Terra-i

A methodology for near real-time monitoring of habitat change at continental scales using MODIS-NDVI and TRMM

Louis Reymondin, Andrew Jarvis, Andres Perez-Uribe, Jerry Touval, Karolina Argote, Alejandro Coca, Julien Rebetez, Edward Guevara, Mark Mulligan

December 2012

Table of Content

Table of figures

1 Introduction

Tropical forests, cover about 47\% of the world's forests which is equivalent to about 1830 million hectares of which about 830 million are located in Latin America, 500 million in Africa and 260 million in Asia (FAO, 2000).Over two-thirds of these forests are located in only three countries: Brazil, the Democratic Republic of Congo and Indonesia (FAO, 2000). The term 'tropical forest' includes different types of forest ecosystem, from the tropical dry broadleaf forests, composed of scattered trees growing under a tropical savannah climate, to the humid and dense rainforest including the relatively cold higher-altitude cloud forests (FAO, 2005). Global forests are essential for conservation of biodiversity but also for the ecosystem services that they supply to humanity (Gibson, Holtz, Tansey, Whitelaw, & Hassan, 2005). Most of the planet's terrestrial biodiversity is found in forests, and at least 50% of it is in tropical forests (Alfonso, Dallmeier, Granek, & Raven, 2001). In addition, degradation of forest cover has direct effects on global warming (Bonan, 2008)because of the role of tropical forests in global carbon budgets. Deforestation is quantitatively most important change to tropical forests in the last years (FAO, 2005; M. C. Hansen et al., 2008) and very few tropical zones showing forest recovery or reforestation. Pantropically, agriculture, cattle ranching, road construction, and unsustainable logging are the main causes of tropical deforestation (Geist & Lambin, 2002; Sierra, 2001). Conventional logging is financially very lucrative practice and has the most severe impact in the environment (Pearce, Putz, & Vanclay, 2002).

Humans have had profound impacts on the Earth's natural ecosystems, contributing to their loss and degradation through a number of activities which negatively affect biodiversity, water resources and the global climate system (Nelson, 2005). However, in many parts of the world the scale and pattern of habitat loss goes unmonitored or infrequently and inconsistently monitored so that conservation and sustainable development decisions to manage the impacts of land use and cover change on ecosystem services have to be made without a complete understanding of the current state and recent history of land cover and use change For instance, the Food and Agriculture Organization recently released several studies based on census data from each individual country (FAO, 1996, 2000, 2005) as well as a global rate estimation based on remote sensing data (Gerrand, Lindquist, & D'Annunzio, 2011). Nevertheless, such studies are contradictory, and it is difficult

to assess which are the most accurate. Indeed, the estimation from the Forest Resources Assessment 2010 (FAO, 2010)shows a decrease of about 3 million hectares globally between the annual rate estimated for the period 1990-2000 and the one estimated for the period 2000-2005. The (Gerrand et al., 2011)released one year later show the opposite figure with an increase in global forest cover of 1.8 million hectares for the same periods of time. This demonstrates well the lack of harmonization of methods and implications for uncertainty of data on vegetation conversion in the tropics.

Decision makers at multiple scales from (local to national to regional) need information on land-cover and use change that is as current as possible in order to prioritise interventions where this change threatens protected areas, is illegal or places important ecosystem services at significant risk. With a number of important policy instruments for halting deforestation being developed as a response to the global climate crisis e.g. REDD+ (Parker, Mitchell, Trivedi, & Mardas, 2008), monitoring of land cover and use, and particularly deforestation, is now high on environmental policy and management agendas. Near real-time information is required to avoid, for example, belated reactions to increased deforestation as observed in Brazil in the final quarter of 2007, when deforestation in the legal Brazilian Amazon sharply increased unexpectedly, (INPE, 2009) and also needs to be of sufficient spatial resolution to detect the loss of modest sized plots (Mayaux, Eva, Brink, Achard, & Belward, 2008). Existing systems are often annual or decadal and aggregated to national level statistics or are high spatial resolution but with only a small area coverage. Near real time products of sufficient spatial resolution to monitor individual deforestation events but which are continental in scale necessary for supporting policy at the regional, national and international scales.

Remote sensing of the Earth from satellites has, over the years, made great progress in monitoring land-cover, although many of the advances have been in the realm of increased spatial resolution and the quality of the land-cover mapping (greater precision, detail in number of land cover classes etc.). These advances have neglected somewhat the temporal dimension. Though there are annual time-series products for basic properties such as NDVI, there are few attempts to convert these to time series of land cover change: most land cover products being updated every 5-10 years-often with different methods and sensors used which makes comparison challenging and error-prone (Mayaux et al., 2005). Many global assessments use satellite data of moderate spatial resolutions

but high temporal resolution, such from the NASA MODIS sensor on the Terra and Aqua satellite platforms which provides a normalised difference vegetation index (NDVI) data product (M. C. Hansen et al., 2008; M. Hansen et al., 2003; Mulligan, 2008). This product measures the greenness of the earth surface every 16 days with a 250m spatial resolution, covering the entire globe (Huete, Justice, & van Leeuwen, 1999). The scale of data produced by such systems generates computer science challenges in both analysing and processing such datasets.

Within this context, the objective of this study is to create a pan-tropical near-real time monitoring system for natural ecosystems based on machine learning algorithms. The approach proposed uses vegetation greenness of the land surface (and its temporal variation) which can be used as a surrogate for land-cover (Shao & Lunetta, 2010; Wang & Tenhunen, 2004; Xiaa, Ruib, Binga, & Qingxia, 2008). Changes in the greenness of a given pixel are considered to be a function of meteorological factors (rainfall, flooding events), climatic factors (soil moisture changes over seasons and inter-annually) and anthropogenic factors (human activity impacting on the vegetation, e.g. deforestation). If these changes in vegetation greenness in relation to meteorological and climatic factors can be accurately modelled, possible anthropogenic impacts on vegetation could therefore be isolated from the meteorological and climatic impacts on vegetation greenness in near-real time and deforestation events thus flagged.

This study is focused on the implementation of a deforestation monitoring systems based on neural network models of the relationship between vegetation greenness and antecedent rainfall. This model is tested in four case studies throughout the tropics along climatic and ecosystem gradients (e.g. Attitudinally from Andes to Amazon and hydrologically from rainforest through savannah).The system is implemented across all Latin America for the period 2004-present and the case studies draw from the significant dataset produced. This implementation was first made public under the name of "Terra-i, an eye on habitat change" during the Rio +20 events. Near-real time, automated monitoring systems such as the model developed in this study have great potential in providing more timely information on habitat conversion and ensuring that conservation and development actions and policies can react and adapt to control emerging fronts of habitat conversion.

Once the methodology was implemented and assessed on relatively small study areas around the tropics, the next step has been to implement it at continental scale from Mexico to the Patagonia. This chapter presents the results of such application. The implementation of the system in Latin-America was first made public under the name of "Terra-i, an eye on habitat change" during the Rio +20 events[1].

To implement Terra-i over all Latin-America involved first to run the system on 35 MODIS tiles, as shown on Figure 1, and to calibrate the results using manually classified Landsat imagery together with satellite data visualization software such as Google Earth.

