

## Hyper-temporal SPOT-NDVI dataset parameterization captures species distributions

Atkilt Girma, C.A.J.M. de Bie, Andrew K. Skidmore, Valentijn Venus & Frans Bongers

To cite this article: Atkilt Girma, C.A.J.M. de Bie, Andrew K. Skidmore, Valentijn Venus & Frans Bongers (2016) Hyper-temporal SPOT-NDVI dataset parameterization captures species distributions, International Journal of Geographical Information Science, 30:1, 89-107, DOI: 10.1080/13658816.2015.1082565

To link to this article: <http://dx.doi.org/10.1080/13658816.2015.1082565>



Published online: 12 Sep 2015.



Submit your article to this journal [↗](#)



Article views: 138



View related articles [↗](#)




View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

## Hyper-temporal SPOT-NDVI dataset parameterization captures species distributions

Atkilt Girma <sup>a,b,\*</sup>, C.A.J.M. de Bie<sup>a</sup>, Andrew K. Skidmore<sup>a</sup>, Valentijn Venus<sup>a</sup>  
and Frans Bongers<sup>c</sup>

<sup>a</sup>ITC faculty, University of Twente, Enschede, The Netherlands; <sup>b</sup>Department of Land Resources Management and Environmental Protection, Mekelle University, Mekelle, Ethiopia; <sup>c</sup>Forest Ecology and Forest Management Group, Centre for Ecosystem Studies, Wageningen University, Wageningen, The Netherlands

(Received 17 December 2014; accepted 10 August 2015)

Hyper-temporal SPOT NDVI images contain useful information about the environment in which a species occurs, including information such as the beginning, end, peak, and curvature of photosynthetically active vegetation (PAV) greenness signatures. This raises the question: can parameterization of hyper-temporal SPOT NDVI images be useful to predict species distribution? A set of SPOT-NDVI images for the whole of Ethiopia covering nine years was classified using the unsupervised ISODATA clustering algorithm to group similar NDVI pixel values. The HANTS (Harmonic ANALYSIS of Time Series) algorithm, that fits series of smoothing cosine waves, was then applied to the time series for each of the NDVI classes to generate seven output Fourier components. These components, together with the topographic parameters slope and elevation, were used as predictors in a species distribution model using MAXENT. Presence-only data of one test species, *Boswellia papyrifera*, were modelled. This species is diminishing at an alarming rate and requires conservation. The performance of the model was evaluated by the area under curve (AUC) of the receiver-operating characteristics value. The output distribution map was tested for its agreement with the NDVI-clustering approach and conventional *B. papyrifera* distribution map using Kappa. The relative contributions of the first four predictors to the MAXENT in sequence were: 2nd harmonic phase, elevation, amplitude of the 1st harmonics, and amplitude of the 2nd harmonics. The average AUC test result for the 100 runs was 0.98 with a standard deviation of 0.002. The probability distribution map clearly shows high correlation with the *B. papyrifera* occurrence data. In addition, the distribution map was found to be in agreement with the NDVI-clustered and conventional map with improved details. Classifying hyper-temporal NDVI images and extracting their parameters through the use of the HANTS algorithm captures the PAV greenness behaviour (parameters) of the environment of the species studied. These parameters have proved successful in predicting the distribution of *B. papyrifera*.

**Keywords:** *Boswellia papyrifera*; HANTS; hyper-temporal; MAXENT; SPOT-NDVI

### 1. Introduction

Selection of appropriate environmental predictors (covariates) increases species distribution prediction accuracy (Phillips *et al.* 2006). Hyper-temporal SPOT NDVI images (for details see <http://www.spotimage.com/>) contain useful information about vegetation cover attributes, such as density/biomass (Seaquist *et al.* 2003), and the start and cessation of

---

\*Corresponding author. Email: [atkiltgirmagebrekidan@utwente.nl](mailto:atkiltgirmagebrekidan@utwente.nl)

photosynthetically active vegetation (PAV) greenness in a season (Wang *et al.* 2005, Lasaponara 2006). It has been noted that an NDVI pattern is a complex resultant of plant phenology and a number of climatic determinants (Immerzeel *et al.* 2005). For example, seasonal variation in NDVI is mainly reported to be due to change in plant phenology, while variation in precipitation and temperature is related to time of the year (Prasad *et al.* 2005). Classified hyper-temporal NDVI images are reported to be successful for mapping crop cover areas (Nguyen *et al.* 2012) using harmonics analysis (Jakubauskas *et al.* 2002), for crop identification and yield estimation (Khan *et al.* 2010), land use mapping (Khan 2011) using ISODATA clustering algorithms, environmental quality using image differencing and principal component analysis (PCA) (Fung and Siu 2000), and for drought assessment using fuzzy sets theory (Rulinda *et al.* 2010). Thus, extraction of relevant parameters of photosynthetically active vegetation (PAV) greenness from hyper-temporal NDVI image signatures could provide useful inputs into species distribution models (SDMs).

Hyper-temporal NDVI signatures can be parameterized using the HANTS (Harmonic Analysis of Time Series) algorithm (Verhoef *et al.* 2005). This algorithm removes cloud contamination and interpolates a hyper-temporal NDVI dataset in order to reconstruct gap-free images. The smooth curve (series of sine/cosine waves, Figure 1) fitted to the NDVI time-series image is described by harmonic functions, each defined by their amplitude and phase (Verhoef *et al.* 2005). Amplitude is equal to half the height of a wave, while phase angle (or simply phase) defines the offset between the origin and the peak of a wave over the range 0 to  $2\pi$  (Jakubauskas *et al.* 2001). The addition of several harmonic terms results in different shapes that mimic the shape of the NDVI curve. Moreover, the HANTS algorithm generates amplitude and phase datasets that can be used as input into a species distribution model (SDM). The number of frequencies (terms) can be determined based on knowledge of PAV greenness over space and time. One of the

