for modelling environmental space–time data. The model was originally applied to predict air concentrations and to deal with error measurements in instruments. In this paper, the state-space model is a spatio-temporal linear model that simultaneously accounts for the spatial and temporal dependence of the NDVI. It is a hierarchical model in two steps defined by two equations: the transition Equation (1) and the state Equation (2). Here, the first equation explains a linear regression between NDVI and the covariates. In this example, the covariates are the meteorological variables and the altitude. The second equation expresses the temporal dependence. More precisely, let us denote z_{st} as the sampled NDVI value at location s and time t . The stochastic process at n locations s_1, \dots, s_n and T time points t_j , from $j = 1, \dots, T$, is represented by $\mathbf{z}_{st} = (z(s_1, t_1), z(s_1, t_2), \dots, z(s_1, t_T), \dots, z(s_n, t_1), \dots, z(s_n, t_T))'$. In this case, $T = 72$ and $n = 561$ locations randomly chosen inside continental Spain. The model is given by

$$\mathbf{z}_{st} = \beta_0 + \beta_1 \mathbf{x}_{1s} + \beta_2 \mathbf{x}_{2st} + \dots + \beta_7 \mathbf{x}_{7st} + \mathbf{v}_t + \boldsymbol{\epsilon}_{st}, \quad \boldsymbol{\epsilon}_{st} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(d)), \quad (1)$$

$$\mathbf{v}_t = \mathbf{G} \mathbf{v}_{t-1} + \boldsymbol{\eta}_t, \quad \mathbf{v}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}), \quad (2)$$

where β_i , $i = 0, \dots, 7$ are the coefficients to be estimated. The first covariate \mathbf{x}_{1s} is time invariant and corresponds to the altitude of the n sampled locations. The rest of the covariates \mathbf{x}_{ist} , $i = 2, \dots, 7$ are the spatio-temporal meteorological covariates: maximum temperature, frost day frequency, precipitation, wet day frequency, cloudy cover percentage, and vapor pressure respectively, depending upon the location s and time t . The unobservable latent temporal process, \mathbf{v}_t , takes account of the temporal dynamics of data through an autoregressive process. It means that the current state \mathbf{v}_t depends on the previous state \mathbf{v}_{t-1} in the state equation through a transition matrix \mathbf{G} . The initial $T \times 1$ state vector, \mathbf{v}_0 is assumed to be normally distributed with mean $\boldsymbol{\mu}_0$ and covariance $\boldsymbol{\Sigma}_0$.

The spatial dependence is accounted for in the covariance structure of the model error, $\boldsymbol{\epsilon}_{st}$, given by $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(d)$. It can be estimated using well-known covariance functions such as the Matérn, the exponential or the spherical covariance functions [36]. The covariance function depends on the Euclidean distance $d = \|\mathbf{s}_i - \mathbf{s}_j\|$, and, therefore, the sampled locations must be Universal Transverse Mercator (UTM) projected. It is invariant to translations, so \mathbf{z}_{st} is assumed to be a second-order stationary process. The additive $T \times 1$ state-estimation errors, $\boldsymbol{\eta}_t$, and the $s \times 1$ measurement errors $\boldsymbol{\epsilon}_{st}$ are uncorrelated Gaussian white noises with zero mean and covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(d)$, respectively. Finally, $\boldsymbol{\eta}_t$ quantifies the uncertainty of the state estimate given the n observations. The transition equation incorporates the spatial dependence and the state equation takes into account

the temporal dependence. Therefore, this state-space model can be interpreted as a spatio-temporal kriging model with a separable spatio-temporal covariance function.

The state-space model is implemented in the R statistical software package `Stem` [37]. Similar state-space models were used by [38] to estimate model parameters and missing data of river basin runoff values, and by [39] to interpolate daily rainfall in Navarre (Spain). This package uses the function `Stem.Estimation` to carry out the iterations of the EM algorithm [40] until convergence. Each iteration calls the function `kalman` to perform both the E-step and the M-step. The exponential covariance function is also assumed for $\Sigma_{\epsilon}(d)$. The maximization process of the likelihood is done with the Kalman filter. We recommend using the coefficient estimates of the multiple linear regression model without assuming spatial dependence as initial values of the coefficients in the EM algorithm. When the state-space model is fitted, the β coefficients are obtained and tested. Additional programming is necessary for calculating *NDVI* predictions for the whole Spain.

Before running the state-space model, the temporal and the spatial dependence need to be explored. Unfortunately, there are no statistical tools for checking jointly the spatio-temporal dependence. Therefore, it can be checked only marginally. Figure 6 shows the autocorrelation function of the first 6 *NDVI* pixels, although similar results are obtained in the rest. Excluding the first vertical bar that is always equal to one because it is the autocorrelation of a pixel with itself, all the locations have at least one vertical bar above the blue dotted horizontal line, showing that the temporal dependence is significant in at least one lag. As expected, a slight seasonality is also present in these data, yet it can change from one pixel to another. To check marginally the spatial dependence, the Moran test [41] is used. This test has been computed for every one of the 72 GIMMS *NDVI*-3g Spanish scenes. It varies between 0.72 and 0.84, indicating a strong spatial autocorrelation. The `acf` function from library `base` [28] and `Moran` function from library `raster` [30] have been used in this step.