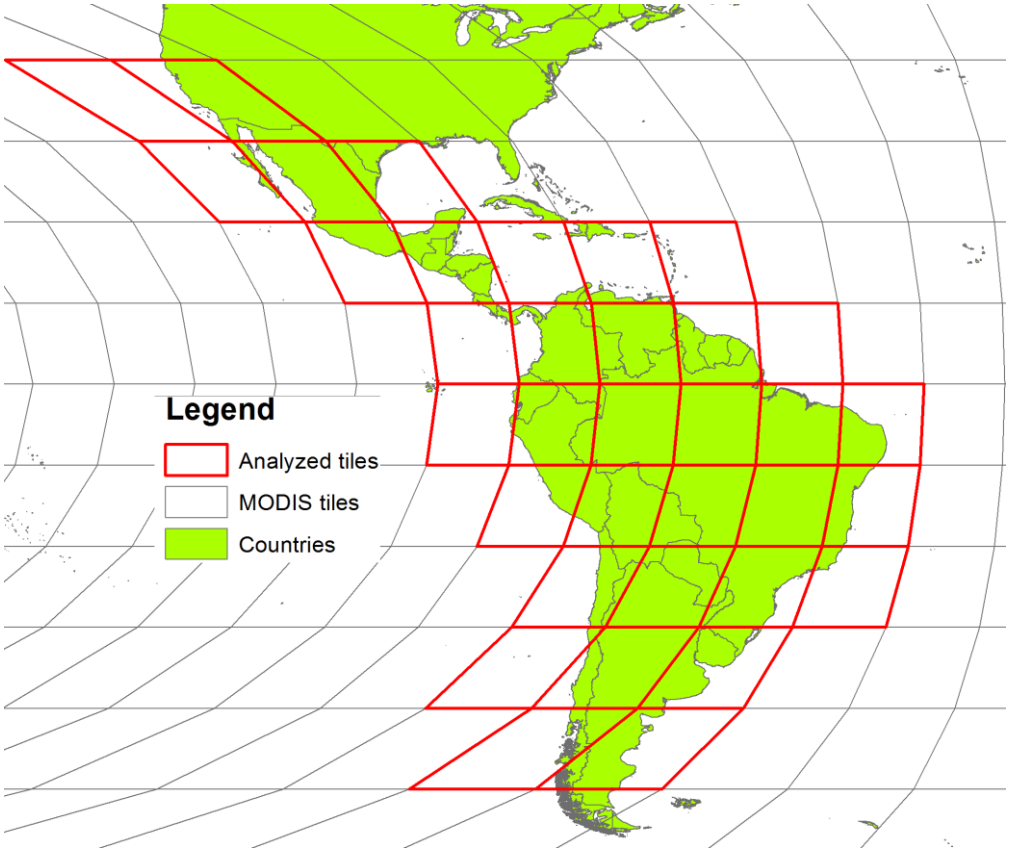

Figure 1, Analyzed MODIS tiles

Terra-i has been operational for the whole of Latin America (Mexico to Argentina) since December 2011 and was officially launched in June 2012. Since then, the database of vegetation change

[1] Thus, the results and achievements of Terra-i were reported around the world as "A New Satellite Tool Tracks Deforestation "- The New York Times (USA), "Terra-i: the most precise real-time deforestation tracker yet "- arstechnica.com (USA), "South America deforestation in one click" - Le Temps (Switzerland), "Terra-i, deforestation under the lens" -eleconomista.mx (Mexico), "Deforestation advances relentlessly in Caquetá "- Portfolio (Colombia).

detections has been updated every 3 months based on the newest available MODIS images. All Terra-i results are available on a web portal www.terra-i.org.

In order to process such large amount of information, a processing framework was developed. It compiles all the tools needed to automatically (1) pre-process the data (download the data from MODIS repository, extract the useful information and re-project the geo-data…), (2) run Terra-i processes in a distributed manner and (3) post-process the model outputs (backup the data, make the data available on the web-portal…). This workflow is described in the annexes *Terra-i short workflow* and *Terra-i extended workflow*.

In Latin-America, few countries have implemented land cover monitoring system. Indeed, the Brazilian PRODES and DETER systems are the only official monitoring systems that are updated regularly. PRODES has provided annual mapping of deforestation events within the Brazilian Amazon since 1988. Moreover, the Brazilian DETER system has provided twice monthly updates since 2004. Imazon's Forest Transparency Initiative also provides regularly updated deforestation maps since 2007. The algorithm they implemented to detect forest lost in the Brazilian Amazon is called SAD. It is important to highlight that these monitoring systems are only operational in the Brazilian Amazon and are not monitoring other eco-regions such as the Cerrado savannahs. The Brazilian territory is about 8,514,876.6 km^2 in area and occupies almost half (47%) of the terrestrial surface of South America. There are about 4,617,915 km^2 of forests in Brazil, which is equivalent to 54.2% of the total area of country. The most widespread forest ecosystem present in Brazil is the Amazon biome (MMA, PNUMA, UNESCO 2007).

Since 1978, the National Institute for Space Research - INPE has produced annual estimates of average deforested areas as part of the Project of Deforestation Estimation in the Amazon (PRODES).

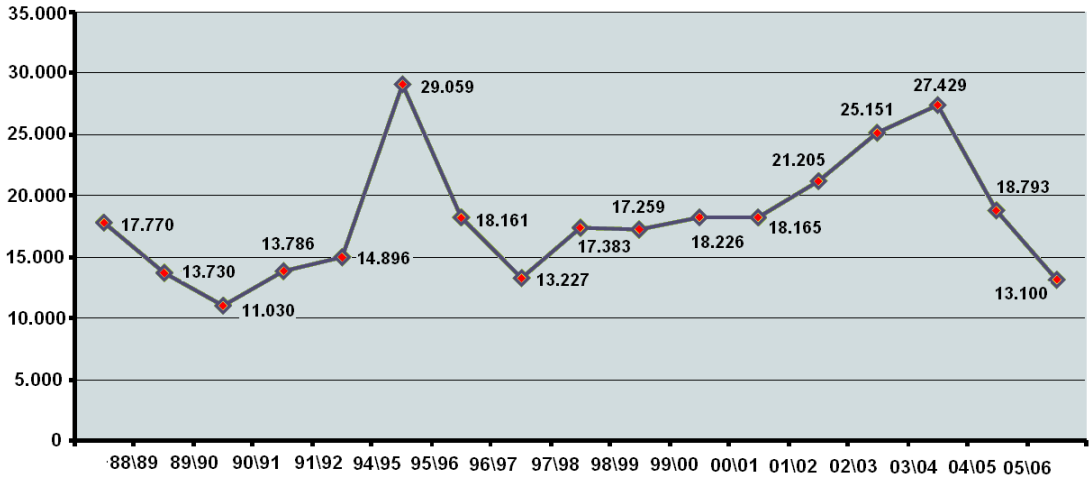

Figure 2. Deforestation in the Brazilian legal Amazon 1988 to 2006 (km^2/year). Source: Data from the "Instituto de Investigaciones Espaciales" (INPE) - PRODES, 2007.

Based on data released by the INPE (Figure 2), deforestation in Legal Amazonia underwent a period of decline, reaching its lowest point during the years 1990/1991, after which it accelerated and grew at an alarming rate to reach its highest historical value (29,059 km^2) during the period 1994/1995. In the subsequent years the annual deforestation rate slowed, gradually increased again until 2004, and since then has fallen sharply. The recent decline can be attributed in large part to the actions taken by the Brazilian government since 2004, within the Action Plan for Prevention and Control of Deforestation in the Legal Amazon (MMA, PNUMA, UNESCO 2007). The System of Deforestation Detection in Real Time (DETER) is another tool developed by INPE for the calculation of deforested areas. This system integrates the Plan of Action against Deforestation in the Legal Amazon and allows the production of updated maps every two weeks showing the location of deforested areas.

Outside of Brazil, countries such as Ecuador, Colombia, Guatemala, Guyana, Mexico, Paraguay and Panama are now implementing systems based on the UN-REDD Measurement Reporting and Verification (MRV) recommendations. In this context, Peru recently published online maps of deforestation at national scale for the years 2010 and 2011. To generate these maps they used Landsat Imagery together with the CLASLite software. This study concluded that between 2009 and 2010 about 108,571 ha of forest were lost in Peru and that 103,380 ha were converted between 2010 and 2011.

2 Methods

2.1 Overall approach for natural vegetation conversion detection

The aim of this study is to develop a near-real time monitoring system that data-'mines' satellite based rainfall and vegetation data to detect deviations from the usual pattern of vegetation change as a monitoring system for possible anthropogenic impacts on natural ecosystems. The model uses a multilayer Perceptron (MLP) neural network combined with Bayesian theory (Bishop, 2007; MacKay, 1992)to identify abnormal behaviour in a time-series of vegetation variation. The operationalisation of the system at the continental scale was a considerable challenge from a computer science perspective as the resolution of the MODIS sensor (250m) means that even only the Amazon basin represents more than one billion individual values for each time-frame (every 16 days). This means more than 26 billion values to process per year for the Amazon. The analysis of such a large dataset necessitated the use of data mining technologies and programming for distributed computing.

Human activities create disturbances that alter the usual cycle of vegetation greenness for an area. Disturbances can be detected when the NDVI of the landscape changes from its baseline values. The general approach adopted here was to build a forecasting model capable of predicting the evolution of vegetation greenness for a site based on the relationship between previous greenness measurements and concurrent climatic measurements at that site. Such a model is then used to predict future NDVI values (16 days ahead, based upon the current climatic conditions) and to identify anomalies or abrupt changes in vegetation where NDVI observations from MODIS differ from the model predictions. The model calculates an anomaly probability based on the difference between predicted and observed values. It was assumed that vegetation evolution (NDVI evolution at a site) is influenced by recent and seasonal rainfall trends. When major changes in the vegetation index are detected (outside of the usual pattern of seasonal evolution), it was assumed that they are due to human intervention (fire, deforestation, conversion). These events are thus flagged in near real-time as events that land managers, conservationists and policy makers should be aware of.

