



## Modeling and forecasting MODIS-based Fire Potential Index on a pixel basis using time series models

Margarita Huesca<sup>a</sup>, Javier Litago<sup>b,\*</sup>, Silvia Merino-de-Miguel<sup>c</sup>, Victor Cicuendez-López-Ocaña<sup>a</sup>, Alicia Palacios-Orueta<sup>a</sup>

<sup>a</sup> Departamento de Silvopascicultura, ETSIM, Universidad Politécnica de Madrid, Spain

<sup>b</sup> Departamento de Estadística, ETSIA, Universidad Politécnica de Madrid, Spain

<sup>c</sup> Departamento de Topografía, EUITF, Universidad Politécnica de Madrid, Spain



### ARTICLE INFO

#### Article history:

Received 10 January 2013

Accepted 9 September 2013

#### Keywords:

Time series analysis

MODIS

Autoregressive models

### ABSTRACT

The aim of this research was to model and forecast MODIS-based Fire Potential Index (FPI), implemented with Normalized Difference Water Index (NDWI), as a proxy of forest fire risk, in Navarre (Spain) on a pixel basis using time series models with a forecasting horizon of one year.

We forecast FPI<sub>NDWI</sub> for 2009 based on time series from 2001 to 2008. In the modeling process, the Box and Jenkins methodology was applied in two consecutive stages. First, several generic models based on average FPI<sub>NDWI</sub> time series from different "fuel type-ecoregion" combinations were developed. In a second stage, the generic models were implemented at the pixel level for the entire study region. The usefulness of the proposed autoregressive (AR) model, using the original data and introducing significant seasonal AR parameters, was demonstrated.

Results show that 93.18% of the estimated models (EMs) are highly accurate and present good forecasting ability, precisely reproducing the original FPI<sub>NDWI</sub> dynamics. Best results were found in the Mediterranean areas dominated by grasslands; slightly lower accuracies were found in the temperate and alpine regions, and especially in the transition areas between them and the Mediterranean region.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

The need for understanding and monitoring long-term vegetation dynamics over large areas is widely accepted by the scientific community. In order to carry out these studies, long time series of spatially continuous and frequent data are essential, as well as efficient tools and advanced techniques for analyzing this information. In this sense, the use of satellite data time series can be considered a sound alternative since satellite observations are objective, comparable, and spatially comprehensive at global to local scales. Remote sensing data are a source of frequently

collected information that makes it possible to obtain updated data and to extract patterns to model vegetation response to either meteorological dynamics or extreme events (e.g. forest fires). Up to now, sensors such as the advanced very high resolution radiometer (AVHRR), VEGETATION or the moderate resolution imaging spectroradiometer (MODIS) have provided data with a high temporal frequency at a moderate spatial resolution that can be used to monitor ecosystem dynamics at several scales. In the near future, new sensors will be developed, increasing the amount of high temporal and spatial resolution information. For instance, it is expected that the Sentinel missions will provide Europe with an autonomous and operational framework for earth observation (Berger et al., 2012), which will ensure long time series of remote sensing data at different scales and resolutions.

Statistical time series analysis (TSA) in its frequency and temporal domains (Box et al., 1994), offers a set of tools and methodologies to understand, model and forecast a variable based on the quantitative identification of temporal patterns, which are therefore based on the history of the variable itself. These techniques have been used widely in economics (Granger and Newbold, 1977), and less frequently in some other disciplines such as hydrology (Modarres, 2007; Gemitz and Stefanopoulos, 2011) agriculture (Mariño et al., 1993), or forestry (Telesca et al., 2005).

**Abbreviations:** ACF, autocorrelation function; Adj-R<sup>2</sup>, adjusted coefficient of determination; AR, autoregressive; ARIMA, autoregressive integrated moving average; B&J, Box and Jenkins; EMR, estimated model from representative series (significant lags + estimated coefficients); EM, estimated model (significant lags + estimated coefficients); GM, generic model (significant lags); L-B Q, Ljung–Box Q; MODIS, moderate resolution imaging spectroradiometer; PACF, partial autocorrelation functions; SARIMA, multiplicative seasonal autoregressive integrated moving average; TSA, statistical time series analysis; U, Theil inequality coefficient U; U<sup>B</sup>, bias proportion; U<sup>C</sup>, covariance proportion; U<sup>V</sup>, variance proportion.

\* Corresponding author at: Departamento de Estadística, ETSIA, Universidad Politécnica de Madrid, Ciudad Universitaria, 28040 Madrid, Spain.  
Tel.: +34 913365840.

E-mail address: [javier.litago@upm.es](mailto:javier.litago@upm.es) (J. Litago).

Conversely, TSA has hardly been applied to remote sensing data. The application of TSA techniques to spatially continuous time series, such as those in remote sensing data, opens a new perspective in terms of environmental variable monitoring. This is because it can be applied to large areas, taking into account the history of each pixel and the spatial coherence of the scene. However, to implement TSA methodologies at the pixel level is a major challenge due to the complexity of some models and the large amount of data.

In the remote sensing domain, TSA has been applied mainly to specific representative pixels or average series from a selected group of pixels. For example, Liang (2001) discriminated 13 cover types based on 12 years of AVHRR-NDVI and brightness temperature average time series. Beck et al. (2011) studied productivity changes in an Alaskan boreal forest based on 88 time series from GIMMS-NDVI (NOAA) and tree rings. Horion et al. (2013) assessed the relationship between 12 average SPOT-VEGETATION-NDVI time series and meteorological data. Huesca et al. (2009) assessed fire seasonality in different ecoregions by means of autoregressive (AR) models applied to Fire Potential Index (FPI) (Burgan et al., 1998) using NDVI or NDWI as remote sensing data. García et al. (2010) analyzed the relationship between representative AVHRR-NDVI time series and rainfall data in order to evaluate the influence of precipitation disturbances on different types of ecosystems.

The application of TSA to remote sensing images in a spatially explicit manner (i.e. at the pixel level) has been applied mainly to evaluate temporal trends in large areas (Verbesselt et al., 2010a; Fensholt et al., 2012; Bunn and Scott, 2006). It has also been used to detect seasonal changes usually associated with phenological dynamics (Fensholt et al., 2012; Verbesselt et al., 2010b). In addition, TSA has been used to analyze the temporal relationship between different environmental variables by means of cross-correlation analysis (Peng et al., 2010).

Generally, the most typical TSA applications are associated with variable modeling and forecasting (Harvey and Andrew, 1981). The first applications to modeling were based on time series decomposition, a technique in which a series is considered as the sum of three components: tendency, seasonality and randomness (Harvey and Andrew, 1981). In the seventies, Box and Jenkins (1970) introduced the autoregressive integrated moving average (ARIMA) models, an approach in which a time series is modeled as a stationary stochastic process. In the remote sensing domain, some authors have applied ARIMA models (Piwowar and Ledrew, 2002); however, due to the highly significant seasonal component usually associated with remote sensing time series, it is more common to use multiplicative seasonal autoregressive integrated moving average (SARIMA) models instead (Xiao et al., 2011; Fernández-Manso et al., 2011; Jiang et al., 2010).

The Box-Jenkins methodology is typically applied in five standard stages (identification, estimation, validation, forecasting, and evaluation). The identification stage involves a 'labor-intensive' design, which makes it difficult to apply it to a large number of pixels (i.e. time series) from spatially continuous remote sensing data. Several approaches have been used to deal with this issue. Xiao et al. (2011) forecasted Leaf Area Index (LAI) using the same estimated SARIMA model (i.e. same structure and coefficients) for all pixels of the MODIS image. They obtained low-accuracy results, possibly due to the lack of model specificity at the pixel level. Han et al. (2010) applied a generic AR (1) model at the pixel level to forecast the Vegetation Temperature Condition Index (VTCI); in this case, the model coefficients were specific for each time series, which resulted in a highly accurate forecast. Jiang et al. (2010) also found accurate predictions using the same approach

with one generic SARIMA model to predict MODIS LAI time series. Fernández-Manso et al. (2011) improved forecast accuracy by applying two different generic SARIMA models to predict conifer forest NDVI-AVHRR time series with a forecast horizon of 10 days.

TSA has a direct and relevant application within the scope of forest fire assessment. According to Pausas (2004), current climate trends in the Iberian Peninsula indicate an increase in annual and summer temperatures, and a slight decrease in summer rainfall, which will amplify the risk of forest fires (IPCC, 2007). Under this scenario, the approach of statistical time series modeling and forecasting with RS data would contribute to a better understanding of the fire risk dynamics and the development of effective early warning methods.

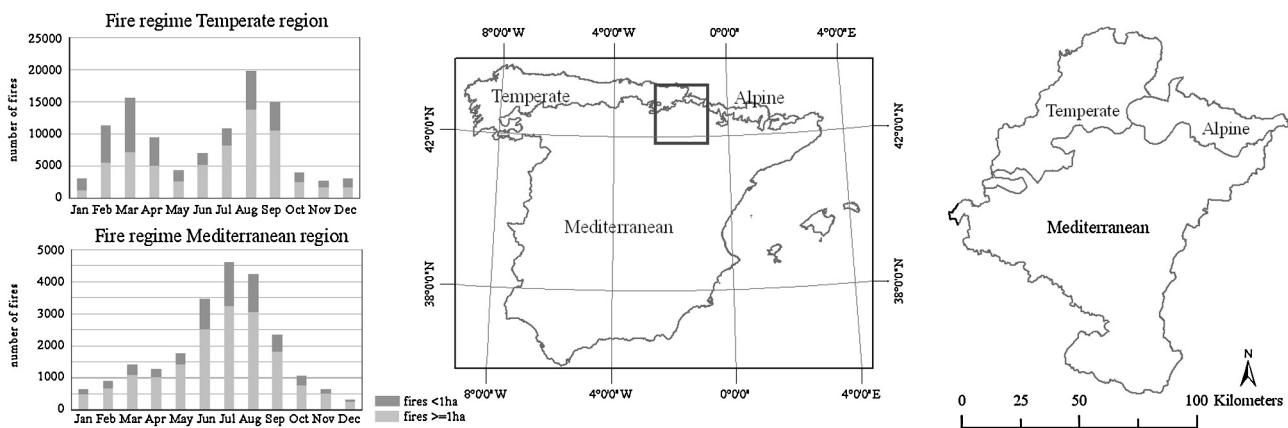
The FPI (Burgan et al., 1998; Sebastián-López et al., 2002; Schneider, 2008) is a dynamic forest fire risk index that is highly specific for fuel type, weather conditions and vegetation status, resulting in values with high spatial variability. This requires the development of pixel-specific models to account for specific environmental characteristics. In addition, forest fire risk estimation models should be able to predict fire risk with a short and medium-term forecast horizon to obtain this information early enough to define fire-prevention plans.

The objective of the present study is to model and forecast MODIS-based FPI<sub>NDWI</sub>, as a proxy of forest fire risk, using pixel-specific autoregressive models based on a large number of generic models and with a forecast horizon of one year.

## 2. Study area

Navarre, which is located in the northern Iberian Peninsula (Fig. 1), occupies an area of 10,420 km<sup>2</sup> and is located on the edge of the temperate, alpine, and Mediterranean ecoregions. The presence of significant climate gradients results in distinct vegetation types and highly variable fire regimes.