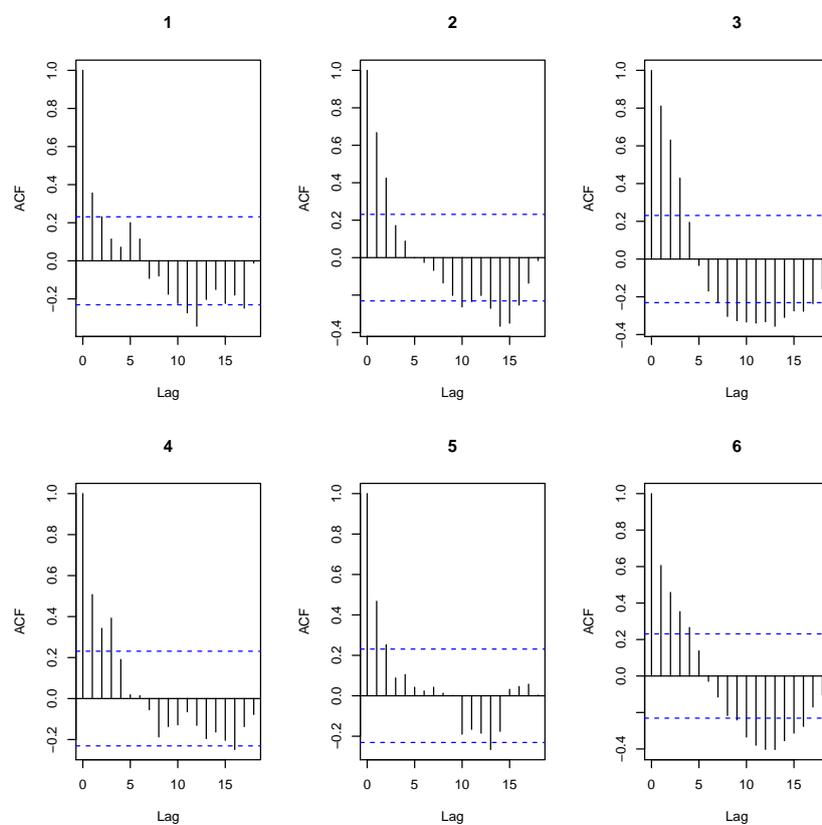


Figure 6. Autocorrelation function in six sampled locations of raw GIMMS *NDVI*3g data.

4. Results

Classical statistical tools are used for checking the statistical significance of the model coefficients. Table 1 shows the estimates, the standard errors, the t -values, and the confidence intervals of the state-space model coefficients. Standard errors are obtained by bootstrapping 10 replicates, but similar results are derived when increasing the number of replicates. All the coefficients are statistically significant because no one of the confidence intervals contain the zero value, except for the *wet* variable. Different random sets of 561 sampled locations have been essayed with similar results. In some cases, the *wet* covariate is statistically significant but with a very small estimate. Therefore, this covariate has been kept in the model, yet we know that it has a negligible impact in the predictions. Interpretation of sign estimates allows to conclude that NDVI is positively correlated with altitude, precipitation, and number of cloud days. However, NDVI decreases when maximum temperature or vapour pressure increase as expected. Meteorological covariates have been divided by 100 and altitude by 1000 because scaling covariates help to avoid singularities in the process of inverting matrices. Maximum temperature could be substituted by the average or minimum temperature without altering significantly the model estimation and the predictions. The model has been statistically validated testing the normality of the residuals.

Table 1. Estimates (Estimate), standard errors (SE), t -values (T-Stat.), lower (CI_low) and upper (CI_upp) limits of the 95% confidence intervals of the state-space model coefficients.

	Estimate	SE	T-Stat.	CI_low	CI_upp
(intercept) β_0	1.1343	0.0086	131.9563	1.1176	1.1435
(height) β_1	0.0471	0.0027	17.2019	0.0425	0.0501
(tmax) β_2	-0.1235	0.0039	-32.0501	-0.1313	-0.1216
(frs) β_3	-0.0153	0.0011	-14.3707	-0.0163	-0.0135
(wet) β_4	-0.0007	0.0008	-0.8950	-0.0010	0.0011
(prec) β_5	0.0190	0.0011	17.7825	0.0176	0.0209
(cld) β_6	0.0142	0.0010	13.7122	0.0116	0.0146
(vap) β_7	-0.0176	0.0070	-2.5172	-0.0208	-0.0013

For checking the model, we firstly compare sampled versus predicted data both in a unique period for all of the 561 locations, and, separately, in every one of the 72 bi-monthly periods. The overall summary of the sampled and predicted values of NDVI from 2011 to 2013 are shown in Table 2, where we can observe that the model does not only provide the same average for sampled and predicted values, but also similar quantile values. The smoothing process crosses over the most extreme values as expected. The state-space model predictions not only follow the pattern of GIMMS NDVI3g data in the overall period (2011–2013), but also in everyone of the 72 bi-monthly periods, as it is shown in the histograms of sampled NDVI values (Figure 7) and the corresponding predictions (Figure 8). Similarity between these figures is evident. In addition, Figure 9 plots sampled versus predicted NDVI data in the 72 periods, exhibiting also a close proximity between them. Therefore, the good performance of the model in sampled data is not only shown in summary statistics but also in all of the sampled locations. Later, an ordinary kriging was applied in every one of the 72 bi-monthly periods to get an overall image of the whole continental Spain. Library *geostatsp* [42] has been used in this step. Figure 10 shows the monthly predictions obtained by averaging the bi-monthly predictions. To complete the validation process, we compare these results to the documented information retrieved from the the Spanish National Agency of Meteorology (AEMET) [43] and the Spanish CRU TS3.10 meteorological data.

Table 2. Minimum (Min.), first quartile (1st Qu.), median (Median), mean (Mean), third quartile (3rd Qu.), and maximum (Max.) of the sampled and state-space smoothed NDVI3g data.

