

# Satellite-based forest monitoring: spatial and temporal forecast of growing index and short-wave infrared band

Caroline Bayr,<sup>1</sup> Heinz Gallaun,<sup>2</sup> Ulrike Kleb,<sup>1</sup> Birgit Kornberger,<sup>1</sup> Martin Steinegger,<sup>2</sup> Martin Winter<sup>3</sup>

<sup>1</sup>Joanneum Research, Statistical Applications Research Group, Graz; <sup>2</sup>Joanneum Research, Remote Sensing and Geoinformation Research Group, Graz; <sup>3</sup>Joanneum Research, Digital, Audiovisual Media Group, Graz, Austria

## Abstract

For detecting anomalies or interventions in the field of forest monitoring we propose an approach based on the spatial and temporal forecast of satellite time series data. For each pixel of the satellite image three different types of forecasts are provided, namely spatial, temporal and combined spatio-temporal forecast. Spatial forecast means that a clustering algorithm is used to group the time series data based on the features normalised difference vegetation index (NDVI) and the short-wave infrared band (SWIR). For estimation of the typical temporal trajectory of the NDVI and SWIR during the vegetation period of each spatial cluster, we apply several methods of functional data analysis including functional principal component analysis, and a novel form of random regression forests with online learning (streaming) capability. The temporal forecast is carried out by means of functional time series analysis and an autoregressive integrated moving average model. The combination of the temporal forecasts, which is based on the past of the considered pixel, and spa-

tial forecasts, which is based on highly correlated pixels within one cluster and their past, is performed by functional data analysis, and a variant of random regression forests adapted to online learning capabilities. For evaluation of the methods, the approaches are applied to a study area in Germany for monitoring forest damages caused by wind-storm, and to a study area in Spain for monitoring forest fires.

## Introduction

Forests and other wooded land which cover over 40% of the European Union's land area, are multifunctional and are serving economic, social and environmental purposes. The urgent need for monitoring forest health and vitality, including forest disturbances (*i.e.*, forest fires, storm damage, drought stress, insect and disease outbreaks) is emphasised *e.g.* by the Green paper (European Commission, 2010) and in the new EU Forest Strategy (European Commission, 2013).

Currently, operational monitoring of forests is mainly applied on the basis of ground sampling data collected in national forest inventories. Typically, for several thousand randomly selected plots with a size below one hectare each, detailed forest parameters are measured in the field. The field measurements are time and cost intensive and typically the measurements, data processing, statistical analysis and reporting requires several years until the final results are provided. Whereas such inventories can provide accurate estimation of forest parameters at the regional level, they do not provide spatially explicit information. Further, in case of forest disturbances such as *e.g.* caused by storm damage or forest fires, up-to-date information on the damages is required. Remote sensing based monitoring methods can complement traditional field based national forest inventories for both aspects. Especially with the upcoming Sentinel2 satellites, remote sensing based forest monitoring capabilities will be improved significantly, as the Sentinel2 satellites will provide wall-to-wall data on forests with high spatial resolution of 10 m, high temporal resolution of 5 days and high radiometric resolution by providing measurements covering the optical, near infrared and short-wave infrared spectral regions.

Current remote sensing methods for monitoring forests are often based on the assessment of one to several image acquisitions. However, to fully utilise the comprehensive information content of up-coming Sentinel2 satellites, methods which exploit the information provided by dense time series are required. Dense time series are currently acquired by medium resolution sensors such as moderate resolution imaging spectrometer (MODIS) or Proba-V vegeta-

Correspondence: Birgit Kornberger, Joanneum Research, Statistical Applications Research Group, Leonhardstraße 59, 8010 Graz, Austria.  
Tel.: +43.316.8761549 - Fax: +43.316.87691549.  
E-mail: birgit.kornberger@joanneum.at

Key words: Satellite images; Forest monitoring; Functional time series analysis; Autoregressive integrated moving average; Online random regression forests.

Acknowledgments: this work was funded by the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT), as part of the project Multi-Dimensional Sensor Data Time Series Analysis (MUST).

Received for publication: 12 January 2015.

Revision received: 3 July 2015.

Accepted for publication: 16 October 2015.

©Copyright C. Bayr et al., 2016

Licensee PAGEPress, Italy

Geospatial Health 2016; 11:310

doi:10.4081/gh.2016.310

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.



tion instrument, but at relatively low spatial resolution. For the current work, we used MODIS as pre-processed data is easily accessible. However, the tested methods can also be used with higher spatial resolution time series imagery, e.g. from Sentinel2 for future forest monitoring applications.

To allow qualitative and quantitative evaluations, the forecasting methods were applied to a test site in Germany for monitoring forest damages caused by wind storm, and to a test site in Spain for monitoring forest fires. The achieved accuracies were evaluated for both test sites for all methods on the basis of a sampling approach.

## Materials and Methods

### Time series data

As basis for the methodological tests, we used time series of the MODIS product MCD43A4, Version 5, which provides 500 m reflectance data adjusted using a bidirectional reflectance distribution function. To reduce noise and to fill data gaps in the original MCD43A4 MODIS data (e.g. cloud covered areas), we applied Savitzky-Golay filtering (Savitzky and Golay, 1964) over n=15 MODIS MCD43A4 8 day mosaics. This leads to a significant noise reduction and provides as result a spatial and temporal consistent wall-to-wall cover also in areas with frequent cloud-cover such as, e.g. in the Alps. This filtering leads to a smoothed signal which is taken into account in the change detection process as described in the results section below. Of the time series data we used band 6, which covers the short-wave infrared spectral region (1628-1652 μm), and the *normalised difference vegetation index* (NDVI) (Kriegler *et al.*, 1969) with:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

where NIR is the reflectance in the near infrared spectral region and RED is the reflectance in the red spectral region. We selected the NDVI as index because of the long term proven robustness in various applications and ecosystem regions. Further, this index only requires data in the red and near infrared spectral regions, which is provided by a large number of operational satellites. We used the data over forested areas over the vegetation period from end of May until end of August in 8-day steps with temporal coverage from 2000 to 2013.

### Study area

Two study areas were selected for testing and evaluation of the methods: Study area 1 is located in the south of Germany where storm damages occurred in the monitoring period. The MODIS time series covers this area as a raster with 541 x 541 pixels with a spatial resolution of 500 x 500 m on the ground. Study area 2 is located in northern Spain where forest fires frequently occur. Also this study area is covered with MODIS time series raster data of 541 x 541 pixels with a spatial resolution of 500 x 500 m.

### Functional time series analysis

Actually, functional data analysis (FDA) is an infinite dimensional (Hall and Hosseini-Nasab, 2006) analysis method. Main characteristics of functional data are repeated measurements, high frequency and multi dimensions, and that it is taken as function of an independent variable. An advantage of FDA is the usage of additional informa-

tion of functions, e.g. the slope, curvature and other characteristics of curves.