The temperate ecoregion is characterized by a warm, temperate maritime climate that is strongly influenced by the Cantabrian Sea, with frequent rain, fog and drizzle. This region is dominated by deciduous broad-leaved forests, mainly *Quercus robur* L., *Quercus petraea* (Matts.) Liebl. and *Fagus sylvatica* L. Fire regimes are characterized by their high frequency and relatively small burned area (Vélez, 2000). The northern part of the alpine ecoregion is characterized by a moist continental climate and is mainly occupied by coniferous and broad-leaved forests. On the other hand, there is a clear influence of the Mediterranean climate in southern part, resulting in a drier continental climate that is dominated by more xerophytic species. Fire regimes are characterized by low frequency and marked seasonal and annual variability resulting in an irregular pattern (Vélez, 2000). Finally, the Mediterranean ecoregion is dominated by a Mediterranean climate, with the Western area clearly influenced by the temperate climate and the Eastern area by the continental characteristics. Forests in this region are characterized by Mediterranean sclerophyllous species such as *Quercus ilex* L. and *Quercus coccifera* L., among others. Fire regimes are intermediate in frequency and may affect medium to large areas. In terms of forest fire patterns there is a clear difference between temperate and Mediterranean regions. Forest fires in the temperate region follows a bi-modal pattern with relative maxima in spring and late summer-fall while forest fires in the Mediterranean region follows a unimodal pattern with an absolute maximum in summer (Vélez, 2000). Fig. 1 shows monthly distribution of forest fires occurrence during the period 2001–2010, for the Spanish temperate and Mediterranean ecoregions (<http://www.magrama.es/>).



**Fig. 1.** Monthly distribution of forest fires occurrence during the period 2001–2010, for the Spanish temperate and Mediterranean ecoregions location of Navarre within the Iberian Peninsula.

### 3. Materials

#### 3.1. Remote sensing data

The remote sensing data used consisted of a set of 454 eight-day composites (i.e. from February 2000 until December 2009) from the MODIS surface reflectance product MOD09A1 (DAAC, <https://lpdaac.usgs.gov/>). This product is made up of seven surface reflectance bands located in the VIS, NIR and SWIR regions (bands are centered on the following wavelengths: 470 (blue), 555 (green), 648 (red), 858 (NIR), 1240 (SWIR1), 1640 (SWIR2) and 2130 nm (SWIR3)) at a spatial resolution of 500 m. The products were downloaded from the NASA Distributed Active Archive Centre.

#### 3.2. Meteorological data

Meteorological information was gathered from the weather stations in the study region and surrounding provinces, in order to ensure good-quality information on the edges of the study area. Daily records from the Navarre weather stations (<http://www.Navarre.es/>) for the period 2000–2009 were provided by the Department of Rural Development and Environment (Government of Navarre). Daily records from 17 weather stations in the surrounding provinces were provided by the Spanish National Meteorological Agency (AEMET, <http://www.aemet.es/>). Since the year 2000, the number of weather stations in Navarre has increased from 53 to 81 in the year 2009, so the maximum number of available stations for each year was used in order to obtain the most accurate results possible.

#### 3.3. Ancillary data

A Digital Elevation Model (DEM) at 1:25,000 scale (10 m pixel size) was provided by the Spanish National Geographic Institute (<http://www.ign.es>). An ecoregion map was provided by the Government of Navarre. The fuel type map used (<http://www.magrama.es/>) at 1:50,000 scale (Fig. 2) was developed by the former ICONA (Spanish National Institute for Nature Conservation) using the 13 fuel types established by the NFDRS (National Fire Danger Rating System), and adapted to Spain. The nine fuel types present in the study region are grouped based on the fire spread element: (1) grasses (i.e. fuel types 1–3), (2) shrubs (i.e. fuel types 4–7), and (3) litter (i.e. fuel types 8–10). Each fuel type is characterized by a specific fuel extinction moisture value, which represents the degree of moisture after which the fuel combustion itself ceases.

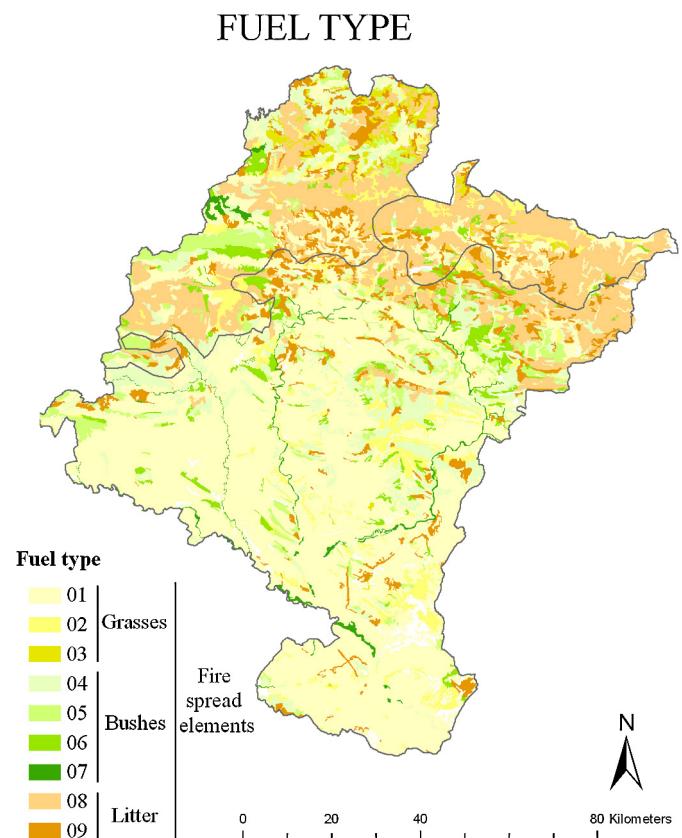
### 4. Methods

#### 4.1. Remote sensing data

MODIS image products were downloaded and projected into UTM Zone 30N WGS-84 coordinate system, by means of the MODIS Reprojection Tool (MRT) software. Time series of the NDWI ([Gao, 1996](#)) (Eq. (1)) were calculated using the following equation:

$$\text{NDWI} = \frac{\rho_{\text{nir}} - \rho_{\text{swir}}}{\rho_{\text{nir}} + \rho_{\text{swir}}} \quad (1)$$

where,  $\rho_{\text{nir}}$  and  $\rho_{\text{swir}}$  represent pixel reflectance in the near infrared and SWIR1 bands, respectively.



**Fig. 2.** Spatial distribution of fuel types presented in the study area.

In order to avoid using anomalous values, the time series were smoothed by eliminating possible outliers, based on thresholds defined by the mean plus two standard deviations. Eliminated values were replaced by the average of the previous and next observations.

#### 4.2. Meteorological data processing

The meteorological variables used were the maximum temperature ( $T_{\max}$ ) and minimum relative humidity ( $H_{\min}$ ). Daily meteorological data were summarized to MODIS frequency, and 8-day mean values were calculated. Maximum temperature was spatially interpolated by inverse distance weighting (IDW). The temperature was previously corrected to sea level using an environmental lapse rate value of 5.5 °C/km. Minimum relative humidity was estimated by a multiple linear regression equation (Eq. (2)) in which  $H_{\min}$  is estimated based on  $T_{\max}$  and elevation (Huesca et al., 2009). The regression model used (Eq. (2)) was validated with independent meteorological observations from the same study region.

$$H_{\min} = 72.1761 - 1.4181T_{\max} + 0.0049H \quad (2)$$

where  $H$  is the elevation obtained from the DEM of Navarre.

#### 4.3. Modified Fire Potential Index computation

The use of FPI<sub>NDWI</sub> (i.e. computed with the NDWI) is based on previous works. Huesca et al. (2009) showed that this index better explained forest fire risk patterns in the temperate region of Navarre, while showing similar results in the Mediterranean area because NDWI is more sensitive to vegetation water content, and Huesca et al. (2010) showed higher forecast ability of FPI<sub>NDWI</sub>.

A general formulation of the FPI<sub>NDWI</sub> can be written as (Eq. (3)):

$$\text{FPI}_{\text{NDWI}} = 100 \times (1 - \text{FMC}) \times (1 - \text{VC}) \quad (3)$$

where FPI<sub>NDWI</sub> is the Fire Potential Index based on NDWI, FMC represents the ratio between the ten-hour-time lag fine fuel moisture (Fosberg and Deeming, 1971) and the extinction moisture, which depends on meteorological data (i.e.  $T_{\max}$  and  $H_{\min}$ ) and fuel type, and VC, calculated from NDWI, is an indicator of live vegetation amount. FPI<sub>NDWI</sub> takes a maximum value of 100 when the risk is very high. A detailed explanation of this index can be found in Huesca et al. (2009).

#### 4.4. FPI<sub>NDWI</sub> time series modeling by means of statistical time series analysis (TSA)

Time series for period 2000–2008 were used for modeling, while data from 2009 were used to validate the forecasting accuracy of the estimated models.

##### 4.4.1. Identification of random time series

The first step was to identify the non-random pixel time series in the image by means of the autocorrelation function (ACF) and the Ljung–Box Q (L–B Q) statistics (Ljung and Box, 1978). The ACF plots allowed the analysis of the FPI<sub>NDWI</sub> dynamics for each pixel of the image. The estimated Q statistic identified pixels where the FPI<sub>NDWI</sub> dynamics were either completely random, seasonal, containing a trend, or even the result of the interaction of several of these components.

##### 4.4.2. Box and Jenkins methodology

The FPI<sub>NDWI</sub> time series analyzed integrate variables related to vegetation status and meteorological conditions that create a significant seasonality in these series, the order of which,  $s = 46$ , is the result of the MODIS acquisition data frequency (i.e. 8 days). This seasonality is a source of non-stationarity in the FPI series.

The Box and Jenkins (B&J) methodology (Box et al., 1994) and some of its subsequent developments (Hamilton, 1994) were used to analyze and to model FPI<sub>NDWI</sub> time series. However, we did not use the regular (1-B) and seasonal (1-B)s first differences transformation proposed in the first stage of the Box and Jenkins general strategy to achieve stationarity in seasonal time series (Box et al., 1994, pp. 353). This transformation assumes the presence of  $s+1$  (i.e. 47) unit roots in all the series, two of them at the zero frequency and  $s-1$  at the seasonal frequencies (Box et al., 1994, pp. 330). However, we cannot verify the presence of seasonal unit roots in series with seasonal order  $s=46$ , because, to the best of our knowledge, it does not exist the required test, equivalent to those proposed by Helleberg et al. (1990), Franses (1991), Beaulieu and Miron (1993), and Darné et al. (2002) for quarterly, monthly, and daily data ( $s=4, 12, 7$ ) respectively.

In addition, the use of double transformation (1-B)(1-B)s may induce overdifferentiation in the series if they do not have  $s+1$  unit roots (see Ghysels and Osborn, 2001, for example). Plosser and Schwert (1977, 1978) indicated that 'seasonal differencing the series ... also induces a non-invertible moving-average process at the seasonal lag'. A more detailed explanation of overdifferencing and its consequences can be found in Osborn (1990), Maddala and Kim (1998), among others.