Summary	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
sampled NDVI	0.0140	0.4130	0.5410	0.5421	0.6740	1.0000
state-space smoothed NDVI	0.0867	0.4211	0.5392	0.5421	0.6604	0.9580

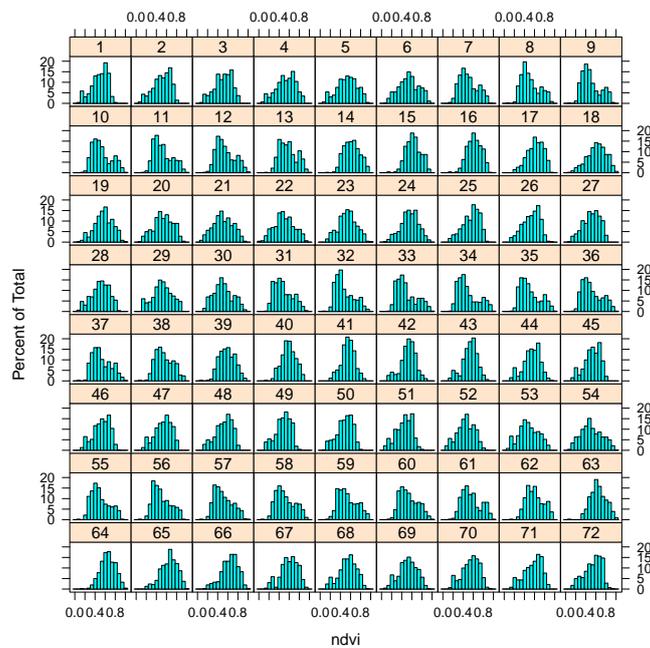


Figure 7. NDVI histograms of the 561 sampled locations in the 72 periods from October 2011 to December 2013.

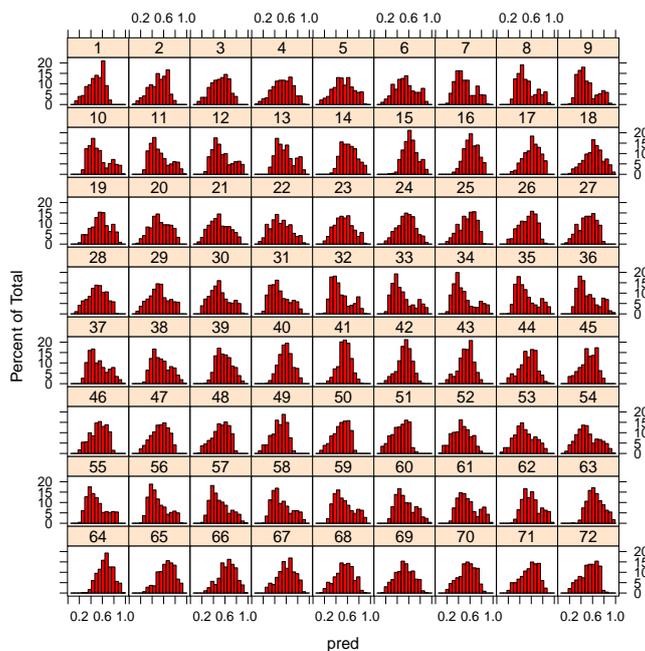


Figure 8. Predicted NDVI histograms of the 561 sampled locations in the 72 periods from October 2011 to December 2013.

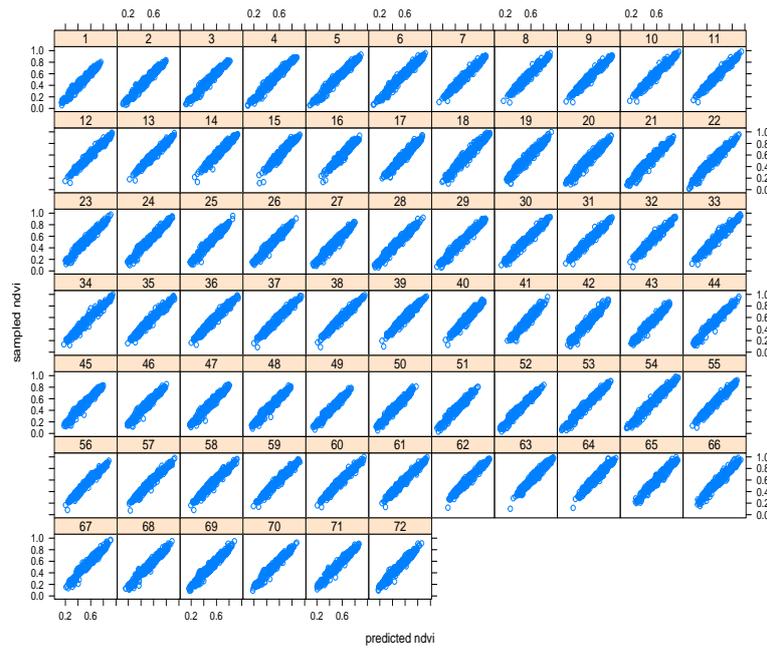


Figure 9. Sampled NDVI vs. predicted NDVI data of the 561 locations in the 72 periods from October 2011 to December 2013.