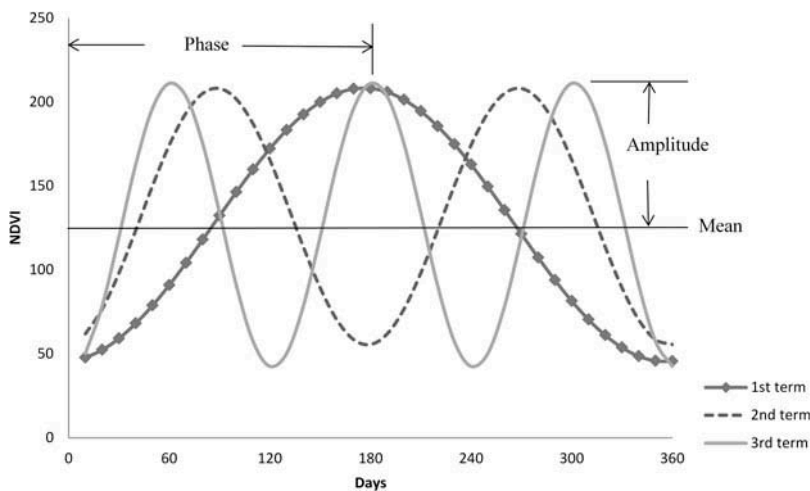


Figure 1. The different Fourier components useful for adaptation of the shape of the NDVI curve and for the removal of clouds. The 1st, 2nd, and 3rd terms (Harmonics) correspond to one, two, and three frequencies, respectively. The zero frequency (straight line) is defined by the mean. The phase angle defines the offset between the origin and the peak of the wave, as shown for the first term only.

disadvantages of this algorithm is that there are no objective rules to determine the HANTS control parameters (Roerink *et al.* 2000). It requires experience with the preceding Fast Fourier Transform (FFT) analysis and the evaluation of several combinations of control parameters.

To conserve an endangered tree species, understanding its current and potential geographic distribution is of importance. Several studies (Pearce and Boyce 2006, Marmion *et al.* 2009, Elith *et al.* 2011) have shown that the generation of species distribution maps can be achieved using appropriate SDMs. Predicting species distribution and abundance with high levels of accuracy is difficult due to the complexity and inherent variation in species, their response to physical and biological factors at multiple scales, and the dynamic nature of environments and species range (Botkin 1990, Scott *et al.* 2002). Many SDM algorithms are available, with the main differences between them being whether systematic or non-systematic survey data are used (Elith *et al.* 2011), and in statistical approach (Guisan and Zimmermann 2000).

*Boswellia papyrifera* (Del.) Hochst, the species used in this study, is a deciduous tree species that produces the widely traded aromatic olio-gum resin called frankincense or olibanum (Fichtl and Admasu 1994). The tree and its olibanum have several medicinal, industrial, household and ecological uses (Adamson 1969, Gebrehiwot 2003, Atta-ur-Rahman *et al.* 2005, Tadesse *et al.* 2007). *Boswellia papyrifera* abundance and distribution information available in Ethiopia mostly consists of estimates and/or sketch maps. The species is mainly found in dry-woodland and wooded-grassland located at low and mid-altitudes between 950 and 1800 m a.s.l. (Fichtl and Admasu 1994). On a local level, the tree is known to grow well on shallow, rocky, and steep slopes (Gebrehiwot *et al.* 2003). The plant has shallow roots, making it susceptible to wind attack. Hence, rocky areas provide *B. papyrifera* with good anchorage for its survival. The species is diminishing at an alarming rate due to fire, insect attack, and human influence, coupled with poor reproduction rates (Gebrehiwot 2003, Girma *et al.* 2013, 2015).

Our approach to modelling the current distribution of *B. papyrifera* exploits the temporal dimension of NDVI (species ecology) into SDM through parameterization of classified hyper-temporal SPOT NDVI datasets. To elaborate, first, the temporal dimension of the hyper-temporal SPOT NDVI datasets was captured using the ISODATA clustering algorithm as used by Ali *et al.* (2013) and de Bie *et al.* (2012). The output of this algorithm was a class map with NDVI signatures. Second, the NDVI signatures were parameterized using the HANTS algorithm to generate NDVI signature peaks (amplitudes) and angles of the curvature of PAV greenness (phase) based on the number of PAV greenness periods. These values contained valuable information about land cover and the environment, and were used as input in MAXENT to model the distribution of *B. papyrifera*. This research was conducted in Ethiopia with the objective of determining *B. papyrifera* tree distribution through hyper-temporal NDVI datasets parameterization.

## 2. Method

### 2.1. Study area

This research was conducted for the whole of Ethiopia (3° 23' to 14° 51' North and 32° 59' to 47° 58' East), a land area of approximately 1.1 million km<sup>2</sup>. Ethiopia spans a wide elevation range (from -130 m in the Afar depression to 4550 m at the peak of Mount

Ras Dashen CSA 2000), and has a diversity of vegetation types (Edwards *et al.* 1976, Azene 2007, Hadgu 2008), agro-ecological zones (Hurni 1998, MoARD 2007), and geological formations (metamorphic, sedimentary, and igneous rock formations).

The climate of Ethiopia varies both spatially and temporally and it is directly related to elevation and topography. The total annual rainfall ranges from less than 150 mm per year in the east of the Afar region to 2200 mm per year around Limu, in the Oromia region. The lowest mean maximum temperature (10°C) occurs over the north-western and central highlands, whereas the highest mean maximum temperature (47°C) occurs in the Afar depression (CSA 2000).

Ethiopia has a very diverse set of ecosystems, ranging from humid forest and extensive wetlands in the west to the desert of the Afar depression. This diversity is attributed to the variation in climate, topography, and vegetation. The forest resources of Ethiopia have significantly and steadily declined in size during the past half century, covering 3.5% of the country and comprising about 6500–7000 species of higher plants, 12% of which are endemic (WBISP 2004). There are several natural gum- and resin-producing tree species in the dry lowlands of Ethiopia, among them are several species of *Boswellia*, *Commiphora*, and *Acacia*. They contribute to the livelihoods of local communities in terms of food security, industrial supply, income generation, and foreign exchange earnings (Gebrehiwot 2003, Lemenih and Teketay 2003, Tadesse *et al.* 2007).