In order to model the evolution of the NDVI at a given point (i.e., a pixel) and time, artificial neural networks were trained using machine learning algorithms exploiting the NDVI data for a given number of preceding measurements, in order to indicate the recent NDVI trend, and the accumulated rainfall (derived from the TRMM daily rainfall product 3b42 (NASA, 1997)for the preceding 16 days in order to fit the MOD13Q1 temporal resolution.
Because the MODIS data are not noise free, NDVI time series smoothing and gap filling algorithms were needed as a first step. As the methods based on Fourier analysis and curve fitting have been shown by (Hird & McDermid, 2009) to perform better than the other methods, the Harmonic Analysis of NDVI Time Series (HANTS) algorithm (Jun, Zhongbo, & Yaoming, 2004; Roerink, Menenti, & Verhoef, 2000; Verhoef et al., 1996) has been chosen as cleaning algorithms to remove the effects of noise, atmospheric distortion and cloud (see section 2.2 for a detailed description).

Although the approach is based on the training of a forecasting model on a per-pixel basis, it is not computationally efficient to train this model on a pixel basis continentally or pan-tropically. To operationalise the model, the forecasting model was therefore trained on a land cover basis rather than a pixel basis on the assumption that within a given climatic region the NDVI response to climate should be fairly consistent within a given land cover. The approach therefore first uses an unsupervised clustering algorithm to find representative prototypes of time-series dynamics which correspond to different land cover types (Shao & Lunetta, 2010; Wang & Tenhunen, 2004; Xiaa et al., 2008). The clustering procedure is applied on MODIS-NDVI time series from the same period of time as the training dataset. This groups together the pixels which had similar trends over these years and which can be modelled by the same forecasting model. Time series from 2000 to 2004 for each cluster are then randomly selected and used as the training dataset for the modelling. As the system should be easily operational on large areas, it should be possible to perform the clustering step without having prior knowledge about the study area. Parameters such as the number of clusters present in a given area are hard to estimate computationally and as the clustering step is computationally expensive, it is ineffective to work on a trial and error basis. A series of unsupervised algorithms have therefore been selected as potential good solutions. These algorithms were implemented and applied to the study area using different distance functions. By

using the index ι (see section 2.3.3) the best result of each algorithm were selected and then compared to the MODIS land cover type (MODIS product MCD12Q1) for the years corresponding to the beginning and the end of the clustered time series in order to find the most appropriate clustering method for land cover clustering using NDVI time-series.

Once the area has been clustered, Bayesian forecasting models were trained using the training dataset previously generated on a pixel-by-pixel basis in order to capture vegetation evolution trends and relationships based on past NDVI and TRMM data. After this 'learning' step, the resulting models were used to predict the greenness of these pixels for subsequent dates. By using Bayesian Neural Networks (detailed in section 2.4.1a probability that the observed value is an anomaly is extracted based on the predicted value and some properties of the dataset found during model training.

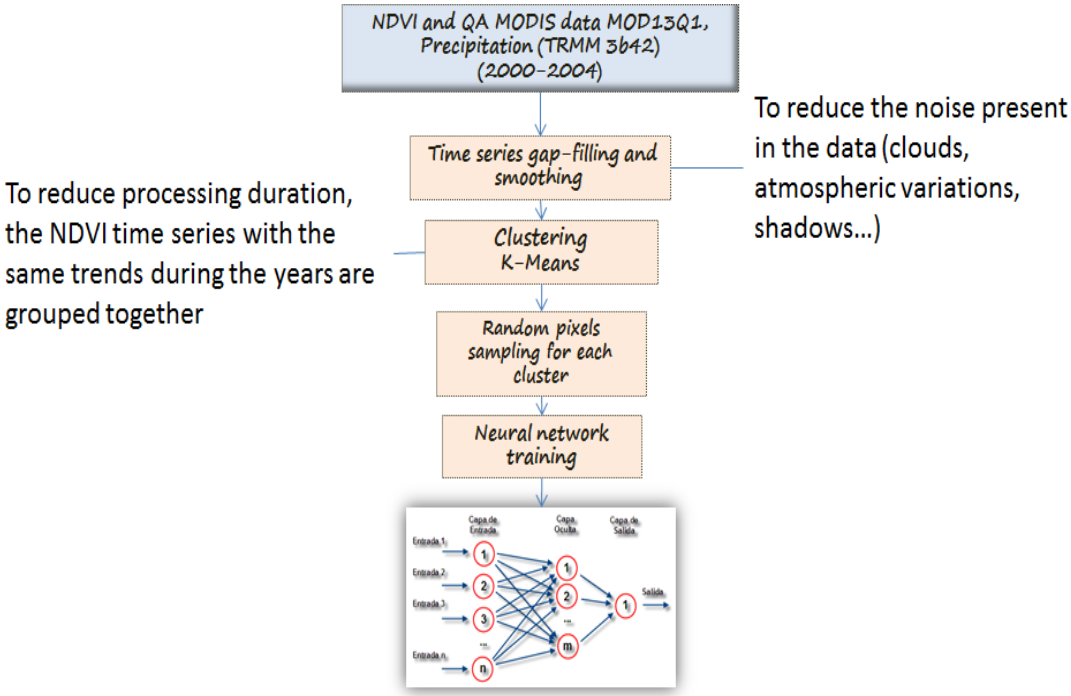

Figure 3 Visual overview of the methodology used to create the models able to detect anomalies in MODIS NDVI time series.

Whenever the probability of being an anomaly exceeds a defined threshold this event is flagged as a potential anomaly. If the anomaly flag is repeated for two consecutive 16 day periods, the system reports an anomaly at the given pixel. It is not only anthropogenic events that are the cause of

unusual long term anomalies in NDVI time series: events such as floods and droughts should also be detected by the system and separated from anthropogenic land cover change. The set of detections resulting from the previous steps should therefore be filtered and the drivers of the disturbance should be identified. To do so, the index, described in section 2.5, using the NIR and SWIR MODIS bands were used. Figure 3 and Figure 4 show a graphical description of the various steps involved from the raw NDVI data processing to the creation of the anomaly probability maps.

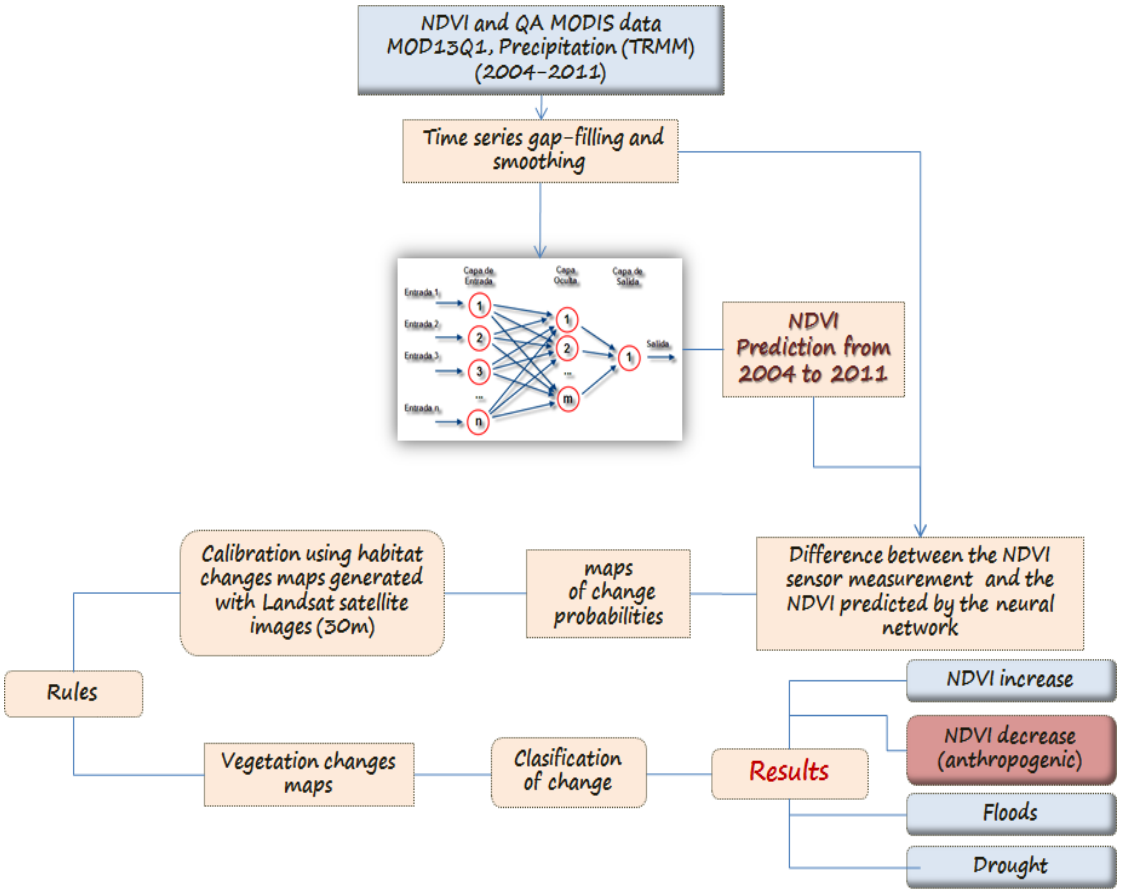

Figure 4 Visual overview of the methodology used to detect anomalies in MODIS NDVI time series.