For all of these reasons, in this study, the B&J methodology is applied to the original FPI<sub>NDWI</sub> time series. The B&J methodology was carried out in the following standard stages:

1. *Identification*: In this stage the dynamic structure of the FPI<sub>NDWI</sub> time series was analyzed and the information obtained was used to specify the dynamic models proposed for each series. Its aim is to identify the characteristic components, trend, cycle, seasonality, irregular or others (structural changes, etc.) of each time series. Identification was accomplished in four main stages:
  - 1.1. Estimation of the regular and partial autocorrelation functions (ACF and PACF).
  - 1.2. Detection of stochastic trends by means of unit roots tests (Dickey and Fuller, 1979).
  - 1.3. Detection of significant periodic components by means of periodogram estimation (Hamilton, 1994) and white noise tests (Fisher's Kappa (Fk) and Barlett's Kolgomoroff-Smirnoff (BKS) tests) (Fuller, 1976). The Barsky and Miron (1989) approach was adopted to explore the deterministic and stochastic nature of the FPI<sub>NDWI</sub> seasonality by estimating a deterministic regression model of each FPI<sub>NDWI</sub> over 45 seasonal dummy variables. The adjusted  $R^2$  (Adj- $R^2$ ) of the regression 'measures the percentage of the variation in the dependent variable due to deterministic seasonality' (Barsky and Miron, 1989).
  - 1.4. Specification of the regular and seasonal autoregressive and moving-average parameters that suitably represent the whole dynamics of each series on the basis of the above preceding identification results. The seasonal nature of the FPI<sub>NDWI</sub> time series was incorporated into the models by including seasonal AR parameters.
2. *Estimation*: The models specified in the previous stage were estimated by nonlinear least squares methods. The Standard AIC and SBC selection criteria (Hamilton, 1994) were applied to select the most suitable model. The individual and joint significance of the model parameters was assessed by means of the Student- $t$  and  $F$  tests.
3. *Validation*: The adequacy of the estimated models was evaluated by means of the L–B Q statistics (Ljung and Box, 1978) that test for autocorrelation in the model residuals. If the test indicates that a significant amount of autocorrelation remains in the estimated

residuals, the model is not valid and it becomes necessary to return to the Identification stage.

**4. Forecasting:** The validated models were used to forecast the historic and future FPI<sub>NDWI</sub> values.

**5. Evaluation of forecast:** The simulation accuracy of the selected models was assessed via ex post static simulation of each FPI<sub>NDWI</sub>, using the Theil inequality coefficient  $U$  (Theil, 1971). A more accurate assessment may be obtained from the analysis of coefficient  $U$  proportions, bias ( $U^B$ ), variance ( $U^V$ ), and covariance ( $U^C$ ).

#### 4.4.3. Methodology implementation

The preceding methodology was applied consecutively at two levels: (1) the ‘representative time series’ of Navarre, and (2) all pixel time series presented in the MODIS image.

**4.4.3.1. Modeling of ‘representative time series’.** Since climate and fuels are the main drivers for fire risk (Vélez, 2000), this work has been based on the hypothesis that different ecoregions and fuel types are characterized by different forest fire behavior, thus each of them should be represented by its own model.

In the image, there are pixels that fall completely within one fuel type and one ecoregion, those pixels were defined as ‘pure pixels’, and were used to define the ‘representative time series’. The 27 ‘representative series’ (i.e. 9 fuel types  $\times$  3 ecoregions) were defined as the average of the ‘pure pixel’ time series within each fuel type – ecoregion class. However, in the same image there are also ‘mixed pixels’ occupied by different combinations of fuel types and ecoregions. In this research, those pixels were characterized as belonging to the dominant fuel type – ecoregion combination.

The methodology was applied completely to the representative time series in the same order as exposed. As a result of that work, an ‘estimated model’ (EMR) (specific lags and coefficients) for each representative time series was developed. These validated models were grouped in several ‘generic models’ (GMs) that summarized the dynamic ‘common factors’ (i.e. common lags) of all original models.

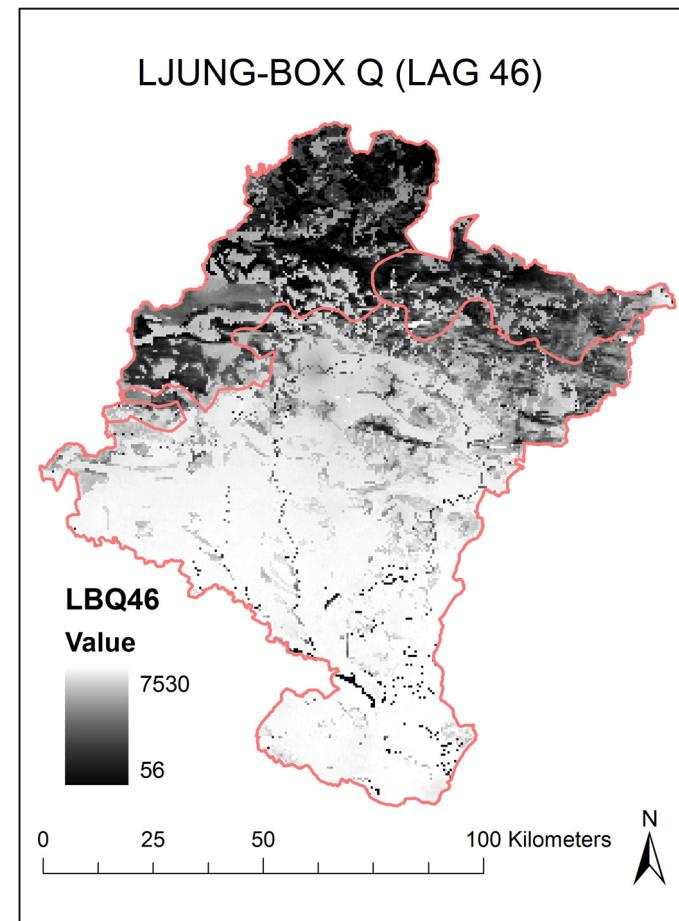
**4.4.3.2. Modeling of spatially continuous MODIS image.** Generic models obtained in the previous steps were implemented at the image level and pixel specific coefficients were estimated. The model validation, FPI<sub>NDWI</sub> forecasting and its evaluation were accomplished in the standard way. Thus, 2009 FPI<sub>NDWI</sub> values have been predicted and evaluated at the pixel level for the whole region.

Statistical analyses were done using SAS9.2 and image processing was accomplished with ENVI 4.2 and IDL language.

## 5. Results

Results of the L–B Q statistic (i.e. white noise test) (Fig. 3) show that all except two pixels present higher values than their 1 and 5% critical values, rejecting the null hypothesis that the time series were white noise. The number of non-random pixels found and analyzed was 40,903.

Fig. 4a shows the FPI<sub>NDWI</sub> time series from three representative zones. Zone (1) corresponds to the Mediterranean region where the fire spread element is mainly grass. Zone (2) is located in the alpine region where the fire is spread primarily through litter. Zone (3) is located in the temperate region in a forest where the fire is spread primarily through shrubs. The FPI<sub>NDWI</sub> in the Mediterranean region follows a clear uni-modal pattern with very low (even negative) values during the non-fire season, and high values during summer. The FPI<sub>NDWI</sub> associated with shrubs in the temperate region also follows a clear annual pattern but presents higher values over the course of the year, as well as a broader cycle with the maximum occurring later than in the Mediterranean ecoregion. On the other hand, the



**Fig. 3.** Spatial distribution of Ljung-box Q values for one year (lag 46). Gray scale, black and white means minimum and maximum Q value respectively.

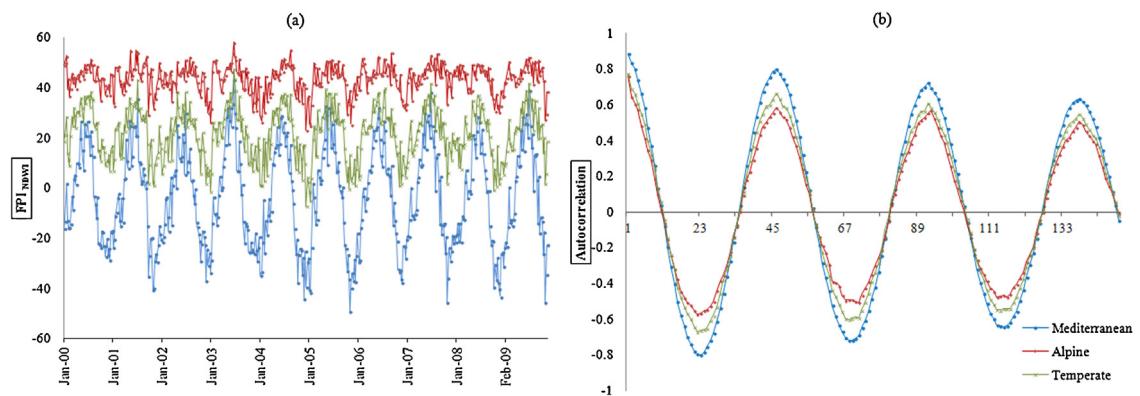
behavior of the FPI<sub>NDWI</sub> in the alpine region associated with litter follows a pattern with two relative maxima, in spring and summer-fall. Their ACF for the first 150 lags (slightly more than 3 years) is shown in Fig. 4b. They present high and positive autocorrelation at lags 1 ( $\approx 8$  days) and s 46 ( $\approx 1$  year), and negative autocorrelation at lag 23 ( $\approx 6$  months).

Fig. 5a and b shows the mean and variance of the FPI<sub>NDWI</sub> time series in the study area. The fuel types whose fire spread element is either litter (i.e. fuel types 8 and 9) or shrubs (i.e. fuel types 4–7) are characterized by a higher FPI<sub>NDWI</sub> mean than those propagated by the grass (i.e. specifically fuel types 1 and 2). Fuel types associated with litter as fire spread element have a low variance while fuel types 1–2 are characterized by high variance. The highly mnegative values during the non-fire season reduce the mean values of FPI<sub>NDWI</sub> time series in the Mediterranean grassland.