If the measurements are dependent on time, we call them functional time series. With the technique of functional time series analysis (FTSA) it is possible to *capture the underlying dynamic of seasonality in the data* (Shang, 2013). Let  $X_t(x)$  be measurements, e.g. the vegetation index in year  $t$  with discrete observation number  $x$ . Let  $f_t(x)$  denote functional data,  $\sigma_t(x)$  describes the quantity of noise varying with  $x$ , and  $\omega_{t,i}$  is a standard normal variable with zero mean and unit variance  $\sigma_\omega^2$ . Then, the smooth function  $f_t(x)$  can be extracted from the following equation:

$$X_t(x_i) = f_t(x_i) + \sigma_t(x_i)\omega_{t,i}$$

where time  $t=1, \dots, n$  with  $n$  as the number of years considered, and  $i=1, \dots, p$  with  $p$  as the number of observations during one year. In Germany and Spain, the number of observations  $p$  equals 33. The smooth function  $f_t(x)$  can be calculated using functional principal component analysis (FPCA), which reduces the infinite dimension of functional data to a finite level, and *points to the most significant components of the data* (Hall and Hosseini-Nasab, 2006). In the literature, there are numerous examples for FPCA with different application fields: sea surface temperature forecasting (Shang and Hyndman, 2011), call center arrivals forecasting (Shen and Huang, 2008), age-specific breast cancer mortality forecasting (Erbas *et al.*, 2007) and fertility rates forecasting (Hyndman and Shang, 2009).

The functional principal component decomposition can be described through the following equation:

$$f_t(x) = \bar{f}(x) + \sum_{k=1}^K \hat{\beta}_{t,k} \hat{\Phi}_k(x) + \hat{\omega}_t(x)$$

where  $\bar{f}(x) = \frac{1}{n} \sum_{t=1}^n f_t(x)$  is the estimated mean function,  $\hat{\Phi}_k(x)$  is

the  $k^{th}$  estimated orthonormal eigenfunction of the empirical covariance operator:

$$\hat{E} = \frac{1}{n} \sum_{t=1}^n [f_t(x) - \bar{f}(x)][f_t(x) - \bar{f}(x)],$$

which maximises the variance,  $\hat{\beta}_{t,k}$  is the  $k^{th}$  principal component score for year  $t$ . This score is given by the projection of  $f_t(x) - \bar{f}(x)$  in the direction of the  $k^{th}$  eigenfunction  $\hat{\Phi}_k(x)$ , that is  $\hat{\beta}_{t,k} = \int [f_t(x) - \bar{f}(x)] \hat{\Phi}_k(x) dx$ . The residual is denoted by  $\hat{\omega}(x)$  as *iid* random functions with zero mean, and the optimal number of principal components is denoted by  $K < n$  (Hall and Hosseini-Nasab, 2006; Shang, 2013). The estimated mean function can also be weighted unequally with geometrically decreasing weights  $w_t = \lambda(1-\lambda)^{n-t}$  with  $0 < \lambda < 1$ , and can be written in the form  $\bar{f}(x) = \sum_{t=1}^n w_t f_t(x)$ . Since we assume temporal autocorrelation, a reason for decreasing weights is the probability that *more recent data affect the results more than data in the distant past* (Hyndman and Shang, 2009).

The main aim of our calculations is to receive h-step-ahead forecasts following the equation:

$$\hat{X}_{n+h|n}(x) = \bar{f}(x) + \sum_{k=1}^K \hat{\beta}_{n+h|n,k} \hat{\Phi}_k(x)$$

by conditioning on the smooth functions  $f(x) = [f_1(x), \dots, f_n(x)]^T$  and



the functional principal components  $\Phi(x) = [\Phi_1(x), \dots, \Phi_k(x)]^T$ . The h-step-ahead forecasts of the principal component scores  $\beta_{n+h,k}$  are denoted by  $\hat{\beta}_{n+h,k}$  and are obtained by using an univariate time series model, e.g. an exponential smoothing (ETS) model. The purpose of the ETS technique is the estimation of a trend and seasonal component with additive and/or multiplicative terms (or none of them).

The error, trend and seasonal component follow the equations (Hyndman *et al.*, 2002):

$$\begin{aligned} \text{Level: } l_t &= \alpha L_t + (1 - \alpha)K_t, \\ \text{Slope: } s_t &= \beta S_t + (\delta - \beta)s_{t-1}, \\ \text{Seasonal component: } c_t &= \gamma C_t + (1 - \gamma)c_{t-m}, \end{aligned}$$

where  $m$  denotes the number of seasons each year,  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  are constants. The values  $L_t$ ,  $K_t$ ,  $S_t$ , and  $C_t$  differ depending on the trend and seasonal component terms (Hyndman *et al.*, 2002; Hyndman and Khandakar, 2008). From the h-step-forecast  $\hat{X}_{n+h|n}$  and the variance of the prediction error  $\hat{\sigma}_{n+h|n}^2$  assuming Normal distribution, the lower and upper limits of the  $(1 - \alpha)$ -prediction interval can be calculated by  $\hat{X}_{n+h|n} \pm q_{\alpha/2} \sqrt{\hat{\sigma}_{n+h|n}^2}$  with  $q_{\alpha/2}$  as the  $\frac{\alpha}{2}$  th quantile of the Standard Normal distribution (Hyndman *et al.*, 2008).

### Autoregressive integrated moving average

A classical and widely used approach to model time series is the autoregressive integrated moving average (ARIMA) model. It is a flexible model, which combines the autoregressive (AR) model, the moving average (MA) model and the autoregressive moving average (ARMA) model. These three main model types can be used if the time series data are stationary. But in practice time series often have a trend over time or a non-constant mean and hence they are non-stationary. Such non-stationary time series can be handled with an integrated ARMA(p,q) model, the ARIMA(p,d,q) model. In the ARIMA approach the non-stationary time series is differenced  $d$  times to obtain a stationary process. Generally an ARIMA model (Shumway, 2006) can be written as:

$$\phi(B) (1-B)^d X_t = \theta(B)\omega_t$$

where  $X_t$  describes the time series,  $\omega_t$  is a Gaussian white noise series with mean zero and variance  $\sigma_\omega^2$  and  $B$  is the backshift operator with  $B^k X_t = X_{t-k}$ .  $\phi(B)$  and  $\theta(B)$  are the autoregressive and moving average operators, respectively, with:

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \end{aligned}$$

The ARIMA(p,d,q) model can be extended to an multiplicative seasonal autoregressive integrated moving average model to account for seasonal effects – so called SARIMA model, where autoregressive and moving average terms for seasonal effects are included in the same manner. The structure of this model – denoted as ARIMA(p,d,q)(P,D,Q) – is:

$$\Phi_P(B^S) \phi(B) (1 - B^S)^D (1 - B)^d X_t = \alpha + \Theta_Q(B^S) \theta(B)\omega_t$$

with autoregressive and moving average operators:

$$\begin{aligned} \Phi_P(B^S) &= 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_p B^{pS} \\ \Theta_Q(B) &= 1 + \Theta_1 B + \Theta_2 B^2 + \dots + \Theta_q B^q \end{aligned}$$

The model selection is done via Akaike information criterion (AIC).