Whilst anomalies are reported at the pixel level every 16 days, this information can then be synthesised to provide summary statistics for administrative units (municipalities, departments, and countries), critical ecosystems or protected areas using the following dataset

- Countries limit: Global Administrative Unit Layers (GAUL) by FAO.
- Protected areas: World Database of Protected Areas Annual Release 2009.

- Indigenous areas: World Database of Protected Areas Annual Release 2009 – Query Indigenous Lands.
- Ecosystems: Global Terrestrial Ecoregions by WWF Release 2003.

In order to assess the validity of the methodology, the results obtained within a series of test areas were compared with different sources such as:

1. Literature assessments over administrative regions (departments, nations)
2. Other monitoring systems (RALUCIAPA, IMAZON, PRODES)
3. Assessment from visible spectrum imagery (e.g. LANDSAT)

Given the lack of ground based data, the methodology was mostly validated using other monitoring systems such as PRODES which have been validated separately. In the following sections, each step of the methods is described in more detail.

2.2 NDVI Time-series smoothing and gap filling algorithms

2.2.1 HANTS Algorithm

The Harmonic ANalysis of Time Series(HANTS) algorithm was introduced by (Jun et al., 2004; Roerink et al., 2000; Verhoef et al., 1996) to perform the screening and removal of cloud affected observations, as wells as a temporal interpolation of the remaining observations to build gapless time-series, as an improvement of the Fourier analysis method used in previous research (Roerink et al., 2000; Verhoef et al., 1996). The method starts with a Fourier analysis on the initial data set and then only keeps the most important frequencies (defined by their amplitudes and phases). The most obvious outliers are removed in an iterative manner and the algorithm stops when the number of such outliers remaining falls below a certain threshold.

Since Fourier analysis requires the samples to be equidistant in time, Fourier analysis has to be run on the whole time series, including the poor quality points. Therefore, after running Fourier analysis, we reconstruct the time series using only the selected harmonics and we apply weighted least square fitting on it against the original time series with the good quality points having higher weights.

At each step, the outlying points (for which the difference with the cleaned series is over a certain threshold) are replaced and a new analysis has to be performed because this replacement may lead to changes in the constant term found by Fourier. We stop iterating either when no more outliers have been detected or after a maximum of 50 iterations.

The HANTS procedure is the following:

1. Calculate the fast Fourier transform. Select the important frequencies (that explain X% of the time series)
2. Fit the inverse Fourier transform to the data that are flagged as 'good' with a multivariate least square optimization
3. Flag as 'bad' all the values that have a difference with the new series greater than a given threshold
4. Repeat 1 to 4 until there is no value flagged as bad or after N steps

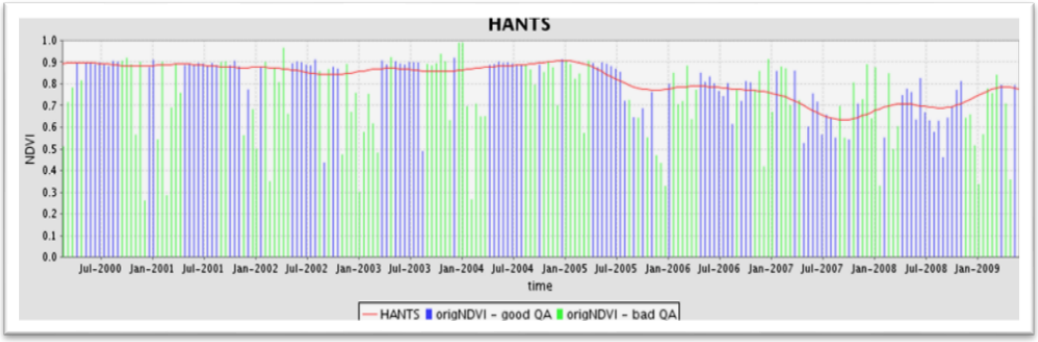

Figure 5 The HANTS algorithm over a NDVI time series. In Blue the original NDVI good quality data, in green the bad quality data and in red the HANTS resulting curve. As one can see, there has been a change in the time series trend since July 2005.

One of the interesting aspects of using HANTS is that one can select the number of harmonics to be considered. In this study, harmonics with a period of up to 6 months were used. Therefore, short-term variations are eliminated and only the long-term trends are kept which are of most interest for anomaly detection.

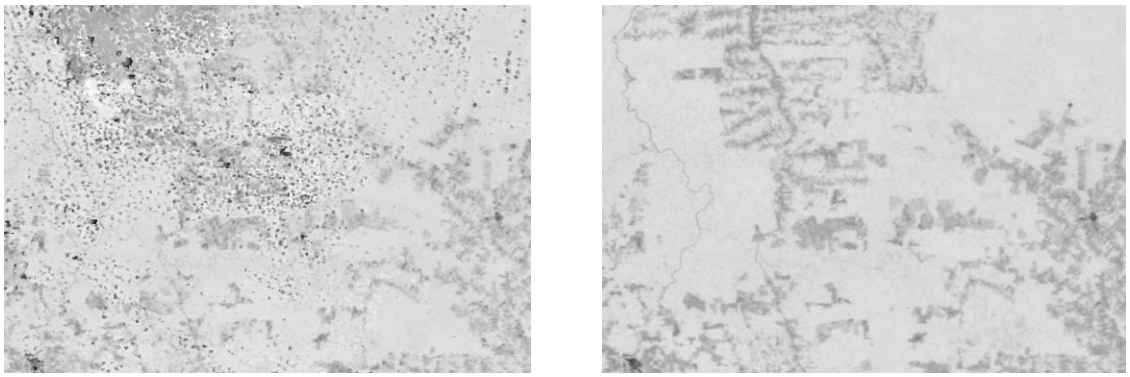

Figure 6 Test area located in Mato Groso, Brazil, with the top-left corner at 59° 30'W, 9° 30'S and the bottom-right one at 58° 30'W, 10° 30'S before and after cleaning.

2.3 The clustering methods

Although the approach is based on the training of a forecasting model on a per pixel basis, it is not computationally efficient to train this model on a pixel basis continentally or pantropically. Consequently, the following algorithm was implemented and applied to the study area using the Euclidian distance function to compare different NDVI time series. By using the index ι (see section the best result of each algorithm was selected and was then compared to the MODIS land cover type (MODIS product MCD12Q1) for the years corresponding to the beginning and the end of the clustered time series. By comparing the results over the same control dataset, each distance function could be assessed in order to find the most appropriate one for land cover clustering using NDVI time-series.

2.3.1 The K-Means algorithm

The k-means algorithm is certainly the most well-known partitioning algorithm (MacQueen, 1967). The data are partitioned into *k* different clusters. To do so the algorithm iteratively partitions the data and recalculates the cluster gravity centre which will therefore move toward the densest area of the cluster. Consequently, data points located in the periphery of the cluster may move to another one during the next iteration. After a certain number of iterations, the algorithm converges to a stable state in which all the data points remain in the same cluster after the recalculation of the gravity centre (MacQueen, 1967). The only parameter necessary to perform the k-means algorithm is the number of clusters *K*. Nevertheless, define the best value for the parameter *K* is a difficult and important step as the quality of the result is directly related to it (MacQueen, 1967).

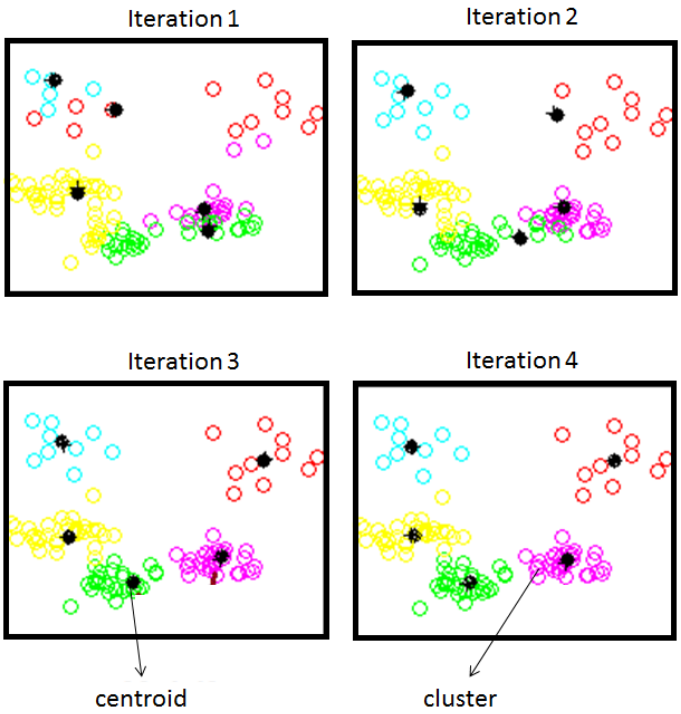

Figure 7 Graphical representation of the different steps of the K-Means algorithm