## 2.2. Species data

A total of 545 geo-referenced presence-only *B. papyrifera* observations were used, of which 343 were collected for this study and 202 were obtained from the Bureau of Agriculture and Natural Resources (BoANR 2003, 2004) for the Kafta Humera (29), Tahtay Adiabo (46), Asgede Tseimbla (92), and Tselemti (35) woredas (a woreda is equivalent to a county administrative unit). These point records were spatially scattered, covering the major areas of occurrence of *B. papyrifera* (Figure 2).

## 2.3. Environmental data

### 2.3.1. NDVI

De-clouded 10-day maximum (S10 MVC) SPOT-NDVI synthesis images at a 1-km spatial resolution were obtained from the [www.VGT.vito.be](http://www.VGT.vito.be) website for April 1998–March 2007 (9 years, 324 images). The data for Vegetations 1 and 2 were obtained from SPOT-4 (April 1998–April 2002) and SPOT-5 (May 2002–March 2007), respectively.

### 2.3.2. Topographic parameters

Elevation and slope variables were derived from the Shuttle Radar Topography Mission (SRTM, USGS 2004) digital elevation model (DEM) and re-sampled from the original, finer resolution (90 m pixel size) to a 1 km × 1 km pixel size for the whole of Ethiopia. Both elevation and slope variables were considered as inputs in the distribution modelling. Elevation is strongly related to *B. papyrifera* presence (Figure 3). No *B. papyrifera* were recorded below 500 m or above 1800 m.

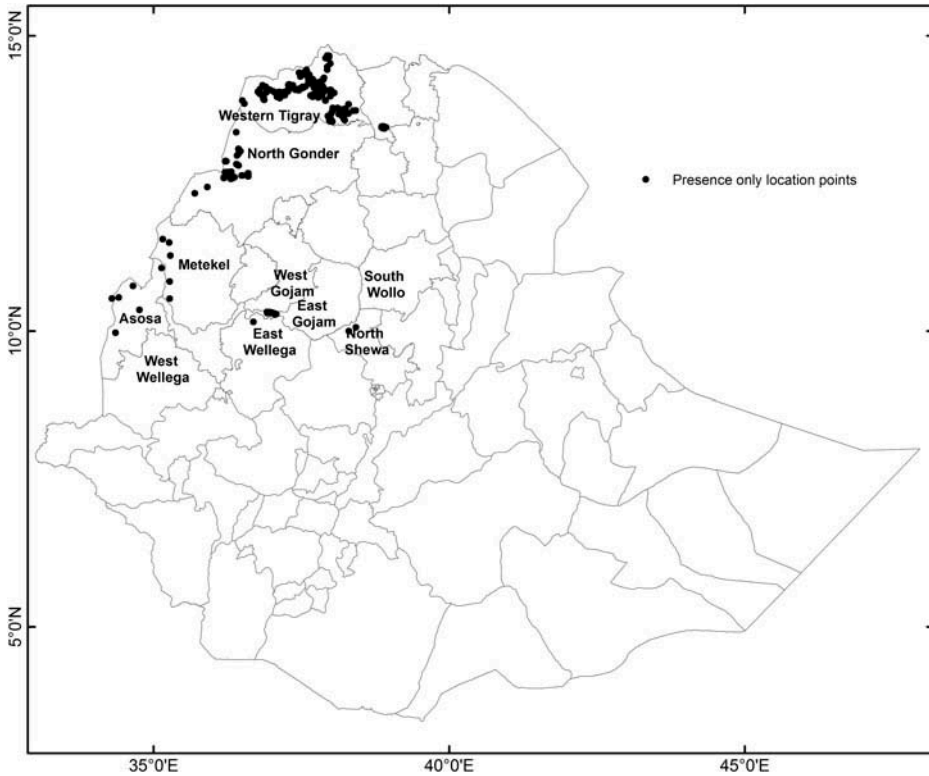


Figure 2. Ethiopia zone boundary map with the *B. papyrifera*-presence-only location points.

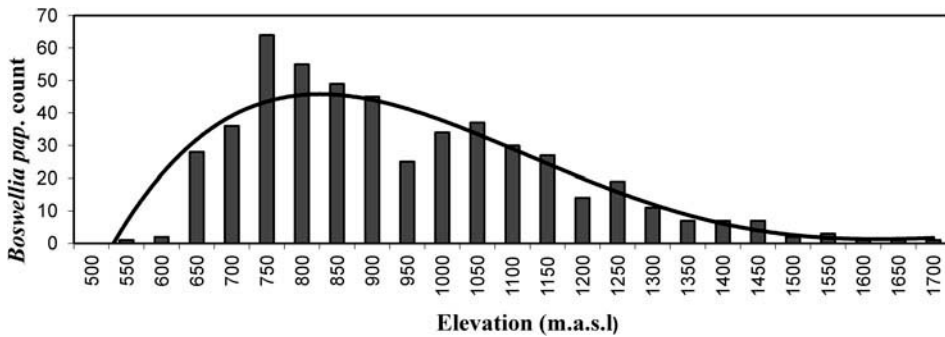


Figure 3. The distribution of *B. papyrifera* presence observations relative to elevation.

#### 2.4. NDVI analysis

Using the ISODATA clustering algorithm in the ERDAS Imagine software, unsupervised classification runs on SPOT-NDVI hyper-temporal datasets were carried out for sets of 10–200 classes using a step size of one, each with 50 iterations. The signatures were analysed to obtain minimum and average divergence statistical indicators (de Bie *et al.*

2008, 2011, Khan *et al.* 2010). These values in turn were used to choose the best number of classes to describe the time-series data. To obtain further insight in possible grouping of NDVI classes, cluster analysis using complete linkage with Euclidean distance as distance measure was implemented in STATISTICA (StatSoft, Inc. [www.statsoft.com](http://www.statsoft.com)). This kind of analysis, which classifies hyper-temporal NDVI datasets into similar environmental units, is referred to below as the 'NDVI-clustering' approach. The signature values extracted for each class (covering nine years) were averaged to form one year of data for subsequent HANTS analysis.