Let the model forecast  $h$  steps ahead be  $\hat{X}_{n+h|n} = E(X_{n+h} | X_n, X_{n-1}, \dots)$  which is expressed through the conditional expectation of  $x_{n+h}$  given all past information up to  $n$ . The variance of the prediction error is  $\hat{\sigma}_{n+h|n}^2 = E((X_{n+h} - \hat{X}_{n+h|n})^2)$ . With the calculation of a prediction interval the precision of the forecast can be assessed. The lower and upper limits of the  $(1 - \alpha)$ -prediction interval are given by:

$$\hat{X}_{n+h|n} \pm q_{\alpha/2} \sqrt{\hat{\sigma}_{n+h|n}^2}$$

where  $q_{\alpha/2}$  denotes the  $X_{\frac{\alpha}{2}}$  quantile of the standard Normal distribution (Shumway, 2006).

The ARIMA model approach based on remote sensing data is applied for various questions. For example, Jiang *et al.* (2010) use a SARIMA model approach to model the leaf area index (LAI) time series of the years 2000 to 2006 and predict LAI values for the year 2007. The model and forecast results of the SARIMA approach are compared with two other methods with respect to different land cover types. Whereas the model performance of these three model approaches varies with land cover type, the SARIMA model and forecast perform very well overall. Han *et al.* (2010) use ARIMA models to predict drought in the Guanzhong Plain in China based on the vegetation temperature condition index (VTCI).

### Online random regression forests

Random forest classifiers (RFC) (Breiman, 2001) are well known and studied classification methods. They show better or at least comparable performance in comparison to other state of the art classification algorithms such as support vector machines (SVM) (Vapnik, 2000) or boosting technologies (Freund and Schapire, 1997). They have been successfully applied to a number of applications. Moreover, they have also recently been adapted to density estimation, manifold learning, semi-supervised learning and regression tasks in a very successful manner. They have also shown their practical feasibility in famous commercial products such as e.g. Microsoft's KINECT® (Microsoft Corp., Redmond, WA, USA). Criminisi *et al.* (2011) proposed a unified approach of random decision forests, which has been applied to a number of machine learning, computer vision and medical image analysis.

Analysing the change of a dependent variable (e.g. a certain forest monitoring parameter) with respect to a set of (multi-dimensional) independent variables (e.g. observations from satellite data) is a typical regression task. Thus it is natural to take into account regression variants of randomised decision trees as an alternative model generation technique. Since satellite data typically arrive sequentially and the processing time of model prediction becomes an important issue, multiple standard offline calculation of the regression is not feasible. Thus streaming or online learning capabilities become an essential requirement for that type of practical application.

We propose a novel regression method termed online random regression forests (ORRF) to build a proper, adaptive model for satellite data. The method follows to some extent the generic approach of

Criminisi's framework (2011) for transferring standard classification forests to regression problems. But coevally some of the concepts introduced by Saffari *et al.* (2009) have to be used to change structure, layout and learning rules of the forest in order to enable the algorithm for having proper online learning capabilities.

### The basics behind random forests

Random (classification) forests are an ensemble combination of several binary decision trees where each binary decision tree is treated independently. The final output decision of the forest is obtained by *e.g.* a simple majority voting of all the individual leaf node predictions or other combinations with respect to probabilities.

Each tree itself consists of a singular root node, which subsequently splits up into two child nodes in a hierarchical manner. A simple test function on a training sample is applied to decide about the path the sample moves down the tree. During the offline training, the *best split* for each individual tree node is calculated by globally optimising several random test functions with respect to the overall information gain obtained by each split.

Taking into account a large number of samples continuously arriving over time, the main disadvantages of such offline optimised classification forests are the increasing calculation time and the necessity for storing all samples. If boundary conditions change or if additional retraining is done the required execution time typically exceeds the processing capabilities of a system.

### Online adaptation of classification forests and introducing regression capabilities

Online adaptation of classification forests has been introduced by Saffari *et al.* (2009). The authors combined the ideas from online bagging and extremely randomised forests. They proposed a novel procedure (algorithm) for growing a decision tree in an online fashion for visual tracking and interactive real-time segmentation tasks.

For our approach we use the basic idea of online bagging in a similar way. In particular we use the basic principle of modelling the sequential arrival of data by Poisson distribution sampling according to Oza and Russell (2001). This allows continuous growing and updating the tree structure.

Although the main difference between classification and regression trees is the change of the output labels from discrete class labelling to continuous prediction values, the scheme of Criminisi's framework (2011) cannot directly be applied due to numerous reasons. In the following we mention the most important ones for our problem and describe the solutions to deal with.

First, the strict *online learning constraints* (strict *online learning* means, that it is not allowed/possible to store any data samples for later usage. Data has to be processed *on the fly* with no buffering as *e.g.* used in *incremental learning*) require for an incremental update of the individual probabilities and statistics for each node of the forest. This is no problem for the classification case, because class-statistics are discrete histograms and their update is trivial.

As regression trees use *continuous densities* it is necessary to store them and to enable incremental update without having access to previous data points. Although it would be possible to sample the density function and store as a regular raster, this approximation of the density function turns out to be not feasible because of a high memory amount required. We approximate the density probability function by a simple Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . The two parameters can be incrementally updated via the following formulas:

$$W_n = \sum_{i=1}^n w_i$$

$$\mu_n = \mu_{n-1} + \frac{w_n}{W_n} (x_n - \mu_{n-1})$$

$$\sigma_n = \sigma_{n-1} + w_n (x_n - \mu_{n-1}) (x_n - \mu_n)$$

where  $w_n$  is the weight of the actual sample  $x_n$  in iteration  $n$  of the node,  $\mu_n$  denotes the mean of sample  $x_n$  and  $\sigma_n$  is the respective standard deviation.

A second important issue of changing online classification to regression behavior is the proper choice of a prediction model for each leaf node. While each node in higher levels of the tree directs an incoming sample down the tree, the leaf node has to deliver a proper prediction of the continuous output as well as the respective confidence. The forms of prediction functions of each leaf node can vary *e.g.* polynomial functions as proposed in Criminisi *et al.* (2011). It makes sense to adapt that function to the expected target function. In general a simple constant model related to  $\mu_n$  is sufficient, if the need for higher forest complexity is less important. The regression forest's overall posterior is then calculated by the median of all individual tree posteriors.

Another important aspect is the proper choice of an objective function for maximising information when splitting intermediate nodes during online training of the forest/tree. In the offline case, the optimal split can be determined during the global training step.

In contrast to the offline case an early split decision has to be made during the samples arriving. The criteria for performing a split are similar to the ones in Saffari's approach (2009). In particular a node only splits if i) a minimum number of samples has already passed the node (ensures statistical significance), ii) the depth of the tree has not exceeded the predetermined maximum model complexity (ensures a final size model), and iii) the minimal information gain required by a split is reached (avoids early growing).