The pseudo code for the K-Means algorithm is the following:

- **Step 0:** Initialize randomly the means m_1, m_2 … m_k
- **Step 1:** Assign each points *P* of the dataset *D* to the cluster *c* with the nearest mean m_c
- **Step 2:** For each cluster *c* calculate its new means m_c by applying $\frac{1}{|c|}\sum_{P \in c} P$
- **Step 3:** If no points have moved from a cluster to another, the clustering is finished, otherwise go to **Step 1**

2.3.2 The time series comparison methods

The most well-known distance measurement is certainly the Euclidean distance. It represents the length of the segment connecting two given data points *a* and *b* in the data space. In this case, the distance between two NDVI time series *a* and *b* with *n* measurements each is calculated as follows:

$$d = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

2.3.3 The clustering results assessment

Each "algorithm - distance function" combination will generate a set of results depending on the set of parameters used to run the algorithm. In order to choose the best set of parameters, the index ι was used. This index is calculated as follows (Liao, 2005):

$$\iota = (\frac{1}{K}\frac{c}{E}D)^p$$

With *K* being the number of clusters and

$$c = \sum_{i=1}^{N} d(x_i, g)$$

$$E = \sum_{k=1}^{K}\sum_{i=1}^{N} u_{ki} d(x_i, g_k)$$

$$D = \max_{i,j=1}^{K} d(g_i, g_j$$

With *N* being the total numbers of data points, U being a $K \times N$ matrix telling if a points is part of a given cluster or not, *g* being the gravity centre of the whole dataset, g_k the gravity centre of a given cluster and *d(x,y)* being the distance function between two data points. *p* is a parameters used to increase the contrast between two indices and is usually set to 2. One can note that the values *g* and *c* are constant for a given dataset and can therefore be calculated once for all. The higher the ι index is the better is the clustering result.

2.4 The models

2.4.1 Bayesian theory and neural networks

Traditional neural networks use algorithms, such as back-propagation, based on maximum likelihood to estimate the model parameters (Rojas, 1996). The Bayesian approach (Bishop, 2007; MacKay, 1992; Thodberg, 1996) for training neural networks considers the parameters, namely, connection weights, *w* , as random variables drawn from a prior distribution *p(w)* and computes the posterior probability given the data.

This approach has a number of advantages.

1. It allows one to automatically infer, using the *Automatic Relevance Detection (ARD)* algorithm (MacKay, 1992; Thodberg, 1996), which inputs are relevant for the models.

2. The uncertainty about network outputs can be defined by inferring error bars which take into account the uncertainty about the parameters and the remaining noise, from atmospheric, cloud and other effects, in the data (MacKay, 1992; Thodberg, 1996).
3. By calculating a measure of the noise of the modelled data, the network can help to reduce overfitting (Bishop, 2007; MacKay, 1992).

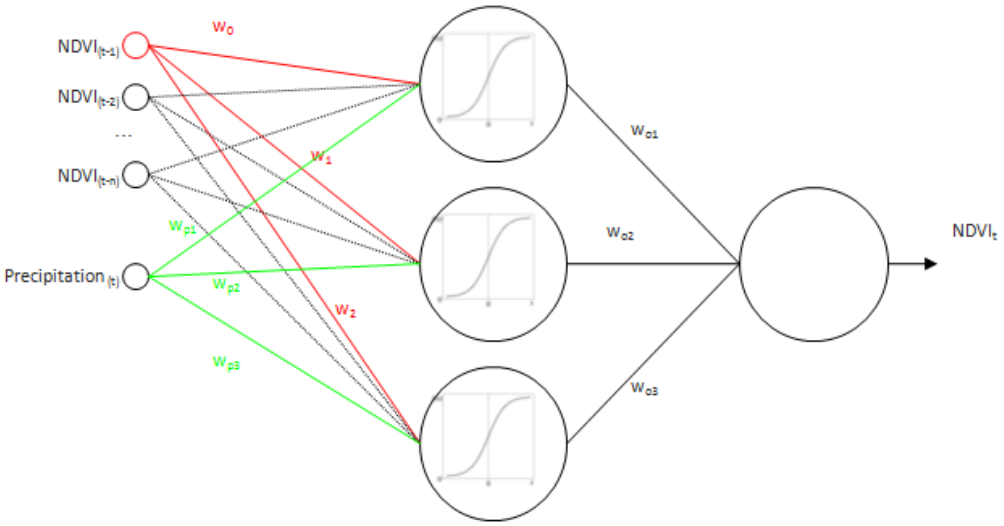

Figure 8 Structure of the artificial neural network implemented

A summary of the procedure follows (after (Bishop, 2007))

2.4.2 Definition of the probabilistic model

The first assumption of this Bayesian representation of neural networks is that the distribution of the target data *t*, in our case the NDVI value at a given date, knowing the input $\mathbf{X}$, the neural network weights $\mathbf{W}$ and the hyperparameter β (which is a scalar) are all Gaussian with the network output $\mathrm{y}(\mathrm{x},\mathrm{w})$, representing the predicted NDVI measurements, as mean and with $\beta^{-\frac{1}{2}}$ as standard deviation. This is given by equation 1.

The notation $\mathrm{N}(\mathrm{x}—\mu,\sigma^2)$ is used to define a normal distribution of mean μ and of variance σ^2.

$$p(t|\mathbf{x},\mathbf{w},\beta) = \mathcal{N}(t|y(\mathbf{x},\mathbf{w}),\beta^{-1})(1)$$

The likelihood function of the dataset $\mathrm{D}=\{\mathrm{t}_1,\ldots,t_N\}$ with N target values t corresponding to N input vectors $\mathrm{X}=\{\mathrm{x}_1,\ldots,x_N\}$ can therefore be defined as

$$p(\mathbf{D}|\mathbf{w},\beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n,\mathbf{w}),\beta^{-1})$$

Let the prior distribution over the weights **w** (which is a vector of size *M*) knowing the hyperparameter α (which is a matrix) be in the form

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|0,\alpha^{-\mathbf{1}})$$

Which yields the following definition of the posterior distribution for the weights.

$$p(\mathbf{w}|\mathbf{D},\alpha,\beta) \propto p(\mathbf{w}|\alpha^{-1})p(\mathbf{D}|\mathbf{w},\beta^{-1}) \quad (2)$$

2.4.3 Automatic Relevance Detection and Weights family

The Automatic Relevance Detection is a technique based on the idea of defining a weights family for each input of the neural network. This means that all the weights in the network don't have the same α parameter. Therefore, in the following development, α is considered as a matrix diagonal. To find the α for weight *n*, one would have to look at the n^{th} element of the αmatrix diagonal, $\alpha_{n,n}$. ARD allows one to determine which inputs, such as the NDVI and the precipitation in the case of Terra-i, are relevant to the neural network output. As shown in section 2.4.6, the most probable α matrix is found given the training dataset. If the α for an input becomes too large, it means this input is not relevant and it can therefore be automatically dismissed.

2.4.4 Model and hyperparameter optimization

The model optimization is divided in two different steps. During the first one, the most probable weights $\mathrm{W}MP$are inferred by maximizing the log of the posterior distribution assuming the hyperparameters α and β are already known. In the second step, the parameters α and β are inferred using the previous result for $\mathrm{W}MP$ and the evidence framework of Bayesian theory (Bishop, 2007; MacKay, 1992). These two steps are conducted in an iterative manner and the more iterations the better approximation of the weights and hyperparameters.

Let first define E_D the error on the data and E_W the error on the weights.

$$E_D(n) = \frac{\beta}{2}(t_n - y(x_n, w))^2$$
$$\mathbf{E_W} = \frac{1}{2}\mathbf{w}^T \alpha \mathbf{w}$$

2.4.5 Maximization of the weights

Since maximizing a function *f* or maximizing *ln(f*) yield the same results, one can ,for simplicity, maximize the log of the posterior distribution (defined by equation 2) instead of the posterior distribution. It is easier to maximize the log of the posterior distribution because term multiplications can be replaced by sums.

$$\ln\, p(\mathbf{w}|\mathbf{D}) = \ln\left(p(\mathbf{w}|\alpha^{-1})p(\mathbf{D}|\mathbf{w}, \beta^{-1})\right) \quad (3)$$

By following (3) and since the hyperparameters are assumed to be known (and can therefore be considered as constant), the function maximizing the log of the posterior is given by equation

$$\ln\, p(\mathbf{w}|\mathbf{D}) = -\mathbf{E_W} - \sum_{n=1}^{N} E_D(n) + const \quad (4)$$

In order to maximize this function (neglecting the constant term) and consequently calculate the most probable weights vector $^{\mathrm{W}}MP$ for the given hyperparameters, the *Scaled Conjugate Gradient* (SCG) can be used as the numerical optimization algorithm (Bishop, 2007).