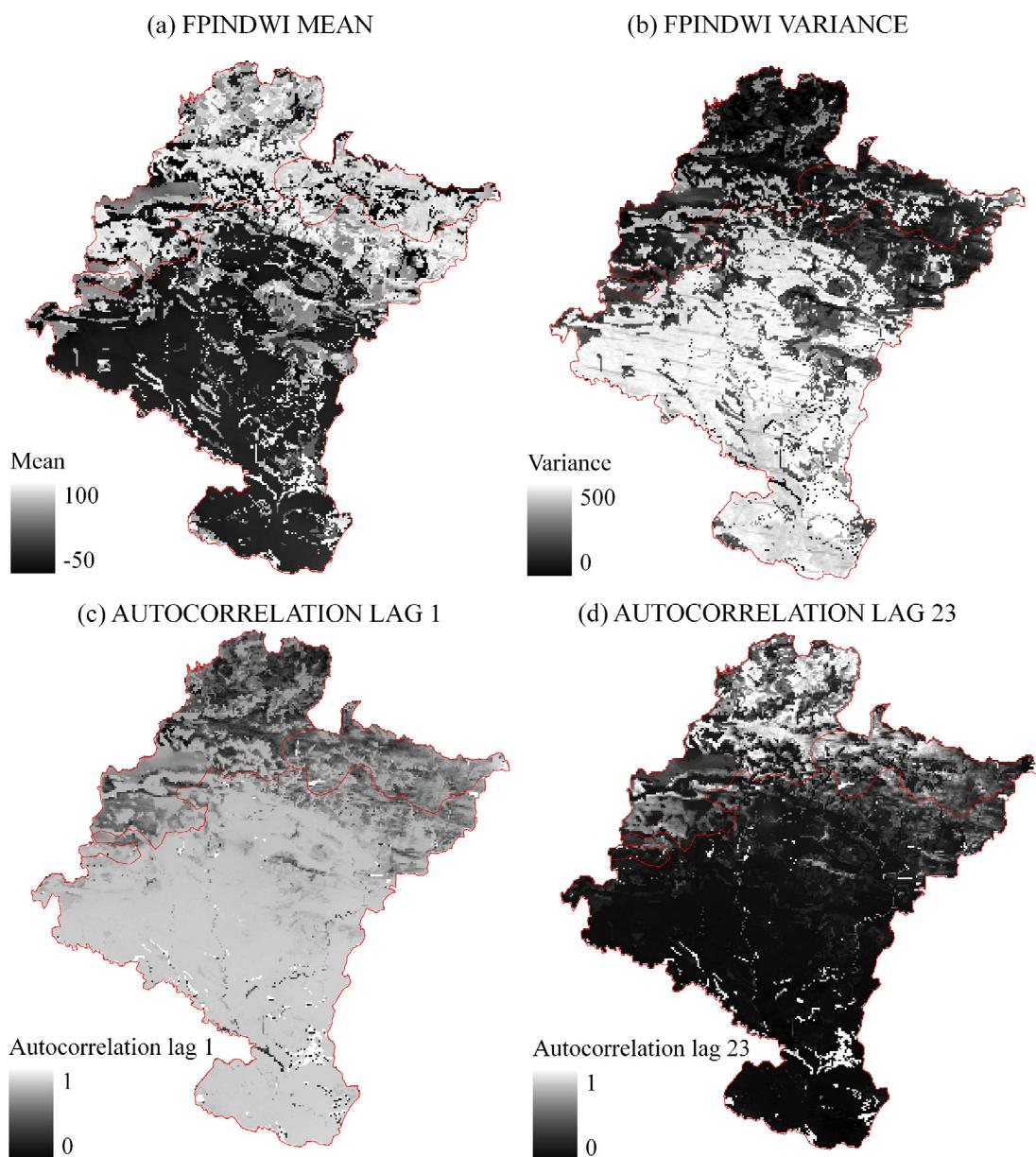
Autocorrelation values at lags 1 ( $\approx 8$  days) and 23 ( $\approx 6$  months) for each pixel are shown in Fig. 5c and d. A clear north–south gradient of FPI<sub>NDWI</sub> autocorrelation values can be observed. The autocorrelation values at lag 46 ( $\approx 1$  year, not shown) presented similar spatial distribution to those at lag 1 with slightly lower values. The southern part of the study area presented higher positive autocorrelation values at lags 1 and 46, and lower negative values at lag 23 than the northern part.

#### 5.1. Modeling of ‘representative time series’

Twenty six ‘representative time series’ corresponding to 26 combinations of “fuel type–ecoregion” present in the study area were analyzed. Fuel type 7 in the alpine region was not considered



**Fig. 4.** (a) Temporal evolution of FPI<sub>NDWI</sub> from February 2000 to December 2009 and (b) autocorrelation values for the first 150 lags for the Mediterranean region where the fire spread element is grass (blue), alpine region where the fire spread element is litter (red), temperate region where the fire spread element are brushes (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



**Fig. 5.** Spatial distribution of mean FPI<sub>NDWI</sub> (a), variance FPI<sub>NDWI</sub> (b), and autocorrelation values at lag 1 (c) and 23 (d).

**Table 1**

Adjusted R-square,  $\bar{R}^2$ , for the estimated regression for  $FPI_{NDWI}$  and for their first differences,  $\nabla FPI_{NDWI}$ .

	Adjusted $R^2$	
	$FPI_{NDWI}$	$DFPI_{NDWI}$
FPINDWI_P1	0.666	0.023
FPINDWI_P2	0.645	0.029
FPINDWI_P3	0.476	-0.002
FPINDWI_P4	0.640	0.011
FPINDWI_P5	0.553	0.000
FPINDWI_P6	0.521	0.015
FPINDWI_P8	0.509	0.005
FPINDWI_P9	0.526	0.001
FPINDWI_T1	0.675	0.033
FPINDWI_T2	0.650	0.026
FPINDWI_T3	0.366	-0.008
FPINDWI_T4	0.590	0.009
FPINDWI_T5	0.585	0.011
FPINDWI_T6	0.490	-0.004
FPINDWI_T7	0.066	-0.014
FPINDWI_T8	0.373	-0.003
FPINDWI_T9	0.412	-0.009
FPINDWI_M1	0.724	0.079
FPINDWI_M2	0.718	0.065
FPINDWI_M3	0.580	0.009
FPINDWI_M4	0.695	0.044
FPINDWI_M5	0.679	0.033
FPINDWI_M6	0.677	0.032
FPINDWI_M7	0.290	-0.014
FPINDWI_M8	0.584	0.005
FPINDWI_M9	0.655	0.026

due to the lack of pure pixels to build the ‘representative time series’. Since this combination is not dominant in any pixel of the study area, modeling was not necessary.

The Dickey–Fuller tests carried out to investigate the stationarity of the representative series rejected the presence of a stochastic trend for all of them. Table 1 shows the Adj- $R^2$  estimates for the deterministic regression for original  $FPI_{NDWI}$ , and for their first differences  $\nabla FPI_{NDWI}$  (1st and 2nd columns, respectively). The estimates ranged from 0.473 to 0.666 in the alpine region, from 0.066 to 0.675 in the temperate region, and from 0.290 to 0.724 in the Mediterranean region. Moreover, the Student- $t$  statistic results indicate that most of the coefficients in the deterministic model were significant. These results indicate a moderate degree of adjustment of the deterministic model to the  $FPI_{NDWI}$  series, evincing a significant deterministic component of the  $FPI_{NDWI}$  annual cycle. However, the estimation of the deterministic regression for the  $FPI_{NDWI}$  first differences yielded very low values of the Adj- $R^2$  in the three ecoregions, with maximums of 0.029, 0.033 and 0.079. In addition, the Student- $t$  statistic results indicate that most of the coefficients in the deterministic model were not significant. These results are evidence of a poor fit of the model, and indicate that the variations of  $FPI_{NDWI}$  are not deterministic in nature.

Based on the results of the identification stage it was decided to use an AR general formulation (Eq. (4)) to model  $FPI_{NDWI}$  time series.

$$FPI = \alpha + \sum_{i=1}^s \beta_i FPI_{t-i} + \varepsilon_t \quad (4)$$

where  $\alpha$  is the intercept term,  $\beta_i$  is the autoregression coefficient for the  $FPI_{NDWI}$  ( $i = 1, 2, \dots, s = 46$ ), and  $\varepsilon_t$  is the uncorrelated residual term.

Seasonal AR parameters at significant lags were introduced into each model to incorporate the seasonal dynamics of the  $FPI_{NDWI}$ . Tables 2–4 show the best EMRs (for each representative time series) and their diagnostic, evaluated by means of the adjusted  $R^2$  and L–B Q statistic for lags 23, 46, and 92. Each table is divided in two

Dependent variable	Intercept	Lags	Model diagnostic													
			$t-1$	$t-3$	$t-4$	$t-10$	$t-21$	$t-22$	$t-23$	$t-34$	$t-42$	$t-45$	$t-46$	$t-55$	$Q_{LB(23)}$	$Q_{LB(46)}$
FPINDWI_A1	-14.38 (-5.04)	0.30 (6.60)	0.18 (4.11)	-0.10 (-2.21)	-0.12 (-2.59)	-0.13 (-2.30)	-0.10 (-2.87)	-0.11 (-2.30)	-0.09 (-1.93)	0.21 (0.30)	-0.07 (-1.64)	0.80	24.78 (0.36)	57.71 (0.12)	125.22 (0.01)	
FPINDWI_A2	-0.59 (-1.47)	0.33 (7.27)	0.16 (3.48)	-0.12 (-3.01)	-0.12 (-2.74)	-0.11 (-2.37)	-0.07 (-1.90)	-0.10 (-2.37)	0.10 (-1.90)	0.20 (2.15)	0.20 (4.08)	0.76	19.69 (0.66)	49.77 (0.33)	108.18 (0.12)	
FPINDWI_A3	13.31 (4.48)	0.43 (9.75)	0.15 (3.47)	-0.08 (-2.32)	-0.08 (-2.32)	-0.13 (-3.06)	-0.13 (-3.06)	-0.11 (-3.06)	-0.11 (-1.88)	0.24 (5.64)	0.58 (0.66)	0.78	19.71 (0.63)	42.35 (0.13)	107.31 (0.06)	
FPINDWI_A4	16.87 (5.23)	0.40 (8.97)	0.16 (3.54)	-0.13 (3.57)	-0.12 (-2.81)	-0.11 (-2.49)	-0.11 (-2.49)	-0.11 (-2.49)	-0.07 (-1.88)	0.23 (5.29)	0.78 (0.36)	0.78	24.84 (0.27)	51.41 (0.27)	113.56 (0.06)	
FPINDWI_A5	17.36 (5.37)	0.36 (8.01)	0.14 (3.15)	-0.11 (-2.87)	-0.13 (-3.00)	-0.12 (-2.54)	-0.12 (-2.54)	-0.12 (-2.54)	-0.08 (-1.88)	0.24 (5.24)	0.67 (0.51)	0.67	22.26 (0.31)	50.22 (0.31)	114.22 (0.06)	
FPINDWI_A6	10.24 (3.25)	0.42 (9.20)	0.13 (2.56)	0.10 (1.97)	-0.07 (-1.85)	-0.12 (-2.86)	-0.12 (-2.86)	-0.12 (-2.86)	0.10 (2.28)	0.16 (3.56)	0.65 (0.91)	0.65	14.63 (0.90)	34.28 (0.91)	103.97 (0.19)	
FPINDWI_A7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FPINDWI_A8	9.12 (2.59)	0.46 (10.47)	0.11 (2.80)	-	-	-	-	-	-	-	-	-	0.65	16.45 (0.84)	31.59 (0.95)	91.48 (0.50)
FPINDWI_A9	8.61 (2.78)	0.40 (9.10)	0.40 (2.68)	-	-	-	-	-	-	0.14 (3.41)	0.21 (3.57)	0.65	18.96 (0.78)	38.33 (0.78)	107.32 (0.13)	

**Table 2**  
The best estimated model and the model diagnostic for each fuel type in the alpine region.

**Table 3**

The best estimated model and the model diagnostic for each fuel type in the temperate region.