The objective function for the information gain is the sum of all  $n$  label's differences between actual ( $y$ ) and the node's labels ( $\mu$ ), for all parent ( $P$ ) and left/right (L, R) child (C) nodes.

$$\sum_{i=1}^{n_p} (y_i - \mu_p)^2 - \sum_{c=L,R} \sum_{i=1}^{n_c} (y_i - \mu_c)^2$$

Another challenge introduced by stream processing tasks is the unknown range of feature dimensions. Extremely random forest approaches select random dimensions and random thresholds of the samples for decision candidates (weak learners). In the offline case, feature ranges can be trivially calculated from the whole dataset, but in the online case this is usually not possible. It is reasonable to carefully select and adapt the feature dimension's boundaries, to avoid useless calculations and a waste of memory. We allow changing the feature dimension boundaries during the growth of the tree, starting with very conservative feature ranges and extending them dynamically to the highest/lowest values seen so far.

One problem introduced by this strategy is the fact, that it changes the tree configuration over time and the actual splitting criterions might no longer be optimal. Thus it is necessary to remove certain trees from time to time. Additionally we allow a novel tree to learn focusing optimally on novel examples with updated tree configura-

tion. Therefore we introduce a *jig-saw criterion*, calculated on the out-of-bag error value calculation (Oza and Russell, 2001). Thus we also are able to react on slightly changing boundary constraints of the problem (e.g. changing geological conditions when drilling down to the earth). It is also possible to derive some measure for the overall *prediction quality* of the forest by averaging or taking the mean of the individual trees' out-of-bag errors already calculated for the *jig-saw criterion* introduced above.

### Detection of unusualness of vegetation parameters

The detection of unusualness of the vegetation index and short-wave infrared band with above described methods follows a common scheme. First of all the NDVI and SWIR are modelled from the satellite data and the expected time series for a certain year is forecasted using the FTSA, ARIMA or ORRF method. Therefore the 80, 95 and 99% prediction intervals, as already mentioned in the respective

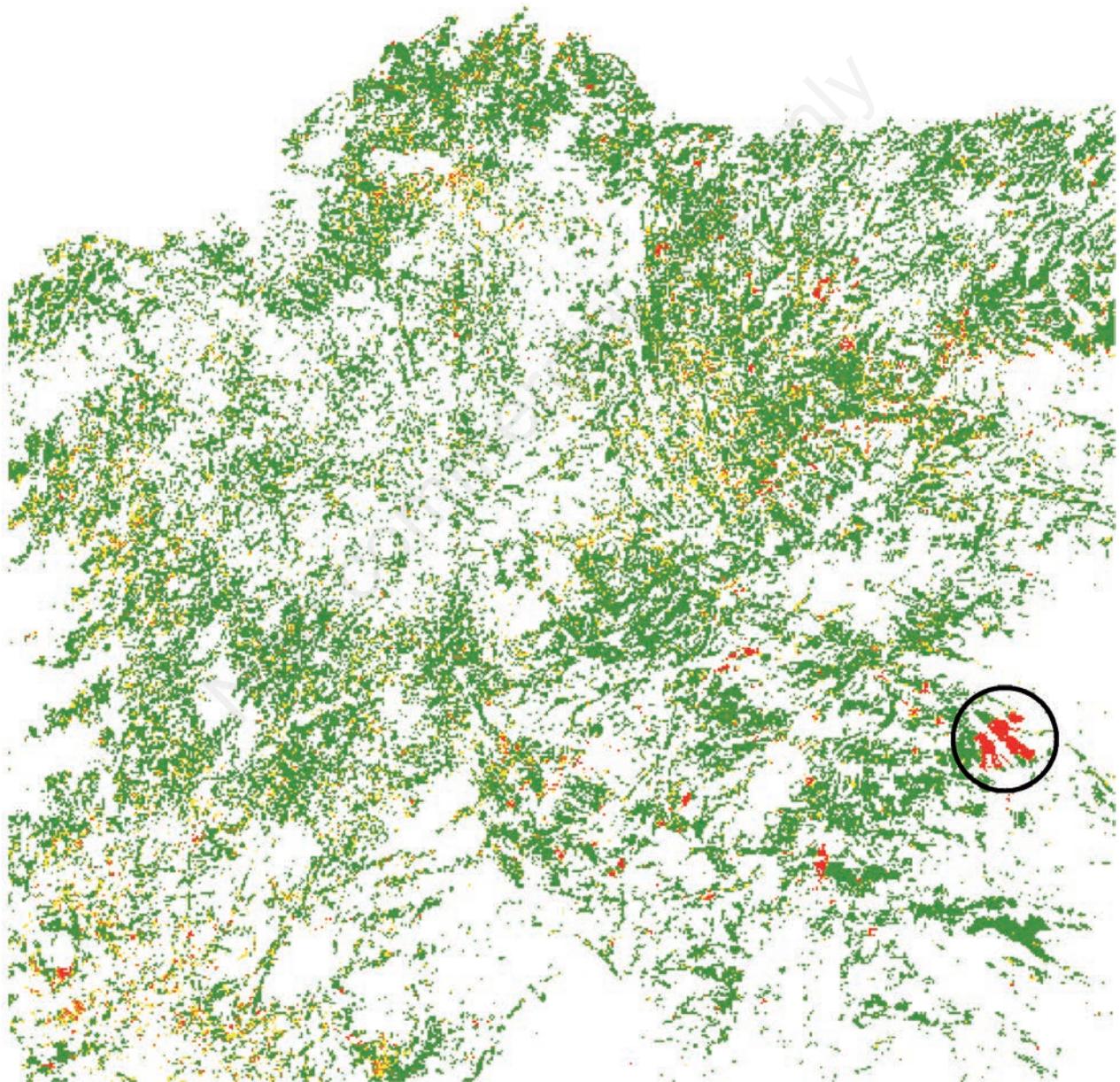


Figure 1. Region with forest fire (circled) in the south-east of Spain study site in 2012.

method's descriptions above, are calculated for the FTSA and ARIMA method. Since ORRF is a non-parametric method, it turns out, that a much simpler estimation of the prediction intervals as the one proposed, leads to more stable and better results. In particular, we validate the ORRF's predictions on all the training/update data of the preceding year by comparing all real training values to the ORRF's model prediction for all pixels and time points. The upper and lower boundaries for including 80, 95 and 99% of the training values (with respect to the ORRF prediction value at each observation time point in the year) can then be treated as estimation for the prediction boundaries

in the actual year of observation. More formal, let  $\hat{X}_{n+h/n}$  be the h-step forecast and  $r = X_{n+h} - \hat{X}_{n+h/n}$  the estimation error, then the empirical  $\alpha$ -quantiles of the prediction error  $q_{r,\alpha}$  are calculated. Hence, the lower and upper boundaries of the  $(1-\alpha)$ -prediction interval follow the equations  $\hat{X}_{n+h/n} - q_{r,\frac{\alpha}{2}}$  and  $\hat{X}_{n+h/n} + q_{r,1-\frac{\alpha}{2}}$  approximately.

Subsequently, for all methods unusualness is detected, if the original time series is outside the prediction boundaries. The forecast models are done solely spatial (FTSA and ORRF), solely temporal (FTSA and ARIMA) and combined spatio-temporal (FTSA and ORRF).



Figure 2. Pixel (squared) within that region in south-east of Spain study site in 2012.