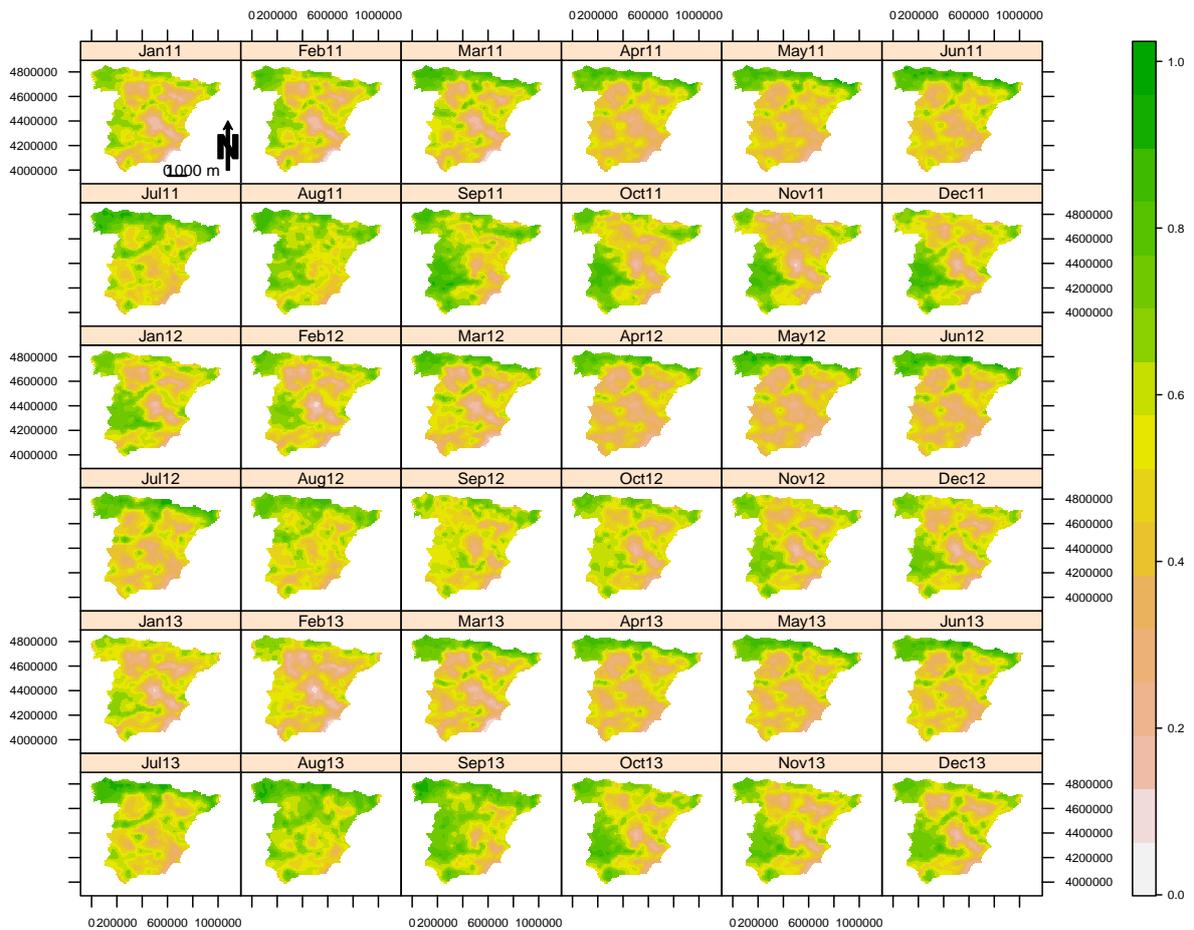


Figure 10. Smoothed NDVI in continental Spain from January 2011 to December 2013.

Spain is the fifth largest country in Europe with an extension of 505,000 km² and an average altitude of 650 m, the third highest country in Europe. It has three climatological regions. The Mediterranean region with dry and warm summers and cool to mild, wet winters. The oceanic region located in the North of Spain and characterised by relatively mild winters and warm summers, and the semiarid region located in the southeastern part of the country. In contrast to the Mediterranean region, the dry season continues beyond the end of the summer. This climatology affects the country vegetation, where differences can be appreciated among and within seasons.

AEMET reveals that the year 2011 was extremely hot with higher temperatures than the historical average (1971–2000). It was also very dry with 25% less rainfall in the North of Spain; however, spring was more humid than normal, particularly in March. The autumn rainfall was 10% lower than usual. The meteorological information drawn from the CRU TS3.10 data is summarized in Figure 11. On Figure 11left, monthly average temperatures are shown, and, on the right panel, the corresponding monthly average rainfall is given. Different colors are used for the different years and the historical mean is plotted in black in both panels. In the spring of 2011, high temperatures and abundant rainfall were also reported, yet the autumn was also very dry. Figure 10 shows the NDVI monthly Spanish predictions obtained by averaging the bi-monthly predictions given by the state-space model. In 2011, low values of NDVI are estimated in autumn but have very high values in spring, in agreement with AEMET and CRU TS3.10 data. The year 2012 was also very hot, especially in summer, and rainfall was 15% less than usual, except for autumn, and the region of Galicia, located in the northwest of Spain, which was extremely humid. These features are also observed in Figure 10, where a blue color is observed in December 2012 in Galicia, a brown color predominates in the main plateau of Spain, and northern regions show high values of NDVI, particularly in spring.

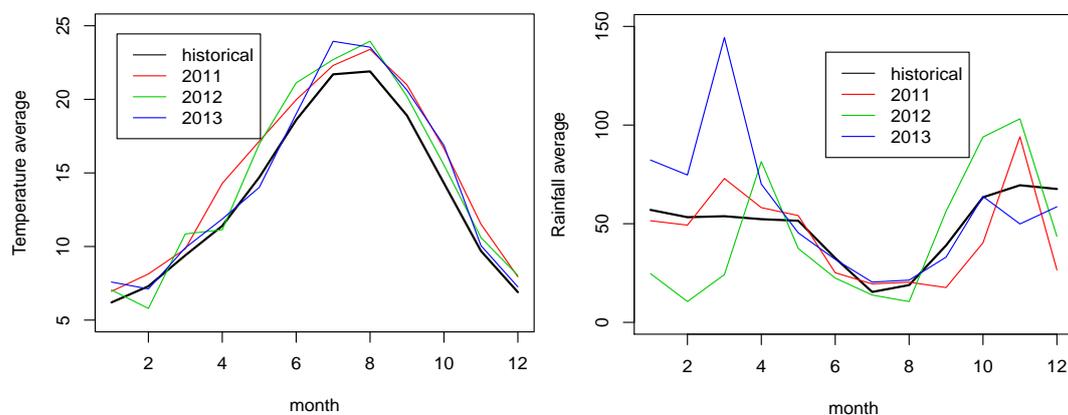


Figure 11. (left) Average temperatures in °C and (right) average rainfall in mm on the sampled locations jointly with the historical data.