Dependent variable	Intercept	Lags											Model diagnostic				
			$t-1$	$t-2$	$t-3$	$t-10$	$t-16$	$t-22$	$t-25$	$t-34$	$t-41$	$t-42$	$t-46$	$Q_{LB(23)}$	$Q_{LB(46)}$	$Q_{LB(92)}$	
FPI <sub>NDWI</sub> T1	-4.96 (-5.86)	0.32 (7.32)	0.21 (5.15)		-0.12 (-4.15)	-0.15 (-3.53)							0.25 (5.83)	0.81 (0.27)	26.67 (0.49)	45.53 (0.17)	103.94
FPI <sub>NDWI</sub> T2	-0.32 (-0.89)	0.31 (7.21)	0.19 (4.73)		-0.12 (-3.91)	-0.16 (-3.65)							0.27 (6.28)	0.79 (0.32)	25.64 (0.52)	44.81 (0.27)	99.87
FPI <sub>NDWI</sub> T3	18.50 (5.40)	0.35 (7.88)	0.15 (3.37)		-0.11 (-2.75)	-0.11 (-2.64)							0.24 (5.44)	0.48 (0.57)	21.15 (0.48)	45.89 (0.19)	103.56
FPI <sub>NDWI</sub> T4	12.44 (5.57)	0.35 (8.11)	0.17 (4.26)		-0.12 (-3.64)	-0.12 (-2.91)							0.27 (6.20)	0.72 (0.51)	22.11 (0.60)	42.92 (0.24)	101.19
FPI <sub>NDWI</sub> T5	10.02 (5.48)	0.36 (8.33)	0.18 (4.44)		-0.12 (-3.81)	-0.11 (-2.73)							0.25 (5.84)	0.71 (0.67)	19.51 (0.81)	37.36 (0.54)	89.97
FPI <sub>NDWI</sub> T6	14.38 (3.53)	0.37 (8.30)	0.16 (3.76)		-0.10 (-2.56)		-0.12 (-2.73)		0.10 (2.26)		0.19 (4.41)		0.61 (0.88)	15.36 (0.87)	35.51 (0.63)	85.80	
FPI <sub>NDWI</sub> T7	31.38 (5.12)	0.38 (8.41)	0.10 (2.25)	-0.10 (-2.26)				-0.12 (-2.65)		0.08 (1.85)	0.09 (2.00)		0.25 (0.97)	12.27 (0.99)	26.01 (0.17)	104.76	
FPI <sub>NDWI</sub> T8	19.94 (3.73)	0.41 (9.43)	0.11 (2.51)		-0.11 (-2.62)	-0.06 (-1.58)		-0.10 (-2.56)		0.13 (3.03)	0.18 (3.99)		0.53 (0.83)	16.60 (0.96)	30.36 (0.46)	92.72	
FPI <sub>NDWI</sub> T9	21.77 (4.50)	0.34 (7.70)	0.11 (2.56)		-0.12 (-2.94)	-0.10 (-2.44)		-0.13 (-3.10)		0.14 (3.31)	0.20 (4.52)		0.55 (0.77)	17.85 (0.91)	33.86 (0.42)	94.17	

**Table 4**

The best estimated model and the model diagnostic for each fuel type in the Mediterranean region.

Dependent variable	Intercept	Lags											Model diagnostic				
			$t-1$	$t-2$	$t-3$	$t-4$	$t-10$	$t-21$	$t-22$	$t-23$	$t-42$	$t-45$	Adj-R <sup>2</sup>	$Q_{LB(23)}$	$Q_{LB(46)}$	$Q_{LB(92)}$	
FPI <sub>NDWI</sub> M1	-0.95 (-2.41)	0.33 (6.77)	0.09 (1.72)	0.18 (3.88)		-0.12 (-4.80)	-0.14 (-3.89)						0.30 (6.98)	0.87 (0.13)	30.93 (0.11)	58.28 (0.04)	117.80
FPI <sub>NDWI</sub> M2	3.73 (4.23)	0.35 (8.09)		0.18 (4.23)	-0.07 (-2.66)	-0.13 (-3.66)							0.11 (2.34)	0.26 (5.33)	27.52 (0.24)	57.44 (0.12)	115.66
FPI <sub>NDWI</sub> M3	14.60 (5.00)	0.39 (8.16)	0.12 (2.33)	0.16 (3.30)		-0.11 (-3.73)	-0.12 (-3.37)						0.24 (5.73)	0.78 (0.88)	15.37 (0.70)	40.63 (0.85)	78.10
FPI <sub>NDWI</sub> M4	6.05 (2.49)	0.32 (7.01)		0.17 (3.56)	0.11 (2.18)	-0.09 (-2.74)		-0.09 (-2.30)		0.12 (-2.90)		0.26 (5.69)	0.84 (0.38)	24.50 (0.40)	47.75 (0.04)	117.08	
FPI <sub>NDWI</sub> M5	5.81 (2.44)	0.32 (7.07)		0.17 (3.65)	0.11 (2.36)	-0.10 (-3.02)		-0.08 (2.15)		0.12 (2.87)		0.25 (5.66)	0.83 (0.50)	22.37 (0.50)	45.42 (0.50)	111.63	
FPI <sub>NDWI</sub> M6	8.74 (2.42)	0.35 (7.72)		0.16 (3.44)	0.10 (2.06)	-0.10 (-3.17)			-0.09 (-2.26)	0.13 (3.07)		0.24 (5.37)	0.83 (0.61)	20.60 (0.56)	43.99 (0.08)	111.44	
FPI <sub>NDWI</sub> M7	36.96 (6.45)	0.47 (10.41)		0.10 (2.39)		-0.17 (-3.91)		-0.11 (-2.39)					0.11 (2.54)	0.55 (0.92)	14.13 (0.88)	34.95 (0.68)	85.12
FPI <sub>NDWI</sub> M8	21.53 (5.50)	0.42 (9.53)		0.10 (2.43)		-0.14 (-3.55)		-0.09 (-2.02)					0.24 (5.63)	0.72 (0.67)	19.52 (0.53)	44.68 (0.12)	108.13
FPI <sub>NDWI</sub> M9	18.76 (5.84)	0.37 (8.48)		0.19 (4.35)	-0.10 (-3.49)	-0.10 (-2.46)		-0.10 (-2.33)					0.27 (6.25)	0.80 (0.65)	19.80 (0.34)	49.38 (0.08)	111.22

**Table 5**

Selected generic models, common factors.

Group code	Description	Grouped models	Significative lags
1	Alpine and temperate grassland	A1, A2, T1 and T2	1, 3, 10, 21, 23, 34, 45, 46, 55
2	Alpine and temperate shrubs	A3, A4, A5, A6, T3, T4, T5, T6	1, 3, 10, 21, 23, 34, 46
3	Mediterranean grassland	M1 and M2	1, 3, 10, 21, 45, 46
4	Mediterranean shrubs-grassland	M3	1, 2, 3, 10, 21, 46
5	Mediterranean shrubs	M4, M5 and M6	1, 3, 4, 10, 22, 42, 46
6	Temperate shrubs-woodland	T7	1, 3, 10, 34, 42, 46
7	Mediterranean shrubs-woodland	M7	1, 3, 21, 23, 46
8	Alpine woodland	A8, A9	1, 3, 22, 42, 46
9	Temperate woodland	T8 and T9	1, 3, 16, 22, 34, 42, 46
10	Mediterranean woodland	M8 and M9	1, 3, 10, 21, 23, 46

Ai, alpine region; Ti, temperate region; Mi, Mediterranean region; i, fuel type.

parts, the first one includes the variable (first column), the model intercept (second column), and the significant lags of the dependent variable (rest of columns). The second part of the table (the last four columns) shows the diagnostic of the model. Furthermore, the model of each variable includes two rows of results. The first row contains the estimated intercept and the lag coefficients (first part), and the adjusted  $R^2$  coefficient and L-B Q values for the lags 23, 46 and 92 (second part). The second row shows, in parentheses, Student t statistics for the estimated intercept and coefficients and the probability of the null hypothesis of L-B Q test.

All the coefficients estimated for the 26 EMRs showed a high statistical significance (i.e. the absolute value of the Student  $t > 2$ ). In addition, the L-B Q test statistics for all model residuals took lower values at lags 23, 46 and 92 (half, one, and two years) than its corresponding critical values. Consequently, the test rejected the hypothesis of residual autocorrelation with a high probability (greater than 0.05 for any lag except for the fuel type 1 in the alpine region at lag 92). It can therefore be affirmed that there is absence of autocorrelation in the residuals for all EMRs.

The developed EMRs were grouped in 10 GMs based on the common dynamic factors found in them (i.e. common lags). Table 5 shows the significant parameter lags of these GM.

## 5.2. Modeling of spatially continuous time series from MODIS image

Fig. 6 shows the location of the time series corresponding to the each GM. A great irregularity dominated the distribution of the number of series corresponding to each generic model (GM) in the study area. Thus, while 45% of pixels belong to the GM 3 (i.e. Mediterranean grassland) only 0.08% and 0.23% of them belong to GMs 4 and 6 (i.e. Mediterranean shrub-grassland and temperate shrub-woodland). Pixels associated with GM 7 (i.e. Mediterranean shrub-woodland) follow Mediterranean river basins, thus this model is associated mostly with riparian vegetation. Most of the pixels associated with GM 10 (i.e. Mediterranean woodland) are located in the transition zone between the alpine and Mediterranean ecoregions.

The amount of useful information to model and forecast FPI<sub>NDWI</sub> time series showed a large variability in the whole study area as it can be inferred from the mean values of the Ljung–Box Q-statistic (Table 6). These values fluctuated from a minimum of 735.63 to a maximum of 7079.37, that largely exceed the 5% critical value of a  $\chi^2$  with 46 degrees of freedom (i.e. 62.83). These high Q values indicate the great amount of available temporal dependency information presented in the series. Thus, this indicates, that series corresponding to pixels associated with GM 6 (i.e. temperate shrub woodland) contained a small amount of useful temporal information and were affected regularly by noisily perturbations, while series corresponding to pixels associated with GM 3 (i.e. Mediterranean grassland) contained a higher amount of useful information.