2.4.6 Hyperparameter optimization

Evidence definition

By using Bayes' theorem, one can write the posterior distribution for α and β as

$$p(\alpha, \beta|\mathbf{D}) = \frac{p(\mathbf{D}|\alpha, \beta)p(\alpha, \beta)}{p(\mathbf{D})}$$

As we have a weak prior knowledge about the noise level and the smoothness of the interpolant, the data dependent term $p(D|\alpha, \beta)$ which is called the **evidence** of α and β will be maximized. The evidence can be reformulated by marginalizing over **w**, which yields.

$$p(\mathbf{D}|\alpha, \beta) = \int p(\mathbf{D}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha)dw \quad (5)$$

Equation 5 can now be integrated as follows

$$p(D|\alpha,\beta) = \left(\frac{|\alpha|}{(2\pi)^M}\right)^{\frac{1}{2}} \left(\frac{\beta}{(2\pi)^N}\right)^{\frac{N}{2}} \int e^{-E_W - \sum_{n=1}^{N} E_D(n)} dw \quad (6)$$

To integrate the exponential term in equation 6, \cite{bishop} recommends to use a Laplace approximation with the covariance **A** given by equation 7.

$$\mathbf{H} = \frac{\partial^2}{\partial \mathbf{w}^2} \sum_{n=1}^{N} \frac{1}{2}(t_n - y(\mathbf{x}_n, \mathbf{w}))^2|_{\mathbf{w}=\mathbf{w}_{MP}}$$
$$\mathbf{A} = \alpha + \beta\mathbf{H} \quad (7)$$

This gives the following result for the integral. This expression is called the **evidence function**.

$$p(\mathbf{D}|\alpha,\beta) = \left(\frac{\alpha}{2\pi}\right)^{\frac{M}{2}} \left(\frac{\beta}{2\pi}\right)^{N} e^{E(\mathbf{w}_{MP})} (2\pi)^{\frac{M}{2}} |\mathbf{A}|^{-\frac{1}{2}} \quad (8)$$

Optimization of α and β

Equation 8 will now be maximized. Indeed, the log of the evidence function is easier to maximize as term multiplications can be replaced by sums.

$$\ln\, p(\mathbf{D}|\alpha,\beta) = \frac{1}{2}\sum_{i=1}^{M} \ln\, \alpha_{i,i} + \frac{N}{2} \ln\, \beta - \sum_{n=1}^{N} E_D(n) - \mathbf{E_W} - \frac{1}{2} \ln\, |A| - \frac{N}{2} \ln(2\pi) \quad (9)$$

To maximize α_c, one takes the derivative of equation 9 with respect to α_c, set equal to 0 which gives equation 10.

$$\alpha_{c,max} = \frac{||C|| - \alpha_c trace_c(\mathbf{A}^{-1})}{\mathbf{w}_{MP}^T \mathbf{w}_{MP}} \quad (10)$$

The same method is used to maximize β which is given by equation 11.

$$\beta_{max} = \frac{N - [M - trace(\mathbf{A}^{-1}\alpha)]}{\sum_{n=1}^{N} [(t_n - y(x_n, w_{MP}))^2]} \quad (11)$$

The Hessian matrix

The **Hessian** matrix has already been defined by equation 7. The Hessian is a crucial component of the hyperparameter optimization and, as we will see, of the anomaly detection process. For this study, the Hessian was evaluated using a generalization to the second order of the *Automatic Differentation Algorithm* (Bishop, 2007)

2.4.7 Anomaly detection

Contrary to common neural networks, Bayesian neural networks don't have a single scalar output. They give the mean (predicted value) and standard deviation (prediction uncertainty) of the predicted value. To detect anomalies, the Bayesian neural network are used to predict, based on a set of *n* previous values $\mathrm{V}=\mathrm{v}_{i-(n+1)}, \ldots, v_{i-1}$, what the next value of the NDVI time series should be. This prediction is then compared to the observed value. The fact that the Bayesian neural network gives a Gaussian probability distribution as output is used to determine the certainty of any anomaly.

More formally, a probability function in the form can be defined.

$$p(t|\mathbf{x}, \mathbf{D}, \alpha, \beta) = \mathcal{N}(t|\mu, var(x)) \quad (12)$$

With the model output as mean.

$$\mu = y(\mathbf{x}, \mathbf{w}_{MP})$$

The input dependent standard deviation arises from the noise of the target data and an x-dependent term which represents the uncertainty in the prediction due to uncertainty in the model's weights.

$$var(\mathbf{x}) = \beta^{-1} + g^T \mathbf{A}^{-1} g$$

with

$$g = \nabla_{\mathbf{w}} y(\mathbf{x}, \mathbf{w})|_{\mathbf{w}=\mathbf{w}_{MP}}$$

By integrating, one can obtain the probability that the observed value, v_t is coherent with the model prediction and that it is therefore not an anomaly.

$$p(t_t = v_t) = \int_{v_t - \omega\sigma_D}^{v_t + \omega\sigma_D} p(t_t|x_t, w, \beta^{-1}) dt$$

Where $\sigma_D = \sqrt{\beta^{-1}}$ and ω is a multiplicative parameter to expand the integration area.

Input-dependent standard deviation

Here we explain how the input dependent standard deviation is calculated. The first term, β^{-1} represents the noise on the target data and is calculated during the hyperparameter optimization step.

The second term is the x-dependent term $\mathrm{g}^T\mathbf{A}^{-1}g$ where $\mathbf{g} = \nabla_{\mathbf{w}} y(\mathbf{x}, \mathbf{w})|_{\mathbf{w}=\mathbf{w}_{MP}}$ and A is equal to the double derivative matrix of $\mathrm{E(w)}$,$\alpha I + \beta\mathbf{H}$. Therefore, the x-dependent term represents the uncertainty on the weights.

To further understand where the x-dependent term comes from, one can define the predictive distribution as

$$p(t|\mathbf{x}, \mathbf{D}) = \int p(t|\mathbf{x}, \mathbf{w}) q(\mathbf{w}|\mathbf{D}) dw \tag{13}$$

Where $q(\mathbf{w}|\mathbf{D})$is the Laplace approximation for the exponential term in 5.

Because of the non-linearity of the network function *y(x,w)* as a function of *w*, this integration is analytically impossible. It is therefore recommended to use a Taylor expansion around W_{MP}.

$$y(\mathbf{x}, \mathbf{w}) = \sum_{n=0}^{\infty} \left(\left[\frac{1}{n!} \frac{\partial^n y(\mathbf{x}, \mathbf{w})}{\partial \mathbf{w}^n} \right]_{\mathbf{w}=\mathbf{w}_{MP}} (\mathbf{w} - \mathbf{w}_{MP})^n \right)$$

And by retaining only the linear term, one gets.

$$y(\mathbf{x}, \mathbf{w}) \approx y(\mathbf{x}, \mathbf{w}_{MP}) + g^T(\mathbf{w} - \mathbf{w}_{MP}) \tag{14}$$

With *g* defined as

$$\mathbf{g} = \frac{\partial}{\partial \mathbf{w}} y(\mathbf{x}, \mathbf{w})|_{\mathbf{w}=\mathbf{w}_{MP}}$$

It is important to note that *g* is easy to numerically calculate using the *Automatic Differentiation Algorithm*.

The distribution of the target data can be easily rewritten:

$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|\mathbf{g}^T(\mathbf{w} - \mathbf{w}_{MP}) + y(\mathbf{x}, \mathbf{w}_{MP}), \beta^{-1}) \tag{15}$$

Using properties of marginal and conditional Gaussians, this can be rewritten as

$$p(t|\mathbf{x}, \mathbf{D}, \alpha, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}_{MP}), \beta^{-1} + \mathbf{g}^T \mathbf{A}^{-1} \mathbf{g}) \quad (16)$$

This is the same as equation 12.

2.5 Identification methods for drivers of change

2.5.1 Flood identification

The LSWI index is based on the fact that the Short-wave infra-red (SWIR) band is sensitive to the moisture in vegetation and in water bodies as has been shown by different studies such as (Gao, 1996). It is calculated as follows in similar manner to NDVI.

$$LSWI = \frac{NIR - SWIR}{NIR + SWIR}$$

The LSWI represents information about the vegetation and its water background while EVI represents the state of the vegetation. (Sakamoto et al., 2007) and (Yan, Ouyang, Guo, Jin, & Zhao, 2010)compute the DVEL index which is the difference between the LSWI value and the EVI one. The DVEL value therefore represents only the information about the background water and hence allows us to define whether a pixel is flooded.