### 2.5. HANTS algorithm

The HANTS algorithm decomposes the time signal for each individual pixel or class into its Fourier components to represent the original NDVI signal. The smooth curve that was fitted to the classified NDVI signatures was described by means of an average NDVI value plus a short series of harmonic (cosine) functions, each based on its amplitude and phase value. The amplitudes and phases computed using this method were more reliable than those based on a straightforward FFT to correct cloudy observations (Verhoef *et al.* 1996, 2005). Fourier analysis can be applied on either continuous or discrete (finite length time period) datasets. FFT, a highly efficient version of Discrete Fourier Transform (DFT), converts the time space into frequency space (Cooley and Tukey 1965). Fourier transformation can be used to remove image noise such as striping and cloud by identifying periodicities.

Curve fitting in the HANTS algorithm is controlled by five parameters. (1) The analysis was fixed at three vegetation growth cycles (three harmonics) with periods of 12, 6, and 4 months, corresponding to PAV greenness. (2) A low-directional-outliers-value (DO-value) was selected to remove low NDVI values (outliers) usually representing cloud. (3) Fit error tolerance (FET) was set to 10. This value determines the absolute difference in the High-Low (Hi/Lo) direction of the remaining (non-rejected) data points with respect to the current curve, after each iteration. When the FET is set too low, the fit will be based on too few data points and becomes unreliable. (4) The degree of over-determinedness (DOD) value was set to 5. The minimum number of extra data points, which have to be used in the ultimate fit, is determined by the DOD (more data points than the necessary minimum). (5) A delta factor (damping factor) of 0.1 was used since the input data had no gaps; however, this number does control the suppression of high amplitudes when there is a long data gap during, for example, the winter period. Accordingly, seven harmonic parameters were computed from the NDVI signature: mean NDVI, 1st amplitude, 1st phase, 2nd amplitude, 2nd phase, 3rd amplitude, and 3rd phase terms. These terms were later converted into Arc/Info ascii grid format for analysis using MAXENT SDM.

### 2.6. Maximum entropy

MAXENT version 3.3.3e (<http://www.cs.princeton.edu/~schapire/maxent>), a general purpose method that estimates probability of distributions based on the principle of maximum entropy (Phillips *et al.* 2006), was used to model the distribution of *B. papyrifera*.

The input data used for MAXENT included nine predictor variables (elevation, slope, and the seven harmonic parameters). These were selected from a larger set of variables including climate (rainfall, evapotranspiration, minimum and maximum temperature), topography (aspect, slope, elevation), soil (such as type, depth, and texture), and NDVI

(amplitude and phases). These were checked for multicollinearity using variance inflation factors (VIFs), from which rainfall was found to be collinear with both NDVI and elevation, while temperature (minimum and maximum) and evapotranspiration were found to be collinear with elevation. Thus, rainfall, temperature (minimum, maximum), and evapotranspiration variables were excluded from the MAXENT analysis. The remainder of the variables had VIF values of less than 5 and were used as inputs. Of these, the three most contributing variables (2nd phase, elevation, and 1st Amplitude) had values of less than 1.5.

MAXENT has a built-in jackknife that measures the importance of each predictor by training the solution with each other predictor or variable first omitted, then used in isolation. In addition, MAXENT provides response curves of all predictors to show how the prediction depends on a particular variable. Moreover, MAXENT has an inbuilt method for adjusting regularization (Elith *et al.* 2011), smoothing the model to make it more regular. A default value of 1 was used for the regularization parameter. Twenty-five per cent (119) of the samples were randomly set aside by MAXENT as test points to model prediction accuracy. A high probability value output from MAXENT indicates that a grid cell contains *B. papyrifera*. Duplicate presence records that fell within a 1 km × 1 km area were discarded, leaving a total of 476 of the 545 observations to be used in MAXENT.

## 2.7. Accuracy assessment

Application of a model will have little merit if the accuracy of the prediction is not tested (Pearson 2007). The MAXENT model was tested to check if it was better than random using (threshold-independent) receiver operating characteristics (ROC) area under curve (AUC). To improve prediction, the MAXENT analysis was repeated (replicated) 100 times. Among the three replicate run type options (Subsample, Crossvalidate, and Bootstrap) available in MAXENT, the 'Subsample' method was used to sub-sample the presence data sets to evaluate model performance. A distribution map was generated from the average of the 100 runs.

In the threshold-independent test, the AUC was calculated for both the training and the test datasets. AUC provides a single measure of model performance, independent of any particular choice of threshold. The ROC curve (sensitivity versus 1-specificity) was plotted for all possible thresholds. Specificity is the proportion of true-positive and false-positive absences, and sensitivity is the proportion of true-positive and false-positive presences. Because only presence-only data were available, the 'fractional predicted area' (the fraction of the total study area predicted present) was used instead of the more standard commission rate (fraction of absences predicted present) (Phillips *et al.* 2006). In addition, the relationship between the MAXENT output distribution map classes (categorized into seven) and *B. papyrifera* presence-only data was studied using box plots.

A conventional statistical accuracy comparison was carried out using overall accuracy and Kappa statistics (Congalton and Green 2009). The conventional and 'NDVI-clustered' maps were compared with the *B. papyrifera* distribution map (MAXENT output) to evaluate agreement and to test if the classification was better than random using Kappa statistic ( $\hat{K}$ ). The conventional map (Figure 10) is crude. It shows only *B. papyrifera* and non-*B. papyrifera* areas. It does not show land cover or spatial differences. Thus, only the Kappa statistic ( $\hat{K}$ ) was used to check the agreement between the maps. Agreement was interpreted as follows:



- <0 Poor agreement
- 0.01–0.20 Slight agreement
- 0.21–0.40 Fair agreement
- 0.41–0.60 Moderate agreement
- 0.61–0.80 Substantial agreement
- 0.81–1.00 Near-perfect agreement

## 2.8. Uncertainty

As with any model, SDMs are subject to uncertainty. Uncertainty in an output map can result from the occurrence data, the modelling method used, and the environmental layers upon which it is based (Fernandez *et al.* 2009, Wiens *et al.* 2009, Convertino *et al.* 2014). Here, the issue of uncertainty was addressed by analysing the ROC curve, Jackknife, and the response curves of each input variable. Focus was on the three input variables that contributed to the highest gain in the model performance, namely the 2nd phase, elevation, and 1st amplitude.