The year 2013 was hot, but not as hot as 2011 and 2012. January and February were 30% more humid than normal, and March was extremely humid, with more than 340% more rain than the normal average. However, December was very dry. In Figure 11, CRU TS3.10 data also show a big pick of rainfall in winter that correspond to high values of smoothed NDVI in spring.

In summary, smoothed NDVI reveals a clear seasonality that intensifies the effect of spring vegetation in 2011, 2012 and 2013, where a higher level of rainfall than average is documented. The images preserve the pattern of the original ones but reduce the larger values of NDVI. As expected, the northern regions of Spain maintain higher values around 0.8 and 0.9, mainly in spring and early summer when temperatures and rainfall are more intense. Mountainous regions are also prone to the highest values, and the main plateau reaches values between 0.3 and 0.5, indicating the presence of bare soils or sparse vegetation. Therefore, the smoothed NDVI obtained through the state-space model is close to the climatological real scenario given in Spain between 2011 and 2013.

Overall, smoothed images are more sensitive to seasonal and specific meteorological changes than the original ones.

Table 3. Mean total surfaces of four NDVI categories in Spain between 2011 and 2013 in thousands of square kilometers.

	ndvi1	ndvi2	ndvi3	ndvi4
Raw GIMMS NDVI3g	10.88	203.08	195.42	95.42
State-space smoothed NDVI	4.58	206.36	209.57	84.29
TIMESAT Savitzky smoothed NDVI	23.83	202.71	191.32	86.94
TIMESAT Gaussian smoothed NDVI	23.30	202.57	193.31	85.62
TIMESAT double smoothed NDVI	23.30	202.57	193.31	85.62

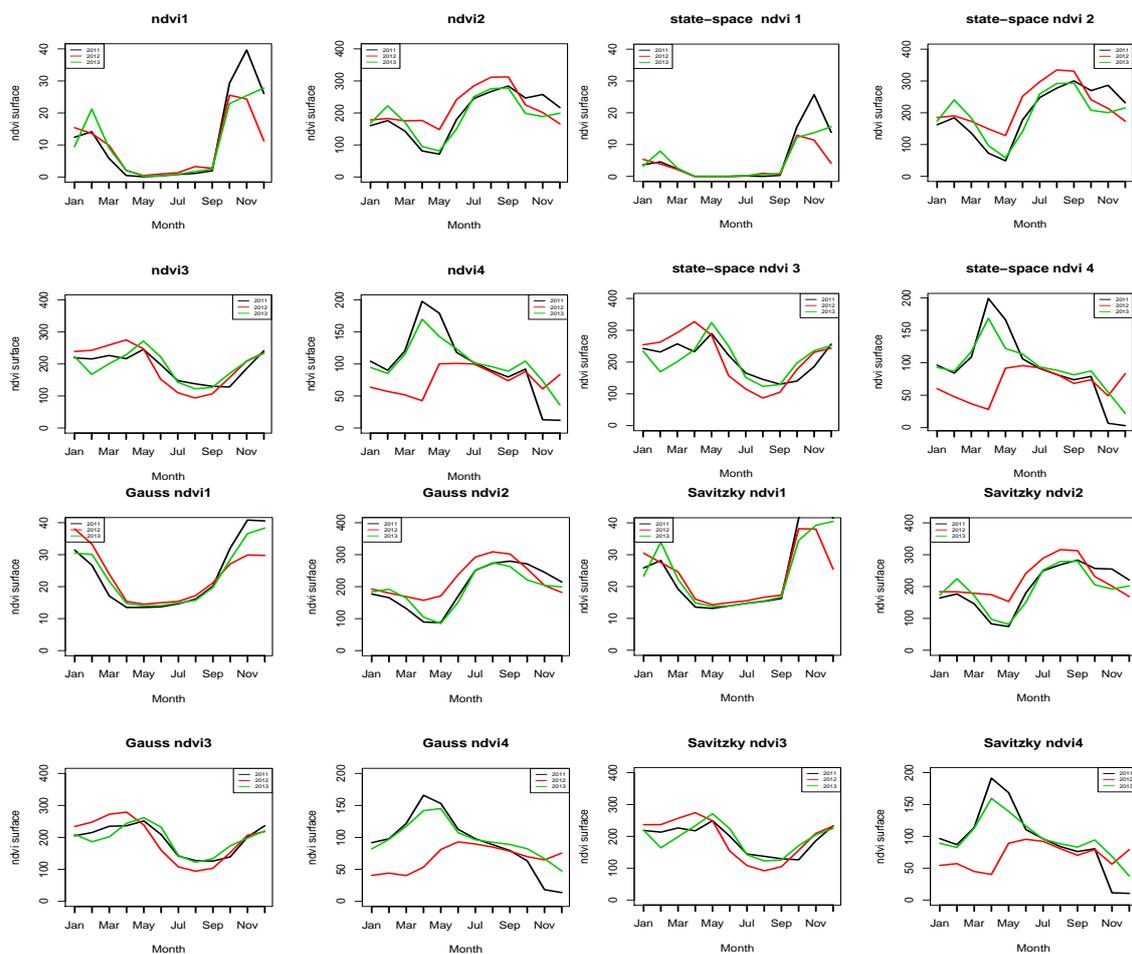


Figure 12. Monthly mean surfaces in the four NDVI categories with raw GIMMS NDVI3g data on the upper left, state-space smoothing on the upper right, Gaussian TIMESAT smoothing on the bottom left, and Savitzky–Golay smoothing on the bottom right.