In order to calculate the LSWI index, bands 5 (NIR) and 6 (SWIR) of the MODIS product MOD09A1 were used. The DVEL index was calculated using the EVI values from the MODIS product MOD13Q1. The index was calculated for each anomaly detected and the pixel identified as flooded are flagged as potential floods.

2.5.2 Drought identification

In order to identify a detected pixel as drought, the precipitation data from the TRMM 3b42 product were compared with a baseline dataset (WorldClim 1950-2000). If the precipitation measured by the TRMM sensor are significantly lower than the baseline for the same date and during at least a trimester, the pixel detected is flagged as having potential for drought. The baseline data are provided by the WorldClim dataset (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005)} and represent long term rainfall means over the last 50 years. WorldClim is a set of interpolated climate layers (e.g. monthly total precipitation) of one kilometre square resolution. It has been built over various climate databases for the 1960-90 periods (Hijmans et al., 2005). As the TRMM product and the WorldClim data have been derived from different types of

technologies, a constant shift exists between the mean of both product. The mean of TRMM data for the training period (from 2000 to 2004) was therefore compared with the WorldClim data in order to define this shift for each pixel and hence to improve the comparison result and utility of the TRMM data for representing anomalies.

2.6 Validation and calibration methods

During the validation phase, three properties of the system were tested. The first one is its ability to detect land cover conversion events, the second is its ability to discriminate the noise from the real trends and therefore its ability to detect only habitat conversion events (as opposed to clouds) and finally its ability in identifying drought, floods and true anthropogenic events.

2.6.1 Validation data and Landsat scenes

Landsat images were used for the validation because of their free availability and high resolution (30 meters), which exceeds that of the MODIS images (250m). Using ERDAS Imagine, we performed the Iterative Self-Organizing Data Analysis Technique (ISODATA) as an unsupervised clustering algorithm in order to classify Landsat satellite images. The algorithm identifies different types of cover in the images such as: clouds, water bodies, primary vegetation, secondary vegetation, areas without vegetation cover, cloud shadow, topographic shadow and urban areas.

2.6.2 Vegetation classification using ERDAS

Due to the characteristics of the satellite collection system, this method has limitations such as banding (SLC-OFF) and poor quality images in regions with frequent cloud cover. The classified images were therefore manually reviewed using ArcGIS software, and all areas with clouds, mist and water, masked manually removing them from the analysis as NoData values.

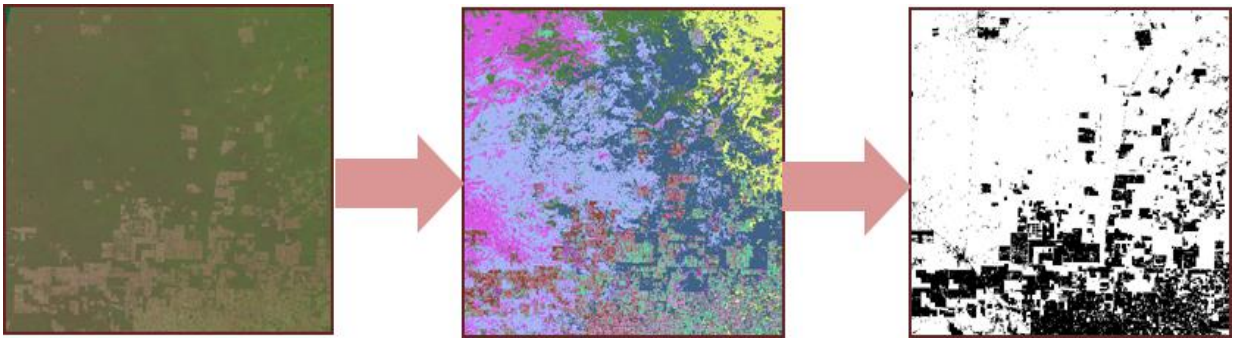

Figure 9 Vegetation classification using ERDAS

The different clusters were finally merged within three different classes: vegetation cover, without vegetation cover and no data. The difference between the two image sets (2004 and 2011) was then calculated. This results in a final image with three values: positive change [NDVI increase](+1), negative change [NDVI decrease](-1) and no change (0).

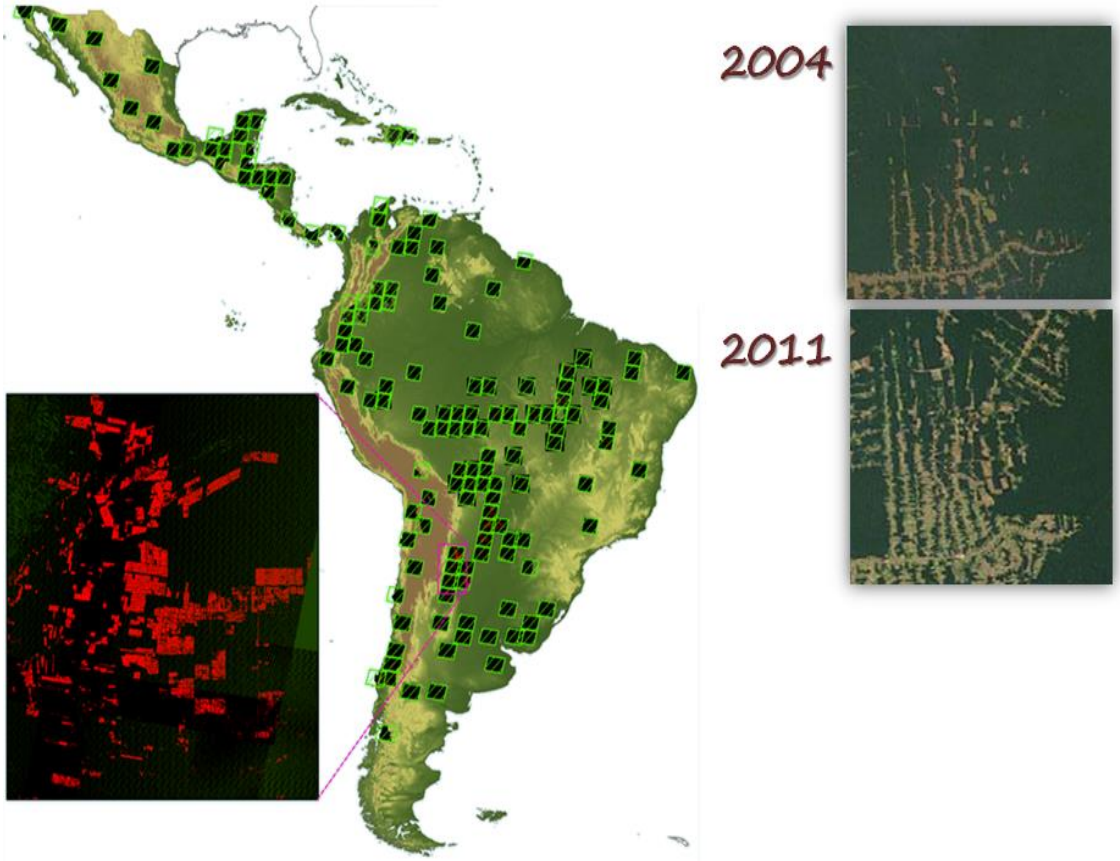

Figure 10 Landsat scenes used for the validation and calibration of the system in Latin America.

Finally, the change detection results were reviewed again in detail, superimposing the original satellite images from 2004 and 2011 over the change product. Where errors were detected in the results (mainly occurring in areas of intensive agriculture) they were corrected manually so that the classified LANDSAT imagery were as true a representation of the LANDSAT scenes as possible.

The results obtained within the test areas were also compared with different sources depending on their availability:

- Literature assessments over administrative regions (departments, nations)
- Other monitoring systems (RALUCIAPA, IMAZON, PRODES)

This database of validated data was split into two independent datasets. The validation dataset which was used to assess the model performance and the calibration dataset which was used in order to find which were the optimal probability thresholds for detection by Terra-i.

2.6.3 Comparison method

As the resolution of the MODIS sensor is relatively low (250m), small scale events are more difficult to detect in MODIS data. Nevertheless, it has been shown by (Koltunov, Ustin, Asner, & Fung, 2009) that events such as selective logging have an effect on phenological markers observable in MODIS NDVI time series. As the resolution of the validation dataset derived from Landsat scenes (see previous section) is much higher (about 30m) it is possible to assess if the MODIS-based model is able to detect sub-pixel events and what proportion of sub-pixel change is necessary for detectability using the MODIS based model. This can be achieved by calculating the proportion of events of a given size that are present in the validation dataset as well as in the MODIS terra-i results. The detections from the comparison dataset were therefore grouped based on sub-pixel disturbance sizes (10%-20%, 20%-30% […] 90%-100% of a MODIS pixel) and the following metrics were calculated for each of them.