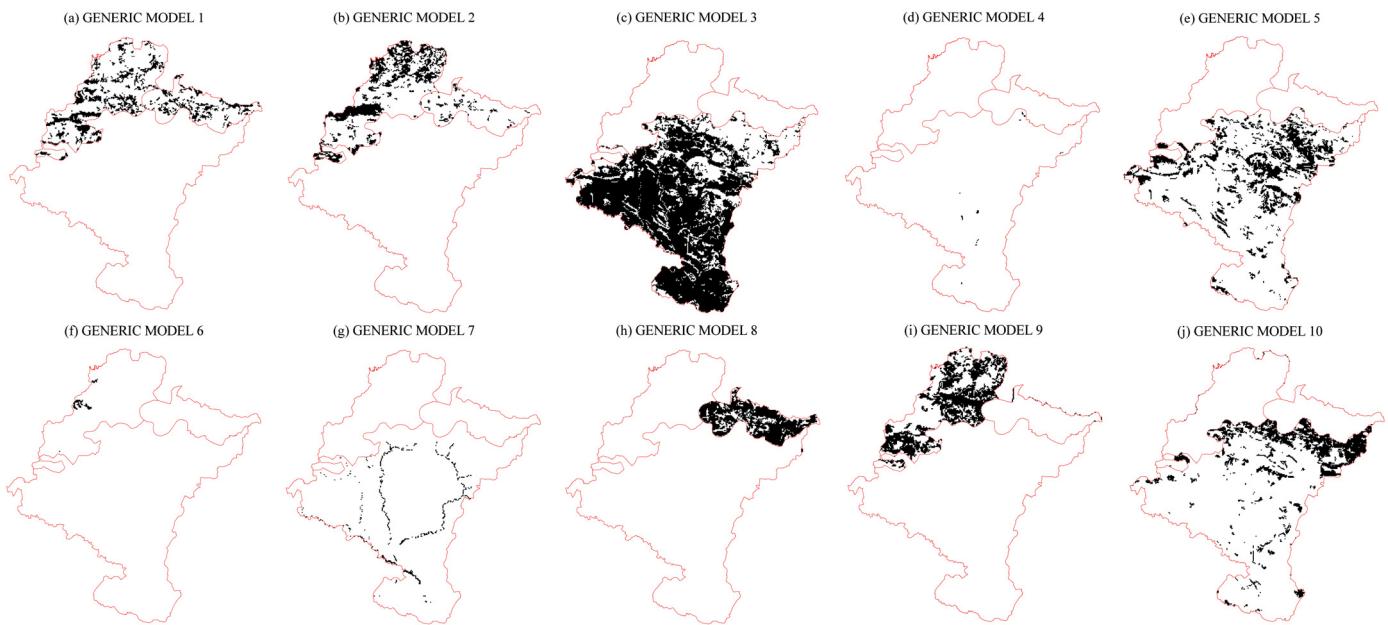
Model coefficients were estimated at the pixel level based on the corresponding GM. Fig. 7a–d shows the spatial distribution of the model intercept and the coefficients at short, medium and long term time scales respectively (approximately 1, 23 and 46 lags). The adequacy of each EM was assessed by the L-B Q statistic of the residuals. The test results indicate that in 95.44% of the cases, residuals are white noise (i.e. they are time-independent). Table 6 shows the percentages of valid EMs per GM. Results from the goodness of fit tests for the EMs were found to be satisfactory for 95% of the modeled time series (39,039). The EMs associated with GM 3 (i.e. Mediterranean grassland) present the highest percentage of valid models (99.15%). More than 85% of EMs associated with each generic model, present white-noise structure in the residuals, except for the EMs associated with 'shrub-woodland' fuel types. In the generic model 6 (temperate shrub-woodland) and 7 (Mediterranean shrub-woodland), valid models represent 65.26% and 79.04% of EMs, respectively.

Fig. 8a and b show the spatial distribution of the Theil inequality coefficient  $U$  for the historic and the 2009 forecasting respectively. For the historic prediction, the  $U$  is smaller than 0.1 in 42% of the valid EMs (i.e. 17,535 EMs), and between 0.1 and 0.24 in 52% of them (i.e. 21,504 EMs). Similar results were found for the 2009 forecasting with slightly higher  $U$  values (i.e. less accurate) in this case. Lower  $U$  values are found in northern Navarre. Fig. 8c and d shows the histograms of the Theil inequality coefficient  $U$  for historic and future predictions respectively. Table 7 presents  $U$ , and the coefficient  $U$  proportions  $U^S$  and  $U^V$  (i.e. bias and variance) average values of the historic prediction per GM. The bias proportion was close to zero in all study regions, but significant differences were found in the variance proportion  $U^V$ , which in 2.26% of the valid EMs is higher than 0.5. In general terms, the EMs from the temperate and alpine regions are characterized by a higher  $U^V$ , specifically those derived from GMs 6 and 9, indicating a higher error in the variance, while in the Mediterranean region only models derived from GM 7 show this problem. Fig. 9 shows some examples of the original

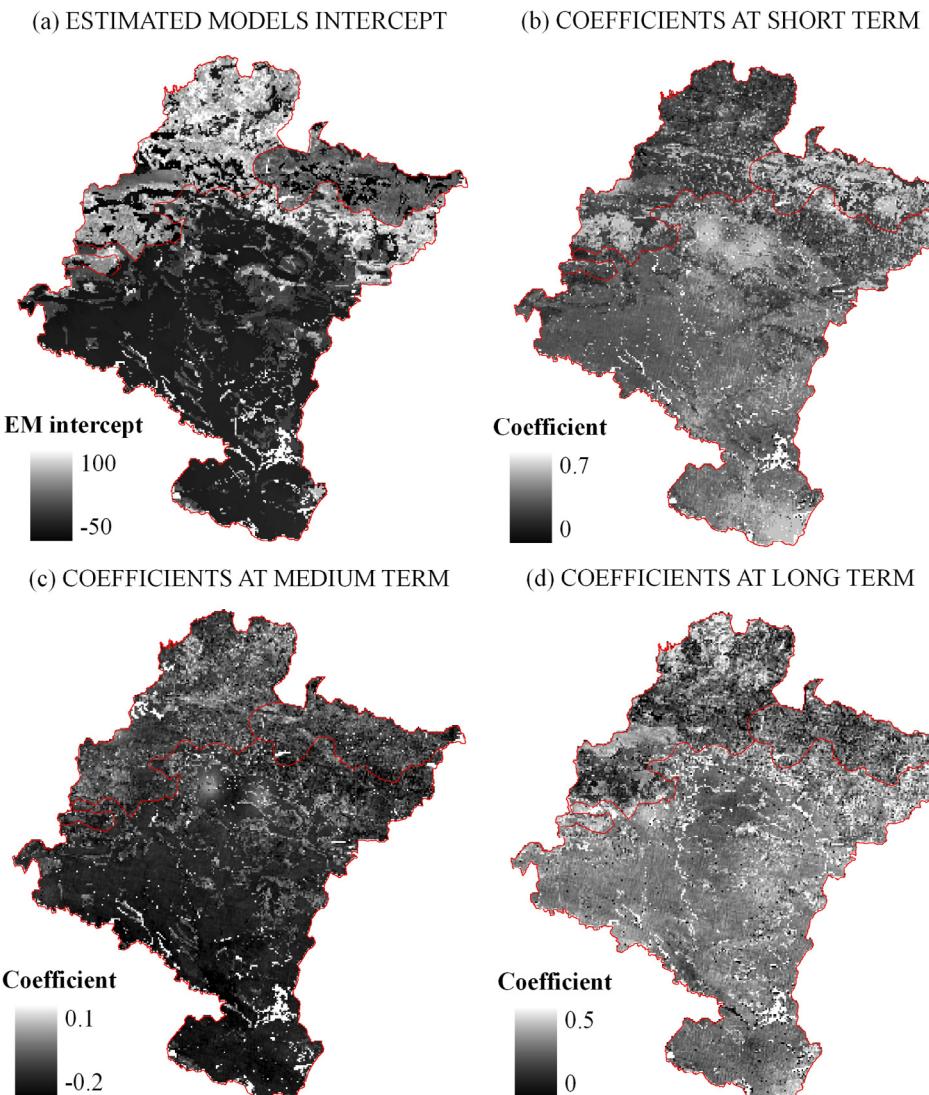
**Table 6**

Mean Ljung–Box Q values at lag 46 (one year), the number of the valid estimated models for each generic model group, and the corresponding percentage in relation to the total number of pixels associated to each generic model.

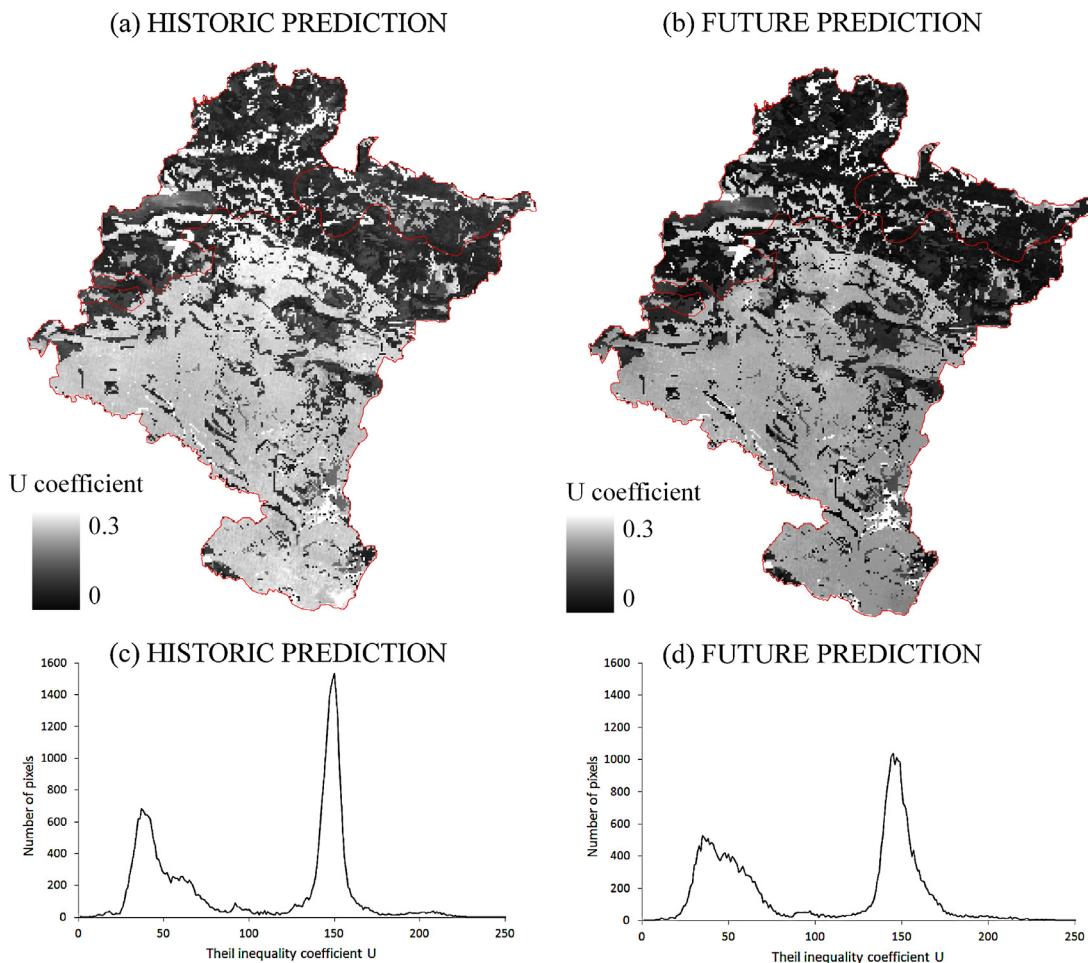
Generic model	Q-Value	Valid estimated models	
		Number	Percentage (%)
1 (Alpine and temperate grassland)	5808.15	2756	98.89
2 (Alpine and temperate shrubs)	3747.99	2848	94.56
3 (Mediterranean grassland)	7079.37	18,274	99.15
4 (Mediterranean shrubs-grassland)	5574.75	33	97.06
5 (Mediterranean shrubs)	6216.74	5003	97.91
6 (Temperate shrubs-woodland)	735.63	62	65.26
7 (Mediterranean shrubs-woodland)	4362.05	360	70.04
8 (Alpine woodland)	3248.64	2597	92.78
9 (Temperate woodland)	2127.24	3559	86.70
10 (Mediterranean woodland)	4553.91	3547	88.32



**Fig. 6.** Spatial distribution of pixels associated to the 10 generic models.



**Fig. 7.** Spatial distribution of the intercept (a), the coefficients at short term (lag 1) (b), medium-term (lags 21–22) (c) and long term (lag 46) (d). White means NODATA.