Checking the performance of the smoothed NDVI with the real data is a difficult task because the NDVI is only estimated through satellite images. In this regard, comparisons of the mean estimated surfaces in four categories of NDVI are presented in Table 3: ndvi1 for data less than or equal to 0.2, ndvi2 for data greater than 0.2 and less than or equal to 0.5, ndvi3 for data greater than 0.5 and less than or equal to 0.7, and ndvi4 for data greater than 0.7. The mean total surfaces have been calculated with the raw GIMMS NDVI3g images, the state-space smoothed NDVI values, and three versions of the TIMESAT smoothed NDVI values from 2011–2013. The smoothing effect of the state-space model is

mainly shown in both *ndv1* and *ndv4* categories where smoothed NDVI mean total surfaces are lower than raw averages. These reductions have been added to the *ndv2* and *ndv3* categories. The three TIMESAT versions behave likewise providing close values to those obtained with the original images in both *ndv2* and *ndv3* categories, but important differences are found in the rest of the categories. Figure 12 shows the monthly mean surfaces of the raw GIMMS NDVI3g data, and the three smoothing versions: the state-space and two versions of TIMESAT, the Savitzky–Golay filtering and the Gaussian filtering data by years. The double logistic smoothing version of TIMESAT has been omitted because it is equal to the Gaussian version. The state-space approach follows the same pattern as the original data, but we can see how the *ndv1* category is smoothed mainly in winter and the *ndv4* category in spring and winter. The TIMESAT smoothing versions do not preserve well the pattern of the raw data in the smallest category, and bigger differences than with the state-space procedure can be found, particularly in the first category. In summary, the state-space approach preserves the monthly pattern of raw data by years and smooths mainly the lowest and upper categories. Additionally, this approach incorporates external information coming from CRU TS3.10 meteorological data and agrees with the information provided by the Spanish National Agency of Meteorology.

5. Discussion

GIMMS NDVI3g data have been widely used during the last decades for studying large scale trend changes over the years, mainly over continental or semi-continental regions. The latest version of the GIMMS NDVI data span the period July 1981 to December 2011 and is termed NDVI3g, but, in this paper, only the last three years are used. The temporal resolution of the 72 images between 2011 and 2013 has been chosen due to two main reasons. One comes with the computational problems in estimating the model that arises when enlarging this period, and the other one comes because auxiliary data at the same resolution is also needed, something difficult to find when we go back a long time. Nevertheless, higher resolutions can be also considered and future work is needed to encompass more years in the proposed model.

The actual resolution of 8 km at the equator is an attractive feature for monitoring changes of vegetation at any scale. Unfortunately, this resolution is not enough to warrant high precision images at smaller scales because images have been pre-processed, and, likely, there is also an important ocean border effect, as in the case of Spain. The Maximum Value Compositing (MVC) algorithm used to suppress atmospheric effects also minimizes significant problems associated with short-wave passive remote sensing of the Earth's surface, but the MVC technique itself has generated a second level of problems that must be addressed for proper interpretation of the NDVI MVC images. These are radiometric effects, which are relevant to the stratification assumption, and engineering effects, which are relevant to the MVC technique [44]. Similar situations can also be found with other NDVI global scenes coming from Terra MODIS or SPOT VGT (see, for example, [45], where evaluation of long trends vegetation coming from these satellites is made, revealing differences among them). High bias can also be found when using MVC in mountain regions (see [46]). Therefore, when down-scaling global scenes to country levels, as in the case of GIMMS NDVI3g in Spain, an adequate smoothing of NDVI data is needed for a proper interpretation of the spatio-temporal NDVI distribution. Similar situations can be found with other image processing techniques that may require smoothing procedures to analyse the data properly.

The main aim of this paper is to show the importance of considering both the spatial and temporal dependence for analyzing and smoothing NDVI data. The stochastic spatio-temporal model used here is a useful tool to capture space and time variability for simultaneously smoothing images. Smoothed images have been compared with TIMESAT that only uses temporal dependence. The state-space method outperforms this alternative, as it is able to reduce the most extreme values preserving the original pattern of raw data. The state-space model also provides the contribution of every covariate to predict NDVI. In this regard, it agrees with other studies such as [47–50], where it is shown that,

among climatic factors, precipitation and temperature influence both temporal and spatial patterns of NDVI.

However, there is an inherent difficulty in checking the performance of the stochastic model. Validation based on comparisons between sampled and predicted data is a common approach that evaluates the model goodness of fit, yet it is not enough to warrant a good performance when predicting new data. In this paper, this approach is satisfactory with the following limitation: sampled data are not necessarily the real data, and, then, when comparing predicted with sampled data, it does not mean that small differences between them correspond necessarily to high-quality predictions. This step can only be done when looking for vegetation changes previously documented. In the Results section, a detailed exploration of documented meteorological data shows a close agreement between the historical information and the smoothed NDVI images. Comparisons of smoothed trend changes with other studies are scarce. The most relevant is [51] where the authors investigate the NDVI trend changes that happened in the Iberian peninsula between 1981 and 2001 using GIMMS NDVI3g data but with a pixel by pixel approach.

6. Conclusions

This paper is focused on showing that a stochastic spatio-temporal model is a useful tool for overcoming random or fluctuations often present in satellite images, and these fluctuations could interfere with the detection of the NDVI trend changes. In the 2011–2013 period where GIMMS-NDVI3g images of continental Spain have been analyzed, we cannot detect trend changes in raw data, yet we have observed that smaller and greater values of NDVI are overestimated. However, in this application and after smoothing, we cannot detect trend changes either. We have shown that smoothed images are more sensitive to seasonal and specific meteorological changes than the original ones, yet they follow a similar pattern. Moreover, the smoothing method used in this paper provides a calibration method of satellite images with real data. A higher spatial and temporal resolution jointly with auxiliary data at the same resolution level could improve the model performance, but the computational cost will also increase. We are currently working on reducing the computational cost while obtaining accurate predictions in a sensible time.

Acknowledgments: The authors would like to thank both the editor and the referees for the constructive comments that led to improving this paper. This research was supported by the Spanish Ministry of Economy and Competitiveness (Project MTM2014-51992-R), the Government of Navarre (Project PI015, 2016), and by the Fundación Caja Navarra-UNED Pamplona (2016).