## 3. Results

### 3.1. NDVI classification and signature analysis

The hyper-temporal NDVI datasets covering nine years that were classified using the ISODATA clustering algorithm produced 140 separable NDVI classes. For reasons of simplicity, only NDVI classes with *B. papyrifera* recorded present are displayed (Figure 4d). The highest number of presence records is contained in NDVI class 123 (Figure 4c). Even though the proportion of sample points among NDVI classes is biased by sampling, the NDVI classes were regrouped into three using natural breaks based on the number of presence counts (Figure 4c). All the NDVI classes with *B. papyrifera* presence records are found in the north and north-west of Ethiopia with an east-to-west gradient (Figure 4d). Cluster analysis (tree diagram) of the hyper-temporal NDVI signatures clearly shows that most of the *B. papyrifera* observed GPS locations fall into two main categories (Figure 4a–c). Of the two main categories, only NDVI class 113 was not sampled due to its small size (25 km<sup>2</sup>). Though the presence data is near comprehensive, the quantity of samples among NDVI units is biased. The map (Figure 4d) thus presents only a first impression of where *B. papyrifera* can be found in Ethiopia.

### 3.2. HANTS analysis

The HANTS algorithm successfully produced a cloud-free NDVI dataset and hyper-temporal NDVI data reduction into seven harmonic components (parameters), the seven components being the mean, and the amplitude and phase of the 1st, 2nd, and 3rd harmonic (Figure 5), respectively. The amplitude of a zero frequency equals the mean value. The environment where *B. papyrifera* was recorded present has a single PAV greenness season (high amplitude), with peak NDVI occurring during August/September.

The phase angle represents approximately the time of the year with highest amplitude (PAV greenness value). For example, most of the *B. papyrifera* areas have a first phase angle of 160°. This means the NDVI peak is approximately 160 days after 1 April, which will be end of August to beginning of September (Figures 5 and 9).

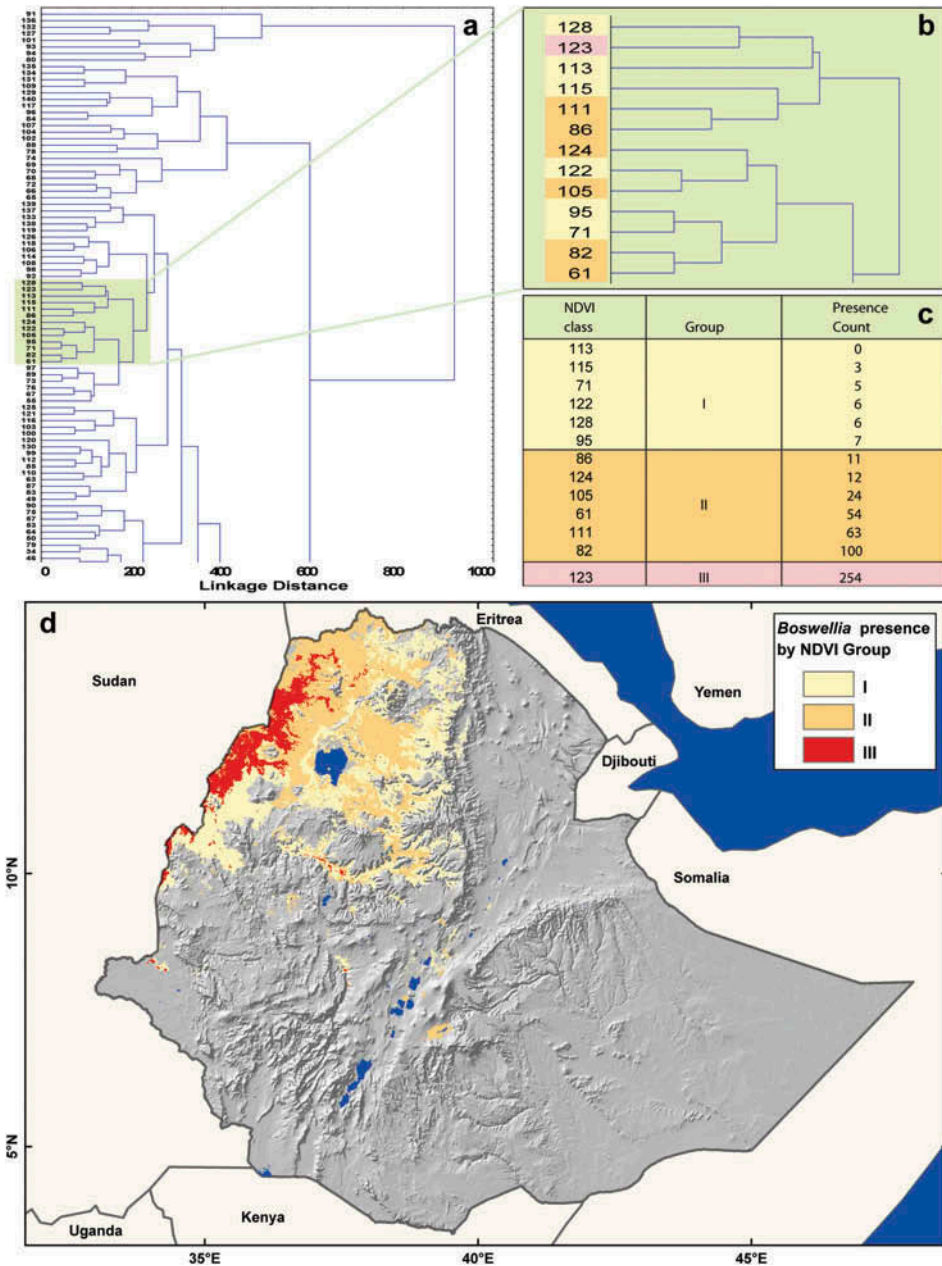


Figure 4. NDVI class signatures (covering nine years) clustering based on Euclidean distance (a, b), which is then regrouped based on the number of presence counts (c) in each NDVI class, to produce a *B. papyrifera* distribution map (d) with three groups. NDVI class 113, being very small in size, was not sampled.