## Results

The results of above methods provide for each pixel for each observation the probability that the reflectance in the SWIR spectral region or NDVI has changed significantly over time. To reduce noise

which is inherent in the time series measurements (such as *e.g.* caused by locally varying atmospheric conditions during image acquisition, sensor noise, *etc.*) change detection is not based on single observations but on the aggregation of three observations, as the forest changes to be observed are in general long term processes (*e.g.* crown cover recovery needs several years after a storm dam-

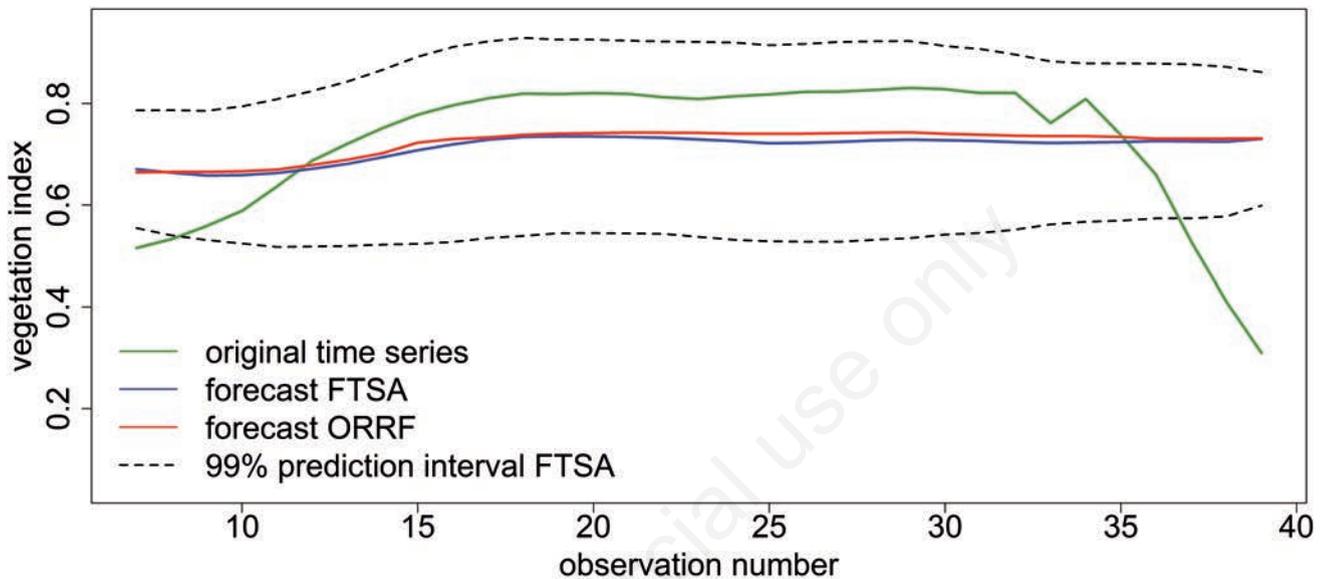


Figure 3. Spatial forecast for year 2012 using the methods functional time series analysis (blue) and online random regression forests (red) in comparison with the observed time series for the year 2012 (green).

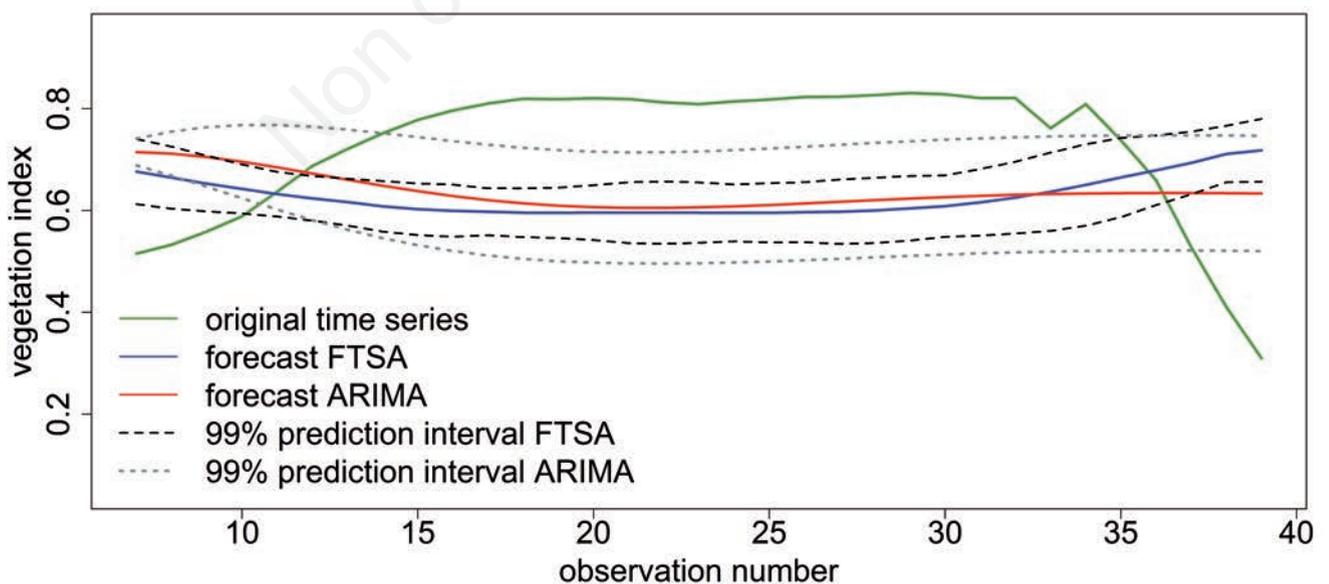


Figure 4. Temporal forecast for year 2012 using the methods functional time series analysis (blue) and autoregressive integrated moving average (red) in comparison with the observed time series for the year 2012 (green).

age). A pixel was therefore classified as changed only in case of at least three consecutive results that the respective pixel is an outlier at the 95% prediction interval.

Figure 1 shows Spain study site where forest fire happened in 2012 (red colored pixels in the black cycle). Pixels with green color mean that no unusualness was detected with the forecasting methods whereby yellow, orange and red mean that the original vegetation

index time series was significantly different from unusualness. From that region we consider one certain pixel corresponding to the region indicated by the yellow bounding box in Figure 2.