**Fig. 8.** Spatial distribution of Theil inequality coefficient  $U$  values for the historic prediction (a) and the future prediction of the year 2009 (b) for the valid estimated models. Gray scale, black (0) and white (0.3) mean minimum and maximum value respectively. Theil inequality coefficient  $U$  histogram for the historical prediction (c) and the future prediction (d).

**Table 7**

Theil inequality coefficient  $U$  and the bias and variance proportion for the historic prediction. In addition, the percentage of valid estimated model with a  $U^V$  between the following ranges: 0–30; 30–50 and >50.

Generic model	$U$	$U^S$	$U^V$	Percentage valid EM		
				$U^V$ (0–0.3)	$U^V$ (0.3–0.5)	$U^V$ (>0.5)
1	0.1567	0.0001	0.1110	98.15	1.52	0.33
2	0.0690	0.0690	0.2363	79.43	17.87	2.70
3	0.1397	0.0006	0.0644	99.53	0.10	0.37
4	0.0621	0.0003	0.1380	93.94	6.06	0.00
5	0.0669	0.0001	0.0832	99.76	0.04	0.20
6	0.0403	0.0005	0.6295	9.68	14.52	75.80
7	0.0780	0.0009	0.2238	75.56	8.05	16.39
8	0.0504	0.0001	0.2855	58.88	36.12	5.00
9	0.0480	0.0004	0.3462	37.18	50.04	12.78
10	0.0498	0.0004	0.1870	87.79	11.39	0.82

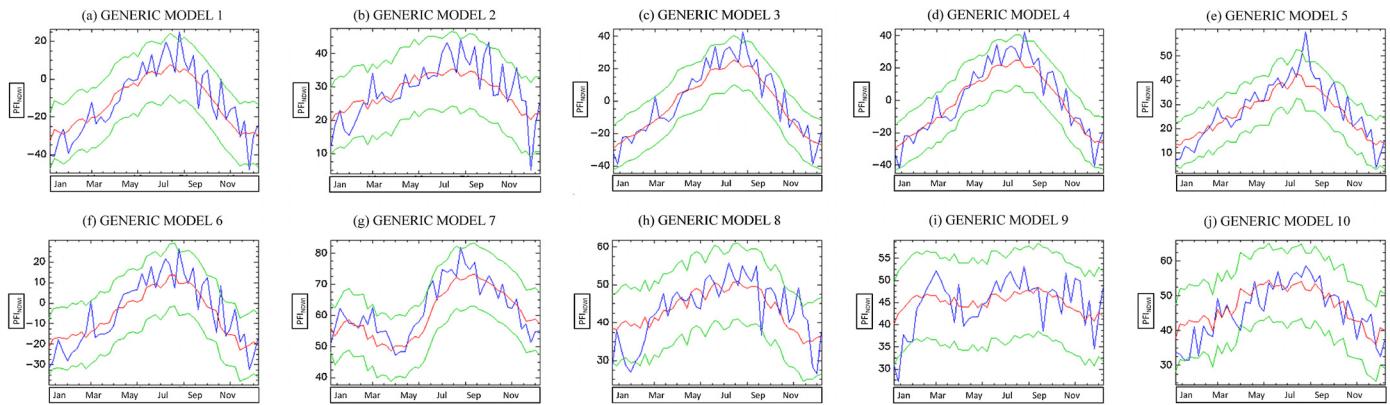
and 2009 forecast  $FPI_{NDWI}$  time series; prediction intervals are also represented.

## 6. Discussion

The  $FPI_{NDWI}$  time series dynamics in Navarre (Spain) was analyzed to build forecasting models in an upcoming year on a pixel basis.

The spatial distribution of the Ljung–Box  $Q$  values (Fig. 3) shows the level of useful information in the  $FPI_{NDWI}$  time series studied. Although some of the time series graphs explored suggest a high level of randomness (i.e. white noise), the results of the  $Q$

test show that most of the series contain relevant information and can be modeled. Temperate and alpine regions have lower  $Q$  values indicating a lower temporal dependency in these time series. Furthermore, the time series information content is higher for fuel types when located in the Mediterranean region. This fact is probably due to the common presence of clouds that could distort the information or mask temporal changes (Barret and Curtis, 1999). Thus, this test is especially appropriate for remote sensing data which is usually affected by artifacts that modify the signal, making the identification of useful information particularly difficult. In addition, results of this work show that time series from ecoregions and fuel types associated with lower  $Q$  values have a higher



**Fig. 9.** Predictive (red) and original (blue) forest fire risk for the year 2009 for the 10 generic models. In green are presented the prediction intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

percentage of non-valid time series models as well as lower forecasting capability. Thus, this test may also indicate the complexity level involved in the modeling process; in general terms, time series with low  $Q$  values will be more difficult to model (Box et al., 1994).

A general dynamic fire risk pattern dependent on short, medium, and long terms has been identified in the initial 26 EMR; however specific selected lags differ among models indicating the existence of distinct unique dynamics. By grouping EMR with common dynamic factors, it was possible to represent the whole dynamics variability using 10 generic models (GMs) (Table 6). This procedure facilitates the modeling process at image level while maintaining relevant information.

The 10 GMs were consistent with bioclimatic regions and fuel types; while  $FPI_{NDWI}$  series from the temperate and alpine regions could be modeled together, those from the Mediterranean region needed to be model independently (i.e. they had different specific lags). These results agree with the differences found in fire risk patterns (Huesca et al., 2009) and fire occurrence (Vélez, 2000, <http://www.magrama.es/>) between ecoregions, underlying the need to predict fire potential with different GMs. Furthermore, the spatial variability found in the study area time series suggested the appropriateness of developing more specific forecasting models within each GM. This was accomplished by estimating specific AR coefficients and intercepts on a pixel basis. Furthermore, the availability of spatially continuous data allows a more detailed assessment of model characteristics and spatial coherence (Fig. 7). The EM intercept values (Fig. 7a) are lower in areas dominated by herbaceous species while fuel types associated with litter as the element of fire spread show higher values. Since this parameter is directly related to the  $FPI_{NDWI}$  time series mean value, differences could be due to the influence of the amount of fuel on the fire potential; thus, in general terms, fuel types with high amounts of fuel (i.e. those associated with litter) result in higher potential mean values.

Model coefficients (Fig. 7b and c) show a high spatial consistency and do not seem to be as strongly related to fuel types, indicating that they are probably able to capture the distinct pixel  $FPI_{NDWI}$  dynamics. Positive AR coefficient values at short-term (at lags 1, 2, or  $3 \geq 8, 16$  or 24 days) indicate that forest fire risk depends largely on the fire risk on the preceding MODIS date ( $\geq 8$  days before) that is, on preceding fuel status and meteorological conditions. While positive coefficient values at long term (AR parameter at lags 45 or  $46 \geq 1$  year) denote the annual fire potential pattern, smaller values respect to the short term coefficients indicate a lower one year  $FPI_{NDWI}$  dependency likely due to inter-annual variability.

Higher coefficients at short and long terms in the Mediterranean area indicate a higher temporal dependence of  $FPI_{NDWI}$ , which was also corroborated by high autocorrelation values at 8 days and one year. This evinces a more regular  $FPI_{NDWI}$  pattern in this region,

probably due to the summer drought regularity (Valladares, 2004). A combination of lower coefficients in the temperate area and lower absolute autocorrelation values at short and long terms denotes a less regular  $FPI_{NDWI}$  pattern in this region. This is probably due to the lack of a marked dry season which results in a highly variable summer moisture regime. Short term model coefficients in the alpine region were clearly higher than long term coefficients, suggesting less temporal dependency of the  $FPI_{NDWI}$  at long term. This result agrees with the irregular pattern characteristic of forest fires in the alpine region (Vélez, 2000) and was also corroborated with the ACF.

Regarding the different fuel types, grasslands model coefficient presented a highly homogenous spatial distribution with high coefficient values at both the short and long terms. On the other hand, the high spatial variability in woodlands may be explained in terms of vegetation composition. Within this type of fuel, pixel time series from evergreen forest showed high coefficient values at short and long terms while those from deciduous forest presented lower coefficient values at long term, indicating a more irregular annual pattern. These results could be explained in terms of the synchrony of the maximum amount of fuel and weather conditions (i.e. maximum temperatures and minimum relative humidity). Extreme fire weather conditions in deciduous forests are coincident with maximum greenness during summertime and this compensatory effect results in a more irregular pattern (Noormets, 2009).

Negative coefficients at lags 21–23 (approximately 6 months) indicate a significant negative relationship at medium-term in most of the study area (except in the temperate shrub woodlands). The negative values, located in the temperate woodlands, may be due to the negative relationship between spring and fall fire risk; wet springs can generate large amounts of biomass that are then available for burning during fall (Huesca et al., 2009). On the other hand in the alpine region there is no a clear difference in the coefficient values among fuel types, showing close to zero values in most of the area. It seems that most of the temporal dependency is accounted for by the short term coefficients due to the irregular  $FPI_{NDWI}$  patterns in this region.

Negative coefficients at lag 10 (approximately 2.5 months) associated with fuel types linked to grasslands and shrubs, show a faster response of grasslands to changes in weather condition than other types of ecosystems (García et al., 2010; Morison and Morecroft, 2006). The occurrence of low  $FPI_{NDWI}$  in springtime due to high biomass will result in high amounts of fuel in summer and higher fire potential.

The high percentage of valid models in the entire study area (i.e. pixel time series whose model residuals are white noise) shows the adequacy of the methodology applied for explaining systematic  $FPI_{NDWI}$  variability. The lower percentage of valid models is found in

pixel time series associated with temperate shrub-woodlands (GM 6) and Mediterranean shrub-woodlands (GM 7). This fact may be due to several factors: (1) both groups are represented by a reduced number of pure pixels, (2) time series associated with GM 6 present a high level of noise (low  $Q$  statistic value), and (3) time series associated with GM 7 are linked to riparian vegetation that present high variability, decreasing the representativeness of the pure pixels.

The presence of non-valid models in pixels associated with litter (GM 8–10) could be due to the presence of evergreen and deciduous forests in the same fuel type, which results in time series showing a mixed pattern. Autocorrelation function showed that, under the same weather conditions,  $FPI_{NDWI}$  amplitude is larger in evergreen than in deciduous forests resulting in a better forecasting ability, similar results were found by Jiang et al. (2010) who reported that more accurate models were obtained with time series that presented more distinct seasonal patterns.

The model forecasting accuracy, historic (i.e. prediction is based on real values) and future (i.e. prediction is based mostly on predicted values) was measured by the Theil's  $U$  coefficient. Small values (Fig. 9a and b) evince high forecasting accuracy in both cases, with moderately higher values in the future forecasting. They show a similar spatial distribution and no significant differences among ecoregion. Although Mediterranean ecoregion shows slightly higher  $U$  values (i.e. higher model accuracy), most of the error is associated with the  $U$  covariance component indicating the error random nature that does not influence the forecasting capability. On the other hand, the model accuracy in terms of variance is statistically different between the Mediterranean and other two ecoregions. Higher errors associated with the variance are found in the temperate and alpine regions may be due to the time series irregularity together with a high level of noise (Alhamad et al., 2007) usually present in remote sensing time series. In this context Jiang et al. (2010) found a direct relationship between data quality and forecast accuracy, Verbesselt et al. (2010b) found that the signal-to-noise ratio significantly affected change detection results, and Liang (2001) obtained less accurate results in the classification procedure using periodograms or ARIMA models due to the influence of cloudy pixels.