### 3.3. *Boswellia papyrifera* distribution in Ethiopia

The MAXENT model generated a distribution map of *B. papyrifera* for the whole of Ethiopia at a 1-km spatial resolution. The testing and training omission and predicted area

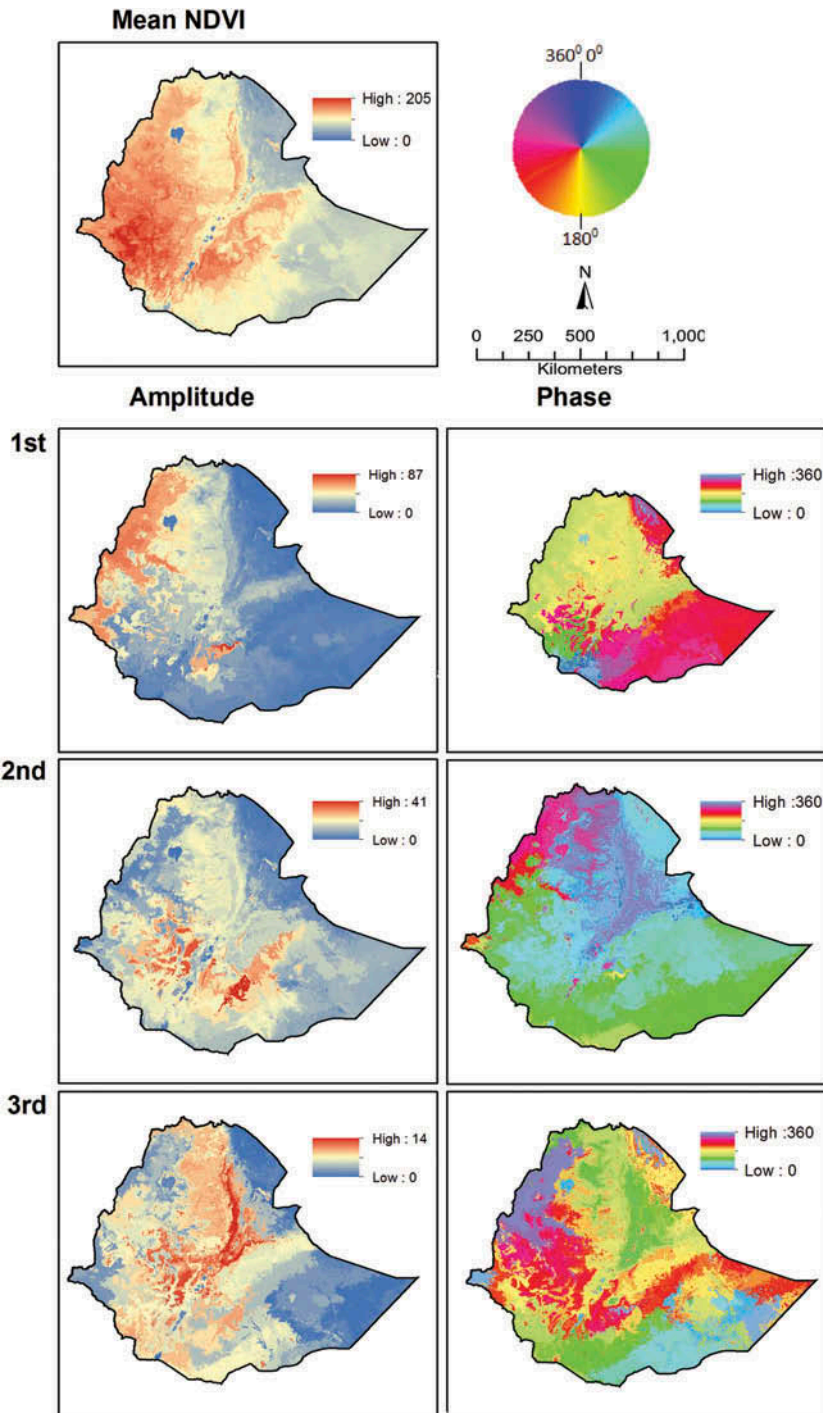


Figure 5. The seven HANTS parameters estimated from 36 (10 days interval) NDVI data covering one year. Higher values are displayed in red, lower values in blue, and intermediate values in yellow, using linear minimum–maximum stretching for amplitudes. The phase angle is represented with rainbow colour, the beginning and end phase angle being represented by the same blue colour. The phase values use 1 April as the reference date of the year.

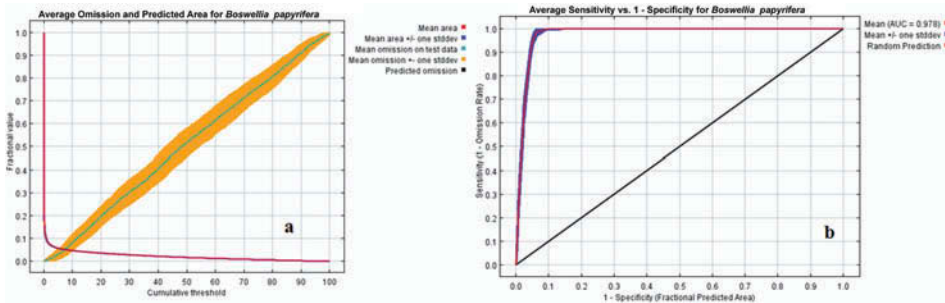


Figure 6. Average omission and predicted area (a), and Sensitivity versus 1-Specificity (b) for *B. papyrifera* after 100 runs (replicates). From both plots the omission test samples forms a very good match with the predicted omission rate. The omission rates for test data were drawn from within the MAXENT distribution. Stdev stands for standard deviation.

plot (Figure 6a, average of 100 replicates) with the choice of cumulative threshold shows that the omission of test samples matches the predicted omission rate for test data drawn randomly from the MAXENT distribution very well. The average area under the ROC curve (AUC, Figure 6b) for both training and testing is almost identical. The average test AUC for the 100 replicate runs is 0.978, and the standard deviation is 0.002.