For this certain pixel, Figures 3-5 show the forecasts of each of the three model types (spatial, temporal and spatio-temporal) using the corresponding methods (FTSA, ARIMA, ORRF) for 2012. Additionally, the prediction intervals are plotted as dotted lines. Unusualness hap-

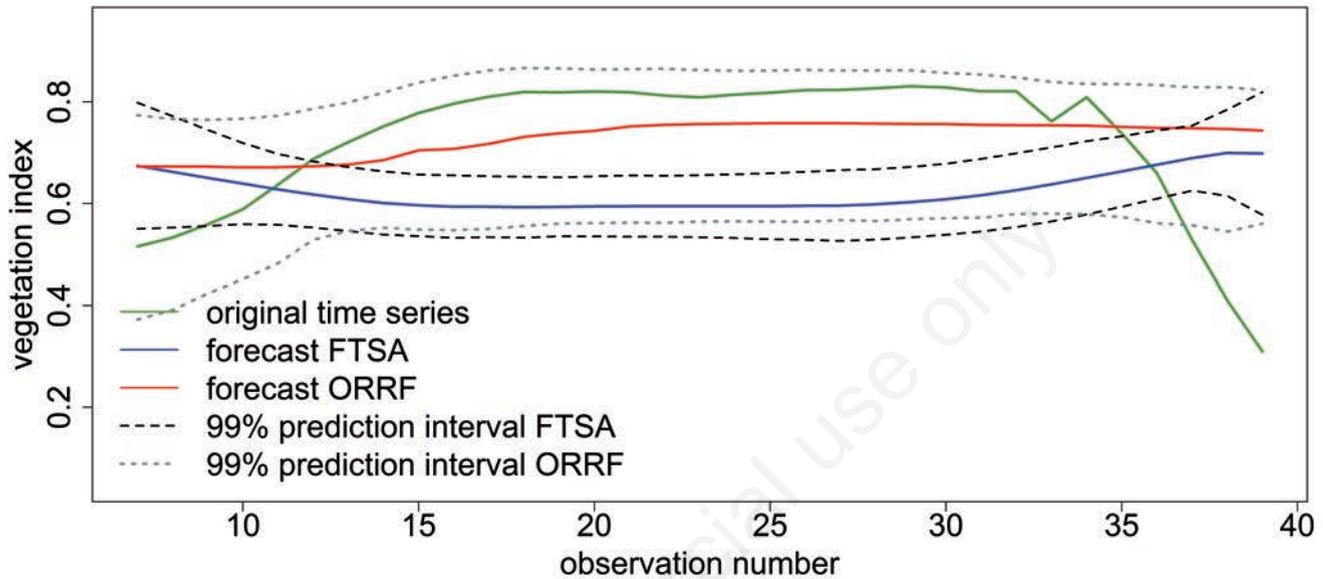


Figure 5. Spatio-temporal forecast for year 2012 using the methods functional time series analysis (blue) and online random regression forests (red) in comparison with the observed time series for the year 2012 (green).

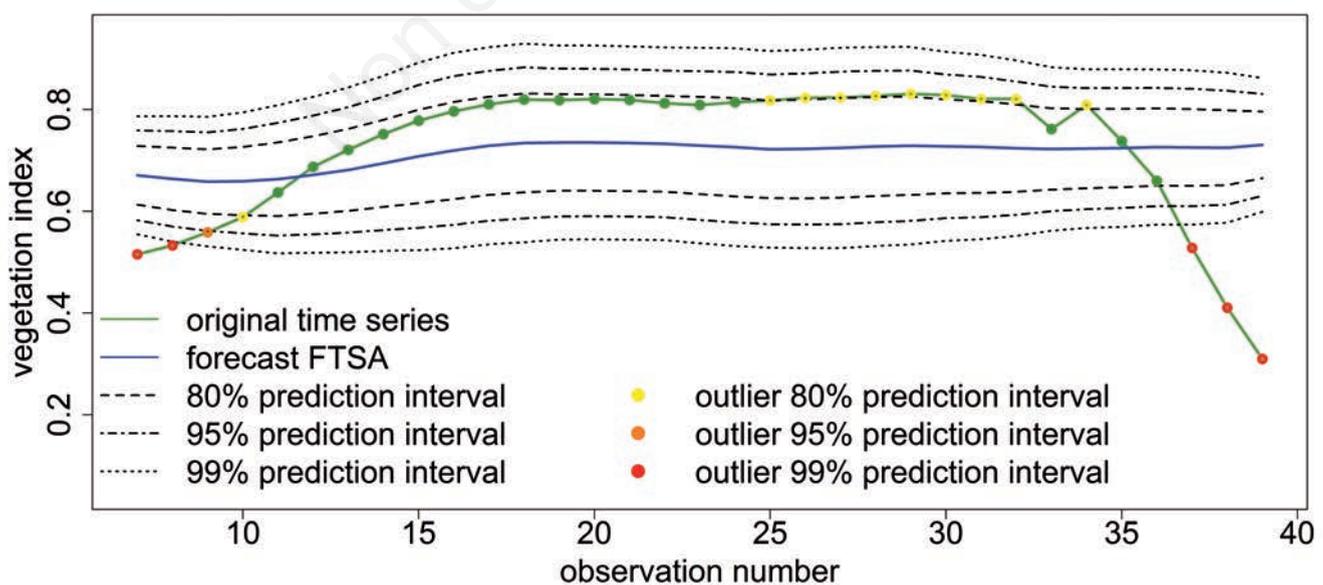


Figure 6. Functional time series analysis forecast (blue line) with original time series (green line) and the 80, 95 and 99% (yellow, orange and red dots, respectively) prediction intervals.

pened if the original time series is above the upper or below the lower prediction boundaries. As the time series data are pre-processed for noise reduction and gap-filling over time (*e.g.* gap-filling in case of cloud cover at the satellite image acquisition) by Savitzky-Golay filtering (Savitzky and Golay, 1964), the change of the signal is not abrupt. We can see that phenomenon in all three plots starting at about observation number 25 when forest fire occurred. The difference between

FTSA and ORRF in the case of the spatio-temporal forecast (Figure 5) is caused by the different adaptation sensitivities of the forecast methods. Figure 6 shows the original time series and the FTSA prediction with the 80, 95 and 99% prediction intervals. The temporal forecast and the three prediction intervals based on ARIMA are shown in Figure 7. The spatio-temporal forecast and its prediction intervals based on ORRF can be seen in Figure 8.