Our results indicate the validity of the proposed hypothesis "that different ecoregions and fuel types are characterized by different forest fire behavior, thus each of them should be represented by its own model" and agrees with findings from other researches using different variables. For instance, Yurekli and Kurunc (2006) found that in order to simulate agricultural drought duration, different ARIMA models for different hydrologic homogeneous sections were needed, while Fernández-Manso et al. (2011) found better results using two SARIMA models to describe NDVI time series for different pines species.

## 7. Conclusions

The present study showed the capability of ARIMA models to forecast  $FPI_{NDWI}$  as proxy of forest fire risk on a pixel basis. Accurate pixel autoregressive models have been estimated based on specific time series identity in 93.18% of the pixels analyzed. In addition the methodology implemented makes it possible to assess information content and forecasting capability of remote sensing time series in a spatially continuous framework.

Specific conclusions are as follows:

- (1) The usefulness of developing autoregressive models using original data and introducing significant seasonal AR parameters for modeling and forecast  $FPI_{NDWI}$  was demonstrated.
- (2) The approach of selecting a few generic models to represent the general  $FPI_{NDWI}$  dynamics patterns made it possible to

implement the Box and Jenkins methodology operatively and reach high model accuracy on a pixel basis.

- (3) The differences found in temperate-alpine and Mediterranean regions suggest the influence of low data quality in terms of information content and model accuracy, which is especially relevant in remote sensing time series.

Results of this work show that the spatio-temporal dimension that remote sensing data provide combined with statistical time series analysis, makes it possible to develop robust applications for environmental monitoring and forecasting. It is expected that models accuracy and specificity will improve as longer time series are available.

## Acknowledgments

This study was co-funded by "Fundación MAPFRE" ("Ayudas a la investigación, Salud, prevención y medio ambiente y seguros"), Tragsatec (FPA110000EST1005) and by the Ministry of Agriculture, Food and Environment (AGL-2010-17505).

## References

- Alhamad, M.N., Stuth, J., Vannucci, M., 2007. Biophysical modelling and NDVI time series to project near-term forage supply: spectral analysis aided by wavelet denoising and ARIMA modeling. *Int. J. Remote Sens.* 28, 2513–2548.
- Barret, E.C., Curtiss, L.F., 1999. *Introduction to Environmental Remote Sensing*. Stanley Thornes (Publishers) Ltd.
- Barsky, R.B., Miron, J.A., 1989. The seasonal cycle and the business cycle. *J. Polit. Econ.* 97, 503–534.
- Beaulieu, J., Miron, J., 1993. Seasonal unit roots in aggregate US data. *J. Econometrics* 55, 305–328.
- Beck, P.S.A., Juday, G.P., Alix, C., Barber, V.A., Winslow, S.E., Sousa, E.E., et al., 2011. Changes in forest productivity across Alaska consistent with biome shift. *Ecol. Lett.* 14, 373–379.
- Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F., 2012. ESA's sentinel missions in support of Earth system science. *Remote Sens. Environ.* 120, 84–90.
- Box, G., Jenkins, G., Reinsel, G., 1994. *Time Series Analysis: Forecasting and Control*, 3rd ed. Prentice-Hall, Englewood Cliffs, New Jersey.
- Box, G.E.P., Jenkins, G.M., 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Bunn, A.G., Scott, G.J., 2006. Trends in satellite-observed circumpolar photosynthetic activity from 1982 to 2003: the influence of seasonality, cover type, and vegetation density. *Earth Interact.* 10, 1–19.
- Burgan, R.E., Klaver, R.W., Klaver, J.M., 1998. Fuel models and fire potential from satellite and surface observations. *Int. J. Wildland Fire* 8, 159–170.
- Darné, O., Litago, J., Terraza, M., 2002. Tests de racines unitaires saisonnières pour des données journalières. *Revue de Statistique Appliquée* 2, 71–91.
- Dickey, D., Fuller, W., 1979. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* 74 (366), 427–431.
- Fensholt, R., Langanke, T., Rasmussen, K., Reinberg, A., Prince, S.D., Tucker, C., et al., 2012. Greenness in semi-arid areas across the globe 1981–2007 – an Earth Observing Satellite based analysis of trends and drivers. *Remote Sens. Environ.* 121, 144–158.
- Fernández-Manso, A., Quintano, C., Fernández-Manso, O., 2011. Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale. *Int. J. Remote Sens.* 32, 1595–1617.
- Fosberg, M.A., Deeming, J.E., 1971. Derivation of the 1- and 10-hour timelag fuel moisture calculations for fire danger rating, vol. 207. Rocky Mountain Forest and Range Experiment Station, Forest Service, US Dept. of Agriculture.
- Franses, P.H., 1991. Seasonality, nonstationarity and the forecasting of monthly time series. *Int. J. Forecasting* 7, 199–208.
- Fuller, W.A., 1976. *Introduction to Statistical Time Series*. John Wiley and Sons, New York.
- Gao, B., 1996. NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266.
- García, M., Litago, J., Palacios-Orueta, A., Pinzón, J.E., Ustin, S.L., 2010. Short-term propagation of rainfall perturbations on terrestrial ecosystems in central California. *Appl. Veg. Sci.* 13, 146–162.
- Gemitz, A., Stefanopoulos, K., 2011. Evaluation of the effects of climate and man intervention on ground waters and their dependent ecosystems using time series analysis. *J. Hydrol.* 403, 130–140.
- Ghysels, E., Osborn, D.R., 2001. *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press, Cambridge.
- Granger, C.W.J., Newbold, P., 1977. *Forecasting Economic Time Series*. Academic Press, New York.
- Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton.

- Han, P., Wang, P.X., Zhang, S.Y., Zhu, D.H., 2010. Drought forecasting based on the remote sensing data using ARIMA models. *Math. Comput. Model.* 51, 1398–1403.
- Harvey, Andrew, C., 1981. *Time Series Models*. Phillip Allen, Oxford.
- Horion, S., Cornet, Y., Erpicum, M., Tychon, B., 2013. Studying interactions between climate variability and vegetation dynamic using a phenology based approach. *Int. J. Appl. Earth Obs. Geoinf.* 20, 20–32.
- Huesca, M., Palacios-Orueta, A., Litago, J., Merino-de-Miguel, S., San-Román-Ortiz, J., Cicuendez, V., Maqueda-Bueno, J., 2010. Metodología de integración de información de teledeteción para la prevención de incendios forestales mediante la predicción estadística. *Seguridad y Medio Ambiente* 125, 22–36.
- Huesca, M., Litago, J., Palacios-Orueta, A., Montes, F., Sebastián-López, A., Escribano, P., 2009. Assessment of forest fire seasonality using MODIS fire potential: a time series approach. *Agric. Forest Meteorol.* 149, 1946–1955.
- Hylleberg, S., Engle, R., et al., 1990. Seasonal integration and cointegration. *J. Econometrics* 44, 215–238.
- IPCC, 2007. *Climate Change 2007: Impacts, Adaptation and Vulnerability: Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*.
- Jiang, B., Liang, S., Wang, J., Xiao, Z., 2010. Modeling MODIS LAI time series using three statistical methods. *Remote Sens. Environ.* 114, 1432–1444.
- Liang, S., 2001. Land-cover classification methods for multi-year AVHRR data. *Int. J. Remote Sens.* 22, 1479–1493.
- Ljung, G., Box, G., 1978. On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.
- Maddala, G.S., Kim, I.M., 1998. *Unit Roots, Cointegration and Structural Change*. Cambridge University Press, Cambridge.
- Mariño, M.A., Tracy, J.C., Taghavi, S.A., 1993. Forecasting of reference crop evapotranspiration. *Agric. Water Manage.* 24, 163–187.
- Modarres, R., 2007. Streamflow drought time series forecasting. *Stoch. Environ. Res. Risk Assess.* 21, 223–233.
- Morison, J.L., Morecroft, M.D., 2006. *Plant Growth and Climate Change*. Blackwell Publishing Ltd.
- Noormets, A., 2009. *Phenology of Ecosystem Processes. Applications in the Global Change Research*. Springer, Dordrecht, Heidelberg, London, New York.
- Osborn, D.R., 1990. A survey of seasonality in UK macroeconomic variables. *Int. J. Forecasting* 6, 327–336.
- Pausas, J.G., 2004. Changes in fire and climate in the Eastern Iberian Peninsula (Mediterranean basin). *Clim. Change* 63, 337–350.
- Peng, D., Huang, J., Huete, A., Yang, T., Gao, P., Chen, Y., et al., 2010. Spatial and seasonal characterization of net primary productivity and climate variables in southeastern China using MODIS data. *J. Zhejiang Univ. Sci. B* 11, 275–285.
- Piwowar, J.M., Ledrew, E.F., 2002. ARMA time series modelling of remote sensing imagery: a new approach for climate change studies. *Int. J. Remote Sens.* 23, 5225–5248.
- Plosser, C.I., Schwert, G.W., 1978. Money, income, and sunspots: measuring economic relationships and the effects of differencing. *J. Monetary Econ.* 4, 637–660.
- Plosser, C.I., Schwert, G.W., 1977. Estimation of a non-invertible moving average process. The case of overdifferencing. *J. Econometrics* 6, 199–224.
- Schneider, P., 2008. A VARI-based relative greenness from MODIS data for computing the Fire Potential Index. *Remote Sens. Environ.* 112, 1151–1167.
- Sebastián-López, A., San-Miguel-Ayanz, J., Burgan, R.E., 2002. Integration of satellite sensor data, fuel type maps and meteorological observation for evaluation of forest fire risk at the pan-European scale. *Int. J. Remote Sens.* 23, 2713–2719.
- Telesca, L., Amatulli, G., Lasaponara, R., Lovallo, M., Santulli, A., 2005. Time-scaling properties in forest-fire sequences observed in Gargano area (southern Italy). *Ecol. Model.* 185, 531–544.
- Theil, H., 1971. *Principles of Econometrics*. John Wiley and Sons, Inc., New York.
- Valladares, F., 2004. *Ecología del bosque mediterráneo en un mundo cambiante*. Ministerio de Medio Ambiente, Egraf, Madrid.
- Vélez, R., 2000. La defensa contra incendios forestales. Fundamentos y experiencias. McGrawHill, Spain, pp. 1.3–9.43.
- Verbesselt, J., Hyndman, R., Newham, G., Culvenor, D., 2010a. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* 114, 106–115.
- Verbesselt, J., Hyndman, R., Zeileis, A., Culvenor, D., 2010b. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sens. Environ.* 114, 2970–2980.
- Xiao, Z., Liang, S., Wang, J., Jiang, B., Li, X., 2011. Real-time retrieval of Leaf Area Index from MODIS time series data. *Remote Sens. Environ.* 115, 97–106.
- Yurekli, K., Kurunc, A., 2006. Simulating agricultural drought periods based on daily rainfall and crop water consumption. *J. Arid Environ.* 6, 629–640.