The average relative contribution of the predictors of *B. papyrifera* distribution varied (Table 1). The model indicated that the contribution of the second harmonic phase was highest (35.8%), followed by elevation (30.8%), and first amplitude (27.4%). The first harmonic phase had the lowest contribution (0.1%). Using ‘permutation importance’ as well as ‘Jackknife regularized training gain analysis’ for the MAXENT model, the relative contribution of the environmental factors was assessed after checking for collinearity. The environmental variable with highest gain when used in isolation is the 2nd phase, which therefore appears to have the most useful information by itself. The environmental variable that most decreases the gain when omitted is elevation, which therefore appears to have the most information that is not present in the other variables.

The output probability of occurrence values was segmented into seven equal area categories to represent the ranges of *B. papyrifera* distribution (Figure 7). The deep reddish colour represents high probability areas, whereas the light grey colour represents low, and the light yellow represents intermediate probability of occurrence areas. For

Table 1. The relative contribution of factors to *B. papyrifera* tree distribution in Ethiopia.

Predictor variable	% contribution
2nd phase	35.8
Elevation (DEM)	30.8
1st amplitude	27.4
2nd amplitude	2.8
Slope	1.8
Mean NDVI	0.9
3rd amplitude	0.4
3rd phase	0.3
1st phase	0.1

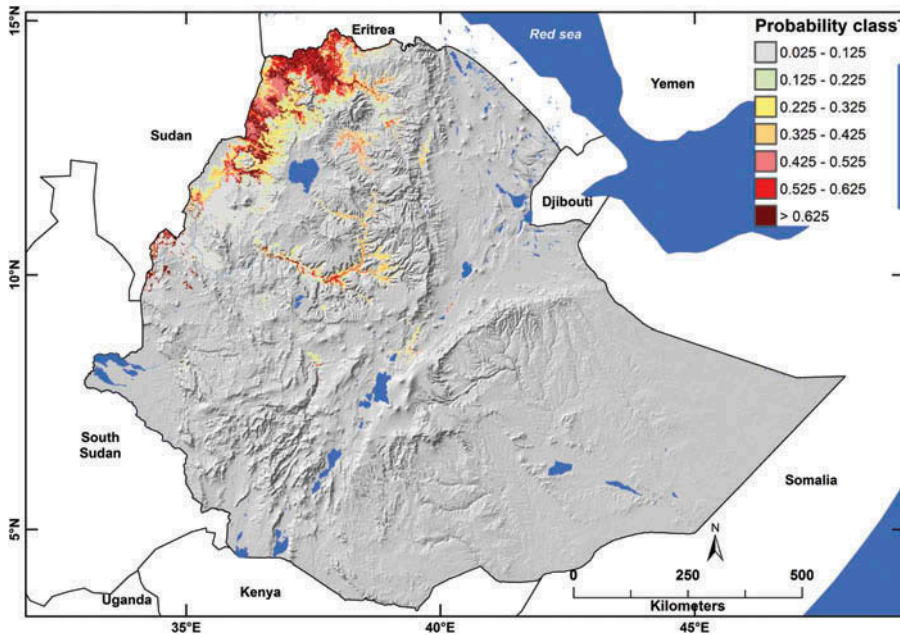


Figure 7. Distribution map of *B. papyrifera* in Ethiopia overlaid on an SRTM DEM-derived hillshade map. The species has a high (probability of) occurrence in the north and north-western part of Ethiopia.

better visualization, the distribution map is overlaid on hillshade derived from the SRTM DEM. The mapped *B. papyrifera* areas are either situated in deep gorges or represent low lands of North and North-West Ethiopia. The areal extent of the seven presence probability classes is provided in Table 2. When plotted against the probability distribution map classes (Figure 8), the presence-only observation points for *B. papyrifera* show a positive correlation between the number of *B. papyrifera* counts and the MAXENT probability values ( $F(6, 545) = 7.65, p < 0.001$ ).

The accuracy comparison of the conventional and NDVI-clustered maps with the distribution map is shown in Table 3. The overall accuracy of the distribution map

Table 2. The total estimated area occupied by *B. papyrifera* in Ethiopia.

S. no.	Predicted class description	Class range	Area (km <sup>2</sup> )	Area (%)
1	None	<0.03	1,029,920	90.98
2	Very low	0.03–0.13	34,869	3.08
3	Low	0.14–0.23	12,886	1.14
4	Relatively low	0.24–0.33	13,505	1.19
5	Moderate	0.34–0.43	11,048	0.98
6	Relatively high	0.44–0.53	12,407	1.10
7	High	0.54–0.63	9046	0.80
8	Very high	>0.63	8398	0.74
Ethiopia total			1,132,079	100.00
<i>B. papyrifera</i> area above 0.5 probability			19,628	1.73
<i>B. papyrifera</i> area between 0.03 and 0.5 probability			82,531	7.29