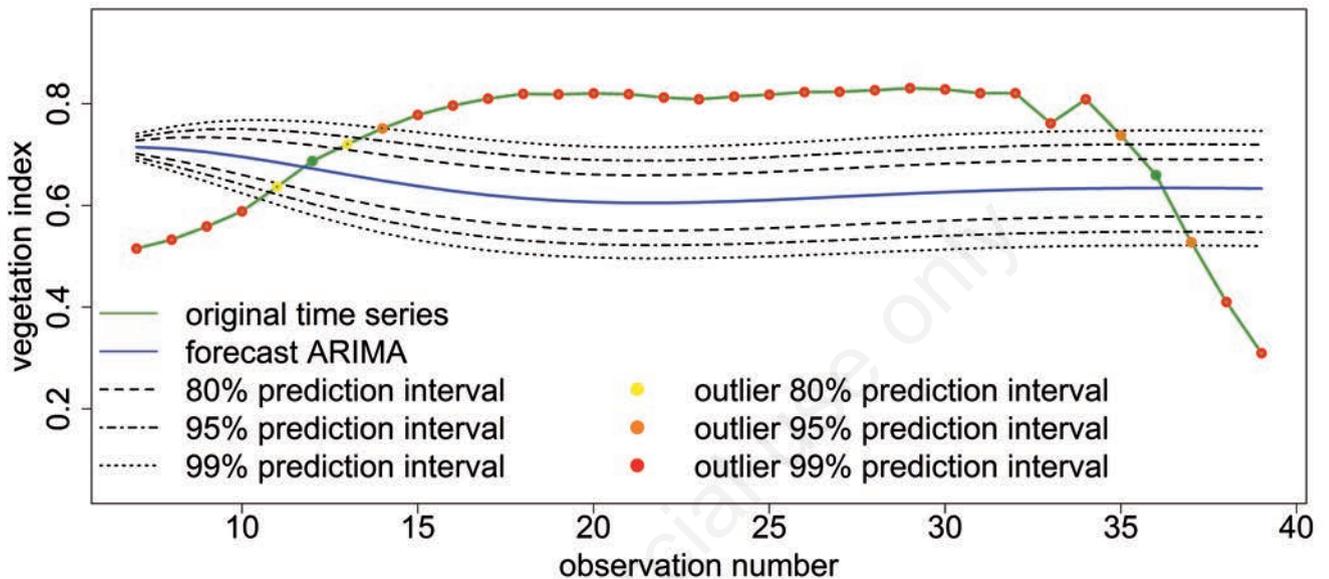


Figure 7. Autoregressive integrated moving average forecast (blue line) with original time series (green line) and the 80, 95 and 99% (yellow, orange and red dots, respectively) prediction intervals.

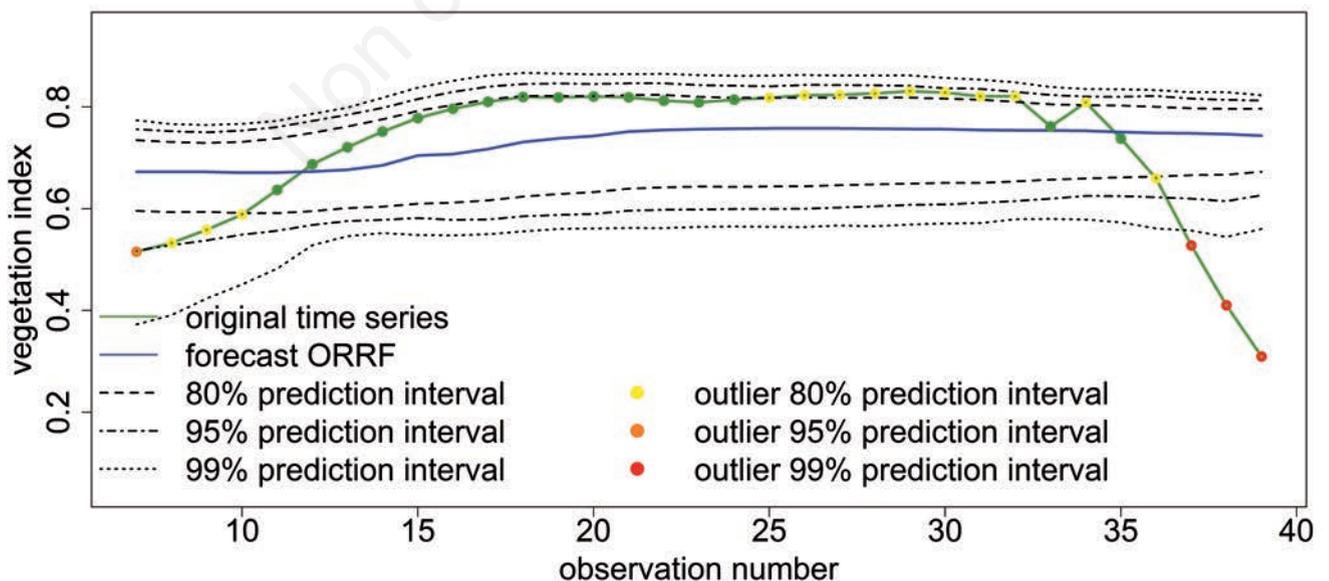


Figure 8. Online random regression forests forecast (blue line) with original time series (green line) and the 80, 95 and 99% (yellow, orange and red dots, respectively) prediction intervals.

## Qualitative and quantitative accuracy assessment

A qualitative evaluation of the monitoring results was performed by comparing the detected changes with time series of higher resolution Landsat thematic mapper (TM) satellite imagery with spatial resolution of 30 m by 30 m per pixel. These comparisons have shown that the derived change detection results correspond well with the changes visible in the TM imagery.

For quantitative evaluation of the different monitoring methods in addition a sampling based approach was applied. As only small parts of the study areas are affected by damages, stratification was performed as a first step. As basis for the stratification, a forest loss map generated by Hansen *et al.* (2013) who mapped forest loss at the Landsat pixel scale was used. These maps were aggregated according to the MODIS spatial resolution (*e.g.* from 30m by 30m pixels to the MODIS resolution). As the qualitative evaluation has shown that the change detection methods are not only capable of indicating changes at the pixel level but also at the sub-pixel level, stratification was applied for the strata *no forest loss*, *forest loss below 20%* and *forest loss above 20%*. Within the stratum *no forest loss* and the stratum *forest loss above 20%* 25 grid cells were selected in both strata randomly. As in the German study site only small areas are affected by storm-damages, the stratum *forest loss above 20%* covered only 23 MODIS pixels. For this stratum, no random sampling was therefore performed but all 23 grid cells were selected for the accuracy assessment. For each of the selected MODIS grid cells, the percentage of forest loss mapped by Hansen *et al.* (2013) was calculated by aggregation. Then, the grid cells were visually checked on the basis of Landsat TM time-series data and the MODIS time series and the Hansen *et al.* (2013) estimate was revised in cases of clear deviations. Such deviations occurred *e.g.*

in case of misclassification in the Hansen *et al.* (2013) map. For both study areas then contingency tables were derived for all selected change detection methods (Table 1). As a measure of the achieved accuracy, the percentage of grid cells, which were classified correct was calculated and is given in percent.

A main factor influencing the achieved accuracy is the nomenclature definition of the category *forest loss*. In the current study we selected a very ambitious threshold for the definition of this category with forest loss indicated already in case that only 20% of the area covered by one pixel are affected *e.g.* by a storm damage or forest fire. The current methods are therefore performing change detection at the sub-pixel level. In the following, the results are shortly discussed for each of the monitoring methods.

## Functional time series analysis

Exceptional high accuracy was achieved for both study areas with the combined spatio-temporal as well as with the temporal approach. No significant errors could be identified. For the spatial approach, the achieved accuracies are in the range of current operational monitoring methods with overall accuracies between 72 and 92 percent. Processing time is a drawback for functional time series analysis methods, which requires processing time of about 8 hours per study area on a usual PC (i7 processor). Software is currently implemented in the R packages *fts* (Shang, 2013) and *forecast* (Hyndman *et al.*, 2008), where each pixel is treated separately. For large area applicability a pre-requisite would be a significant optimisation of the implementation, *e.g.* via graphics processing unit (GPU) processing.

**Table 1.** Evaluation results as contingency tables with reference category in columns, classified category in rows, and achieved accuracy given in percent.