1. True positives (TP) : the pixel indicates a detection in both images.
2. False positives (FP) : the pixel indicates a detection in the result of Terra-i, but not in the human tagged image.

3. True negatives (TN) : the pixel indicates "not detected" in both images.
4. False negatives (FN) : the pixel indicates "not detected" in the result of Terra-i, but not in the human tagged image.

With these four measurements, one can compute two measures quite often used in information retrieval and pattern recognition: the precision and the recall. The formulas of these two measures are given by the following equations, where tp is the number of true positive pixels, fp the number of false positive pixels and fn the number of false negative pixels:

$$pecision = \frac{tp}{tp + fp}$$

$$recall = \frac{tp}{tp + fn}$$

The precision measures the probability that a pixel detected by the Terra-i processing is really a deforested one according to LANDSAT and the recall measures the probability that a deforested pixel in LANDSAT is given as detected by the Terra-i algorithms. We cannot use these measures independently, because it is possible to have a recall of 1 by setting all the pixels to detected (in this case, fn = 0 and the recall becomes 1). However, in this case precision will be low (and fp excellent) .

Therefore one needs a third measure that takes into account both the precision and the recall. The F-Measure is perfectly suitable in this case, because it is based on a harmonic mean. This implies that, to get a good value, the precision and the recall must both be good. If only one of them is good, the result will be bad. The generalized formula of the F-Measure is given by the following formula:

$$\frac{(1+\beta^2) D_1 D_2}{\beta^2 D_1 + D_2} \quad (17)$$

where β is a parameter indicating how many times one attaches as much importance to the recall (D_1) as the precision (D_2).

2.6.4 Model calibration methods

The calibration approach is to compare the results of the terra-i MODIS model with the LANDSAT based calibration dataset to find the optimal probability threshold for detection. This analysis is

repeated for each class of the analysed area, allowing the definition of the optimal threshold for each cluster. For each map of detections of the calibration dataset and for each cluster of the analysed area, the F-Measure is calculated as defined by equation 17. The purpose of the calibration algorithm is to maximize the F-Measure. This is easily done by calculating it with different thresholds: the one with the highest result is chosen as the most appropriate threshold for the given cluster.

3 References

Alfonso, A., Dallmeier, F., Granek, E., & Raven, P. (2001). Biodiversity: Connecting with the Tapestry of Life. *Smithsonian Institution/Monitoring and Assessment of Biodiversity Program and Presidents Committee of Advisors on Science and Technology*.

Bishop, C. M. (2007). *Pattern Recognition and Machine Learning* (2nd ed.). Springer.

Bonan, G. B. (2008). Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science*, *320*(5882), 1444–1449.

FAO. (2010). Global Forest Resources Assessment 2010. *FAO forestry paper, Rome, Italy.*, (163).

FAO, F. and A. O. of the U. N. (1996). Global Forest Resources Assessment 1996. *FAO forestry paper, Rome, Italy.*, *130*.

FAO, F. and A. O. of the U. N. (2000). Global Forest Resources Assessment 2000. *FAO forestry paper, Rome, Italy.*, *140*.

FAO, F. and A. O. of the U. N. (2005). Global Forest Resources Assessment 2005. *FAO forestry paper, Rome, Italy.*, *147*.

Gao, B. C. (1996). NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266.

Geist, H. J., & Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical deforestation. *Bioscience*, *52*(2).

Gerrand, A., Lindquist, E., & D'Annunzio, R. (2011). Remote sensing survey updates forest-loss estimates. *Unasylva*, *62*(238), 14 – 15.

Gibson, R. B., Holtz, S., Tansey, J., Whitelaw, G., & Hassan, S. (2005). *Sustainability Assessment: Criteria and Processes*. Earthscan Ltd.

Hansen, M. C., Stehman, S. V, Potapov, P. V, Loveland, T. R., Townshend, J. R. G., DeFries, R. S., … DiMiceli, C. (2008). Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 9439–9444.

Hansen, M., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., & Sohlberg, R. A. (2003). Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. *Earth Interactions*, *7*(10), 1–15.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

Hird, J. N., & McDermid, G. J. (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, *113*.

Huete, A., Justice, C., & van Leeuwen, W. (1999). *MODIS Vegetation Index (MOD 13) Algorithm Theoretical Basis Document*.

INPE. (2009). *Estimativas Anuais desde 1988 at{é} 2009*. Retrieved from http://www.obt.inpe.br/prodes/prodes{_}1988{_}2009.htm

Jun, W., Zhongbo, S., & Yaoming, M. (2004). Reconstruction of a Cloud-free Vegetation Index Time Series for the Tibetan Plateau. *Mountain Research and Development*, *24*, 348–353.

Koltunov, A., Ustin, S. L., Asner, G. P., & Fung, I. (2009). Selective logging changes forest phenology in the Brazilian Amazon: Evidence from MODIS image time series analysis. *Remote Sensing of Environment*, *113*, 2431–2440.

Liao, T. W. (2005). Clustering of time series data-a survey. *Pattern Recognition*, *38*, 1857–1874.

MacKay, D. J. C. (1992). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*, *4*(3), 448–472.

MacQueen, J. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Berkeley Symposium on Mathematical Statistics and Probability*, *5*, 281–297.

Mayaux, P., Eva, H., Brink, A., Achard, F., & Belward, A. (2008). Chapter 5, Remote Sensing of Land-Cover and Land-Use Dynamics. In E. Chuvieco (Ed.), *Earth observation of global change: the role of satellite remote sensing in Monitoring the Global Environment* (pp. 85–108). Springer Netherlands.

Mayaux, P., Holmgren, P., Achard, F., Eva, H., Stibig, H.-J., & Branthomme, A. (2005). Tropical forest cover change in the 1990s and options for future monitoring. *Philosophical Transactions: Biological Sciences*, *360*(1454), 373–384.

Mulligan, M. (2008). *RALUCIAPA : Rapid assessment of land use change in and around protected areas (2000-2005)*. Retrieved from http://www.unep-wcmc.org/

NASA. (1997). *Algorithm 3B42 - TRMM Merged HQ/Infrared Precipitation*. Retrieved from http://trmm.gsfc.nasa.gov/3b42.html

Nelson, G. C. (2005). Drivers of Ecosystem Change: Summary Chapter. *Millenium Ecosystem Assessment, Current State and Trends Assessment*.

Parker, C., Mitchell, A., Trivedi, M., & Mardas, N. (2008). *The little REDD book*. Global Canopy Programme.

Pearce, D. W., Putz, F., & Vanclay, J. K. (2002). Is sustainable forestry economically possible? In *Valuing the Environment in Developing Countries: Case Studies.* (pp. 447–500).

Roerink, G. J., Menenti, M., & Verhoef, W. (2000). Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *Internonational Journal of Remote Sensing*, *21*, 1911–1917.

Rojas, R. (1996). The Backpropagation Algorithm. In *Neural Networks* (pp. 151–184). Springer-Verlag Berlin.

Sakamoto, T., Nguyen, N. Van, Kotera, A., Ohno, H., Ishitsuka, N., & Yokozawa, M. (2007). Detecting temporal changes in the extent of annual flooding within the Cambodia and the Vietnamese Mekong Delta from MODIS time-series imagery. *Remote Sensing of Environment*, *109*, 295–313.

Shao, Y., & Lunetta, R. S. (2010). The use of MODIS NDVI data for characterizing cropland across the great lakes basin. *International Journal of Applied Earth Observation and Geoinformation*, *12*, 81–88.

Sierra, R. (2001). The role of domestic timber markets in tropical deforestation and forest degradation in Ecuador: implications for conservation planning and policy. *Ecological Economics*, *36*, 327–340.

Thodberg, H. H. (1996). A review of Bayesian neural networks with an application to near infrared spectroscopy. *IEEE transactions on neural networks*, *7*, 56–72.

Verhoef, W., Menenti, M., Azzali, S., W.Verhoef, M.Menenti, & S.Azzali. (1996). A colour composite of NOAA-AVHRR-NDVI based on time serie analysis (1981-1992). *International Journal of Remote Sensing*, *17*, 231–235.

Wang, Q., & Tenhunen, J. D. (2004). Vegetation mapping with multitemporal NDVI in North Eastern China Transect (NECT). *International Journal of Applied Earth Observation and Geoinformation*, *6*, 17–31.

Xiaa, Z., Ruib, S., Binga, Z., & Qingxia, T. (2008). Land cover classification of the North China Plain using MODIS EVI time series. *ISPRS Journal of Photogrammetry {&} Remote Sensing*, *63*, 476–484.

Yan, Y.-E., Ouyang, Z.-T., Guo, H.-Q., Jin, S.-S., & Zhao, B. (2010). Detecting the spatiotemporal changes of tidal flood in the estuarine wetland by using MODIS time series data. *Journal of Hydrology*, *384*, 156–163.