Author Contributions: Ana F. Militino did the design, the research and the review of the manuscript. Maria Dolores Ugarte contributed to the interpretation of the results and wrote the manuscript. Unai Pérez-Goya did the computational part and produced the Figures.

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

Abbreviations

The following abbreviations are used in this manuscript:

AEMET	Spanish National Agency of Meteorology
AVHRR	Advanced Very High Resolution Radiometer
CRU	Climatic Research Unit
EM	Expectation–Maximization
GIMMS NDVI3g	Third generation of Normalized Difference Vegetation Index of the Global Inventory
	Modeling and Mapping Studies
LST	Land Surface Temperature
MODIS	Moderate Resolution Imaging Spectroradiometer
MODIS11A2	Land Surface Temperature and Emissivity 8-Day L3 Global 1km from MODIS
MODIS13Q1	Vegetation Indices 16-Day L3 Global 250m from MODIS/TERRA
MVC	Maximum Value Compositing
NDVI	Normalized Difference Vegetation Index

SPOT VGT Vegetation of the Satellite Pour l'Observation de la Terre
 UTM Universal Transverse Mercator

References

1. Slayback, D.A.; Pinzon, J.E.; Los, S.O.; Tucker, C.J. Northern hemisphere photosynthetic trends 1982–1999. *Glob. Chang. Biol.* **2003**, *9*, 1–15.
2. Tucker, C.J.; Pinzon, J.E.; Brown, M.E.; Slayback, D.A.; Pak, E.W.; Mahoney, R.; Vermote, E.F.; El Saleous, N. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *Int. J. Remote Sens.* **2005**, *26*, 4485–4498.
3. Rouse, J., Jr.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309.
4. Sobrino, J.; Julien, Y. Global trends in NDVI-derived parameters obtained from GIMMS data. *Int. J. Remote Sens.* **2011**, *32*, 4267–4279.
5. Forkel, M.; Carvalhais, N.; Verbesselt, J.; Mahecha, M.D.; Neigh, C.S.; Reichstein, M. Trend change detection in NDVI time series: Effects of inter-annual variability and methodology. *Remote Sens.* **2013**, *5*, 2113–2144.
6. Tüshaus, J.; Dubovyk, O.; Khamzina, A.; Menz, G. Comparison of Medium Spatial Resolution ENVISAT-MERIS and Terra-MODIS time series for vegetation decline analysis: A case study in Central Asia. *Remote Sens.* **2014**, *6*, 5238–5256.
7. Klisch, A.; Atzberger, C. Operational Drought Monitoring in Kenya Using MODIS NDVI Time Series. *Remote Sens.* **2016**, *8*, 267.
8. Wang, R.; Cherkauer, K.; Bowling, L. Corn Response to Climate Stress Detected with Satellite-Based NDVI Time Series. *Remote Sens.* **2016**, *8*, 269.
9. Liu, Y.; Li, Y.; Li, S.; Motesharrei, S. Spatial and temporal patterns of global NDVI trends: Correlations with climate and human factors. *Remote Sens.* **2015**, *7*, 13233–13250.
10. Maselli, F.; Papale, D.; Chiesi, M.; Matteucci, G.; Angeli, L.; Raschi, A.; Seufert, G. Operational monitoring of daily evapotranspiration by the combination of MODIS NDVI and ground meteorological data: Application and evaluation in Central Italy. *Remote Sens. Environ.* **2014**, *152*, 279–290.
11. Li, Z.; Huffman, T.; McConkey, B.; Townley-Smith, L. Monitoring and modeling spatial and temporal patterns of grassland dynamics using time series MODIS NDVI with climate and stocking data. *Remote Sens. Environ.* **2013**, *138*, 232–244.
12. De Jong, R.; de Bruin, S.; de Wit, A.; Schaepman, M.E.; Dent, D.L. Analysis of monotonic greening and browning trends from global NDVI time series. *Remote Sens. Environ.* **2011**, *115*, 692–702.
13. Wu, D.; Wu, H.; Zhao, X.; Zhou, T.; Tang, B.; Zhao, W.; Jia, K. Evaluation of spatiotemporal variations of global fractional vegetation cover based on GIMMS NDVI data from 1982 to 2011. *Remote Sens.* **2014**, *6*, 4217–4239.
14. Sobrino, J.A.; Julien, Y.; Morales, L. Changes in vegetation spring dates in the second half of the twentieth century. *Int. J. Remote Sens.* **2011**, *32*, 5247–5265.
15. Neeti, N.; Eastman, J.R. A contextual mann-kendall approach for the assessment of trend significance in image time series. *Trans. GIS* **2011**, *15*, 599–611.
16. Xu, L.; Li, B.; Yuan, Y.; Gao, X.; Zhang, T. A Temporal-Spatial Iteration Method to Reconstruct NDVI Time Series Datasets. *Remote Sens.* **2015**, *7*, 8906–8924.
17. Jönsson, P.; Eklundh, L. TIMESAT—A program for analyzing time series of satellite sensor data. *Comput. Geosci.* **2004**, *30*, 833–845.
18. Eklundh, L.; Jonsson, P. *TIMESAT 3. 2 with Parallel Processing—Software Manual*; Lund University: Lund, Sweden, 2012.
19. TIMESAT. A Software Package to Analyse Time-Series of Satellite Sensor Data. Available online: <http://www.nateko.lu.se/TIMESAT/> (accessed on 10 January 2017).
20. Cressie, N.; Wikle, C.K. *Statistics for Spatio-Temporal Data*; John Wiley & Sons: New York, NY, USA, 2015.
21. Hengl, T.; Heuvelink, G.B.; Tadić, M.P.; Pebesma, E.J. Spatio-temporal prediction of daily temperatures using time series of MODIS LST images. *Theor. Appl. Climatol.* **2012**, *107*, 265–277.