		Spain		Germany	
		Change	No change	Change	No change
FTSA - Temporal	Change	24	1	20	2
	No change	1	24	3	23
	AA (%)		96.0		89.6
FTSA - Spatial	Change	15	1	22	3
	No change	10	24	1	22
	AA (%)		78.0		91.7
FTSA - Spatio-temporal	Change	22	0	20	0
	No change	3	25	3	25
	AA (%)		94.0		93.8
ARIMA - Temporal	Change	18	0	13	0
	No change	7	25	10	25
	AA (%)		86.0		79.2
ORRF - Spatial	Change	11	1	22	3
	No change	14	24	1	22
	AA (%)		70.0		91.7
ORRF - Spatio-temporal	Change	14	1	21	1
	No change	11	24	2	24
	AA (%)		76.0		93.8

FTSA, functional time series analysis; ARIMA, autoregressive integrated moving average; ORRF, online random regression forests; AA, achieved accuracy.

## Autoregressive integrated moving average

The qualitative evaluation by comparison of the monitoring result with the Landsat time series showed that the ARIMA method is less sensitive to sub-pixel changes compared to the functional time series analysis approach. However, the quantitative accuracies with 86% in the Spain study area and 79% in the German study area are high as we perform change detection at the sub-pixel level. In case that the nomenclature specification of the *forest loss* category is defined with a minimum change of 50% instead of 20%, overall accuracies for both study sites are above 96%. Main drawback for operational application is the processing time of about 3 hours per study area on a usual PC (i7 processor). The R-package *forecast* is used (Hyndman *et al.*, 2008). As with the functional time series analysis method, significant optimisation with (GPU) processing would be required for operational application.

## Online random regression forests

Accuracy of the spatial approach is 70% for the German and 92% for the Spain study site. Qualitative evaluation has shown a local bias of the change detection result (overestimation of changes in the southern area of the Spain study site). For the spatio-temporal approach, high accuracies were achieved with overall accuracies of 76% in the Spain and 94% in the German study site. As with the ARIMA method, accuracy increases significantly in case of changed nomenclature specification. In case that the nomenclature specification of the *forest loss* category is defined with a minimum change of 50% instead of 20%, overall accuracies for both study sites are above 94%. A main advantage of the online regression forest method is that the required processing time is low, in the range of several seconds per study area on a usual PC. This method allows therefore near real time processing.

## Discussion

Exceptional high accuracies were achieved with the temporal and spatio-temporal functional time series analysis methods, which however need significant optimisation of the software implementation to allow large area applicability (GPU processing). It is expected that near real time monitoring based on upcoming Sentinel2 satellite data can be achieved in this case at the local to regional level. Implementation can be recommended for such applications but also for applications based on medium resolution satellite data, *e.g.* MODIS or upcoming Sentinel3 time series data. For near real time, continuous forest monitoring at the national to continental level based on upcoming Sentinel2 satellite data the implementation of spatio-temporal online random forest method is recommended, however because of sensitivity issues, monitoring should focus on severe changes only (*e.g.* monitoring at the pixel level with change fraction above 50%).

## Conclusions

The results demonstrate the applicability of dense time series of remote sensing data for large area forest monitoring with spatial-explicit wall-to-wall coverage. Compared to more *simple* methods, which are often based on bi-temporal comparison of satellite

imagery, the methods applied in this publication are based on wall-to-wall, fully consistent time series and allow fully automatic processing. This automation is a specific requirement for future forest monitoring applications based on satellite imagery with high spatial and high temporal resolution such as *e.g.* Sentinel2 time series.

## References

- Breiman L, 2001. Random forests. *Machine Learn* 45:5-32.
- Criminisi A, Shotton J, Konukoglu E, 2011. Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Available from: [research.microsoft.com/pubs/155552/decisionForests\\_MSR\\_TR\\_2011\\_114.pdf](http://research.microsoft.com/pubs/155552/decisionForests_MSR_TR_2011_114.pdf)
- Erbas B, Hyndman RJ, Gertig DM, 2007. Forecasting age-specific breast cancer mortality using functional data models. *Stat Med* 26:458-70.
- European Commission, 2010. Green paper on forest protection and information in the EU: preparing forests for climate change. SEC(2010)163 final. European Commission ed., Brussels, Belgium.
- European Commission, 2013. A new EU forest strategy: for forests and the forest-based sector. SEC(2013)659 final. European Commission ed., Brussels, Belgium.
- Freund Y, Schapire RE, 1997. Decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55:119-39.
- Hall P, Hosseini-Nasab M, 2006. On properties of functional principal components analysis. *J Roy Stat Soc B* 68:109-26.
- Han P, Wang PX, Zhang SY, Zhu DH, 2010. Drought forecasting based on the remote sensing data using ARIMA models. *Math Comput Model* 51:1398-403.
- Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Townshend JRG, 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342:850-3.
- Hyndman RJ, Khandakar Y, 2008. Automatic time series forecasting: the forecast package for R. *J Stat Softw* 27:1-22.
- Hyndman RJ, Koehler AB, Ord JK, Snyder RD, 2008. Forecasting with exponential smoothing. The state space approach. Springer Verlag, Berlin-Heidelberg, Germany.
- Hyndman RJ, Koehler AB, Snyder RD, Grose S, 2002. A state space framework for automatic forecasting using exponential smoothing methods. *Int J Forecasting* 18:439-54.
- Hyndman RJ, Shang H L, 2009. Forecasting functional time series (with discussions). *J Korean Stat Soc* 38:199-221.
- Jiang B, Liang S, Wang J, Xiao Z, 2010. Modeling MODIS LAI time series using three statistical methods. *Remote Sens Environ* 114:1432-44.
- Kriegler FJ, Malila WA, Nalepka RF, Richardson W, 1969. Preprocessing transformations and their effects on multispectral recognition. In: *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, Institute of Science and Technology of the University of Michigan. University of Michigan ed., Ann Arbor, MI, USA, pp 97-131.
- Oza N, Russell S, 2001. Online bagging and boosting. In: *Proceedings of Artificial Intelligence and Statistics*. Morgan Kaufmann Publ., Key West, FL, USA, pp 105-12.
- Saffari A, Leistner C, Santner J, Godec M, Bischof H, 2009. On-line random forests. In: *Proceedings of the 3rd IEEE ICCV Workshop*



- on Online Computer Vision. IEEE Publ., Kyoto, Japan, pp 1393-400.
- Savitzky A, Golay MJE, 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627-39.
- Shang HL, 2013. ftsa: an R package for analyzing functional time series. *R Journal* 5:64-72.
- Shang HL, Hyndman RJ, 2011. Nonparametric time series forecasting with dynamic updating. *Math Comput Simulat* 81:1310-24.
- Shen H, Huang JZ, 2008. Interday forecasting and intraday updating of call center arrivals. *M&SOM-Manuf Serv Op* 10:391-410.
- Shumway RH, 2006. *Time series analysis and its applications*. 2<sup>nd</sup> ed. Springer Verlag, Berlin-Heidelberg, Germany.
- Vapnik VN, 2000. *The nature of statistical learning theory*. 2<sup>nd</sup> ed. Springer Verlag, Berlin-Heidelberg, Germany.

Non commercial use only