22. Wang, J.; Dong, J.; Liu, J.; Huang, M.; Li, G.; Running, S.W.; Smith, W.K.; Harris, W.; Saigusa, N.; Kondo, H.; et al. Comparison of gross primary productivity derived from GIMMS NDVI3g, GIMMS, and MODIS in Southeast Asia. *Remote Sens.* **2014**, *6*, 2108–2133.
23. Pinzon, J.E.; Tucker, C.J. A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote Sens.* **2014**, *6*, 6929–6960.
24. Zhang, R.; Ouyang, Z.T.; Xie, X.; Guo, H.Q.; Tan, D.Y.; Xiao, X.M.; Qi, J.G.; Zhao, B. Impact of Climate Change on Vegetation Growth in Arid Northwest of China from 1982 to 2011. *Remote Sens.* **2016**, *8*, 364.
25. Yuan, X.; Li, L.; Chen, X.; Shi, H. Effects of precipitation intensity and temperature on NDVI-based grass change over Northern China during the period from 1982 to 2011. *Remote Sens.* **2015**, *7*, 10164–10183.
26. Ecocast. Monitoring, Modeling and Forecasting Ecosystem Change. Available online: <http://ecocast.arc.nasa.gov/data/pub/gimms/3g.v0/> (accessed on 17 January 2017).
27. Erasmi, S.; Schucknecht, A.; Barbosa, M.P.; Matschullat, J. Vegetation greenness in northeastern Brazil and its relation to ENSO warm events. *Remote Sens.* **2014**, *6*, 3041–3058.
28. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
29. Detsch, F. Gimms: Download and Process GIMMS NDVI3g Data, R Package Version 0.5.1. Available online: <https://cran.r-project.org/web/packages/gimms/gimms.pdf> (accessed on 10 January 2017).
30. Hijmans, R.J. Raster: Geographic Data Analysis and Modeling, R Package Version 2.5-2. Available online: <https://cran.r-project.org/web/packages/raster/raster.pdf> (accessed on 10 January 2017).
31. CRU TS3.10. Climatic Research Unit. Available online: <http://www.cru.uea.ac.uk/data> (accessed on 10 January 2017).
32. Harris, I.; Jones, P.; Osborn, T.; Lister, D. Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 Dataset. *Int. J. Climatol.* **2014**, *34*, 623–642.
33. Ugarte, M.D.; Militino, A.F.; Arnholt, A.T. *Probability and Statistics with R*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2015.
34. Durbin, J.; Koopman, S.J. *Time Series Analysis by State Space Methods*; Oxford University Press: New York, NY, USA, 2012.
35. Fassò, A.; Cameletti, M. The EM algorithm in a distributed computing environment for modelling environmental space–time data. *Environ. Model. Softw.* **2009**, *24*, 1027–1035.
36. Militino, A.F.; Ugarte, M.D. Assessing the covariance function in geostatistics. *Stat. Probab. Lett.* **2001**, *52*, 199–206.
37. Cameletti, M. Stem: Spatio-Temporal Models in R, R Package Version 1.0. Available online: <https://cran.r-project.org/web/packages/Stem/Stem.pdf> (accessed on 10 January 2017).
38. Amisigo, B.; Van De Giesen, N. Using a spatio-temporal dynamic state-space model with the EM algorithm to patch gaps in daily riverflow series. *Hydrol. Earth Syst. Sci. Discuss.* **2005**, *9*, 209–224.
39. Militino, A.; Ugarte, M.; Goicoa, T.; Genton, M. Interpolation of daily rainfall using spatiotemporal models and clustering. *Int. J. Climatol.* **2015**, *35*, 1453–1464.
40. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B B (Methodol.)* **1977**, *39*, 1–38.
41. Moran, P.A. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23.
42. Brown, P.E. Model-Based geostatistics the easy way. *J. Stat. Softw.* **2015**, *63*, 1–24.
43. AEMET. Agencia Estatal de Meteorología. Available online: <http://www.aemet.es/es/portada> (accessed on 10 January 2017).
44. Holben, B.N. Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.* **1986**, *7*, 1417–1434.
45. Fensholt, R.; Rasmussen, K.; Nielsen, T.T.; Mbow, C. Evaluation of earth observation based long term vegetation trends—Intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data. *Remote Sens. Environ.* **2009**, *113*, 1886–1898.
46. Fontana, F.M.; Trishchenko, A.P.; Khlopenkov, K.V.; Luo, Y.; Wunderle, S. Impact of orthorectification and spatial sampling on maximum NDVI composite data in mountain regions. *Remote Sens. Environ.* **2009**, *113*, 2701–2712.
47. Wang, J.; Price, K.; Rich, P. Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains. *Int. J. Remote Sens.* **2001**, *22*, 3827–3844.

48. Ichii, K.; Kawabata, A.; Yamaguchi, Y. Global correlation analysis for NDVI and climatic variables and NDVI trends: 1982–1990. *Int. J. Remote Sens.* **2002**, *23*, 3873–3878.
49. Schultz, P.; Halpert, M. Global correlation of temperature, NDVI and precipitation. *Adv. Space Res.* **1993**, *13*, 277–280.
50. Potter, C.; Brooks, V. Global analysis of empirical relations between annual climate and seasonality of NDVI. *Int. J. Remote Sens.* **1998**, *19*, 2921–2948.
51. Julien, Y.; Sobrino, J.A.; Mattar, C.; Ruescas, A.B.; Jimenez-Munoz, J.C.; Soria, G.; Hidalgo, V.; Atitar, M.; Franch, B.; Cuenca, J. Temporal analysis of normalized difference vegetation index (NDVI) and land surface temperature (LST) parameters to detect changes in the Iberian land cover between 1981 and 2001. *Int. J. Remote Sens.* **2011**, *32*, 2057–2068.



© 2017 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/>).