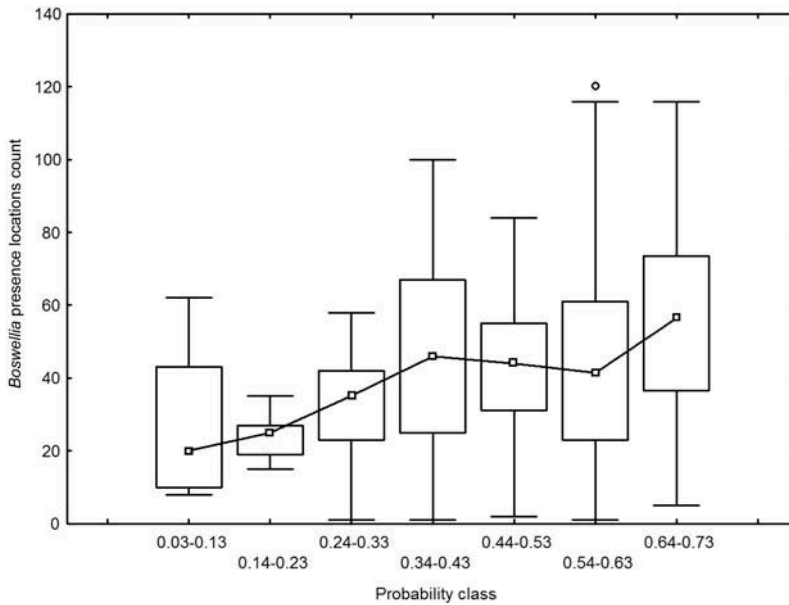


Figure 8. Relationship between *B. papyrifera* probability distribution classes and presence location counts. The highest probability classes contain high presence counts ( $F(6, 545) = 7.65, p < 0.001$ ). The line connects the median points of the classes.

Table 3. Accuracy comparison of the 'conventional' and NDVI-clustered map with the probability distribution map for *B. papyrifera* in Ethiopia.

<i>B. papyrifera</i> maps	Distribution map	
	Overall accuracy	Kappa
Conventional map	0.90	0.54
NDVI-clustered map	0.93	0.63

(Figure 7) is 90% compared to the conventional map (Figure 10), and 93% compared to the NDVI-clustered map (Figure 4). The Kappa result of the distribution map shows moderate agreement (0.54) with the conventional map, and substantial agreement (0.63) with the NDVI-clustered map.

#### 4. Discussion

Parameterization of classified hyper-temporal SPOT-NDVI datasets based on the HANTS algorithm is useful for mapping and modelling species distributions.

NDVI images integrate the effects of many environmental variables. Here it is demonstrated that classifying hyper-temporal NDVI data provides useful information about the ecology of a species across broader geographic space and can capture the temporal dimension of hyper-temporal images. Unsupervised classification of these datasets using the ISODATA algorithm resulted in 140 NDVI classes. This way, uniform land

use land cover units were successfully classified (de Bie *et al.* 2008, 2011, Khan *et al.* 2010). The main advantage of this classification is that similar NDVI units are grouped together, resulting in fewer data (data-reduction) and reduced noise. Moreover, it provides a first indication of where *B. papyrifera* trees occur. Significant positive relationships were reported between variation in species composition and NDVI (He *et al.* 2009).

The HANTS algorithm is successful not only in removing clouds from hyper-temporal NDVI datasets, but also in extracting NDVI shape parameters that correspond to the amount of photosynthetically active vegetation (PAV) greenness and the seasons. Ethiopia has eight generalized agro-climatic (ACZ) regions (Figure 9a) corresponding to crop-growing seasons, which are extremely variable (Figure 9b) from site to site (Hurni 1998). The NDVI signature in these areas varied from unimodal to bimodal. The phase angle indicates a change in the time (start and end) of PAV greenness; change within a single amplitude (frequency) indicates a change in vegetation condition such as the amount and duration of PAV greenness; and differences in amplitude (from one to two frequencies) indicate changes in PAV greenness mode (changes in season). *Boswellia papyrifera* areas depicted a single frequency with a slightly variable peak and width in PAV greenness (Figure 9b). This was in line with reports by Jakubauskas *et al.* (2002). These *B. papyrifera* areas were successfully parameterized into seven Fourier components, which were used in MAXENT together with elevation and slope for *B. papyrifera* tree potential distribution mapping in Ethiopia.

The conventional map of *B. papyrifera* available in Ethiopia (FWC 1984) is crude and produced at a small scale (1:9,500,000). The north side (e.g. Axum and Adwa areas) of the conventional map covers a broad area and extends into the highlands, while the north-west side (Metema) covers a very narrow area that lacks current presence records. The presence-only data indicate that the distribution of the species is restricted to between 500 and 1800 m (Figure 3) and predominantly between 700 and 900 m. *Boswellia papyrifera* was reported to be found in large amounts in Tigray, Gondar, Wollega and Gojam and thinly spread in Wollo and Shewa province in Ethiopia (Girma 1998) but not mapped (Figure 2 shows approximate locations). The new *B. papyrifera* map produced here (Figure 7) using MAXENT confirms the report. Tadesse *et al.* (2007) estimated the total cover of *B. papyrifera* in Ethiopia to be 15,000 km<sup>2</sup> (again unsupported by map). The reported area is somewhat similar to our potential distribution map of above the 0.5 probability cut-off (19,628 km<sup>2</sup>).

Mapping *B. papyrifera* has never been easy, mainly because it grows amongst other species in natural environments that are difficult to access. Different environmental variables derived from satellite sensors including MODIS, QSCAT, SRTM, and TRMM allow an increased understanding of species distribution, land-cover classification, and near real-time conservation planning across multiple spatial scales (Gillespie *et al.* 2008) using MAXENT (Saatchi *et al.* 2008). MAXENT improved the conventional map of *B. papyrifera* through the use of presence-only wide area data coverage, the main predictors being the 2nd phase harmonics (35.8%), elevation (30.8%), 1st amplitude (27.4%), and 2nd amplitude (2.8%). MAXENT has built-in routines that provide threshold-dependent and threshold-independent tests of its performance by examining the omission rate and the area under the ROC curve (AUC). The 'Subsample' and 'Crossvalidate' options in MAXENT produced similar high AUC results. AUC values tend to be higher for species with narrow ranges, relative to the study area described by environmental data (Phillips *et al.* 2004). Results indicated that meaningful model prediction is obtained concerning *B. papyrifera* distribution. In addition, the comparison between the presence-only records and the computed probability values suggest an increase in the *B. papyrifera* probability

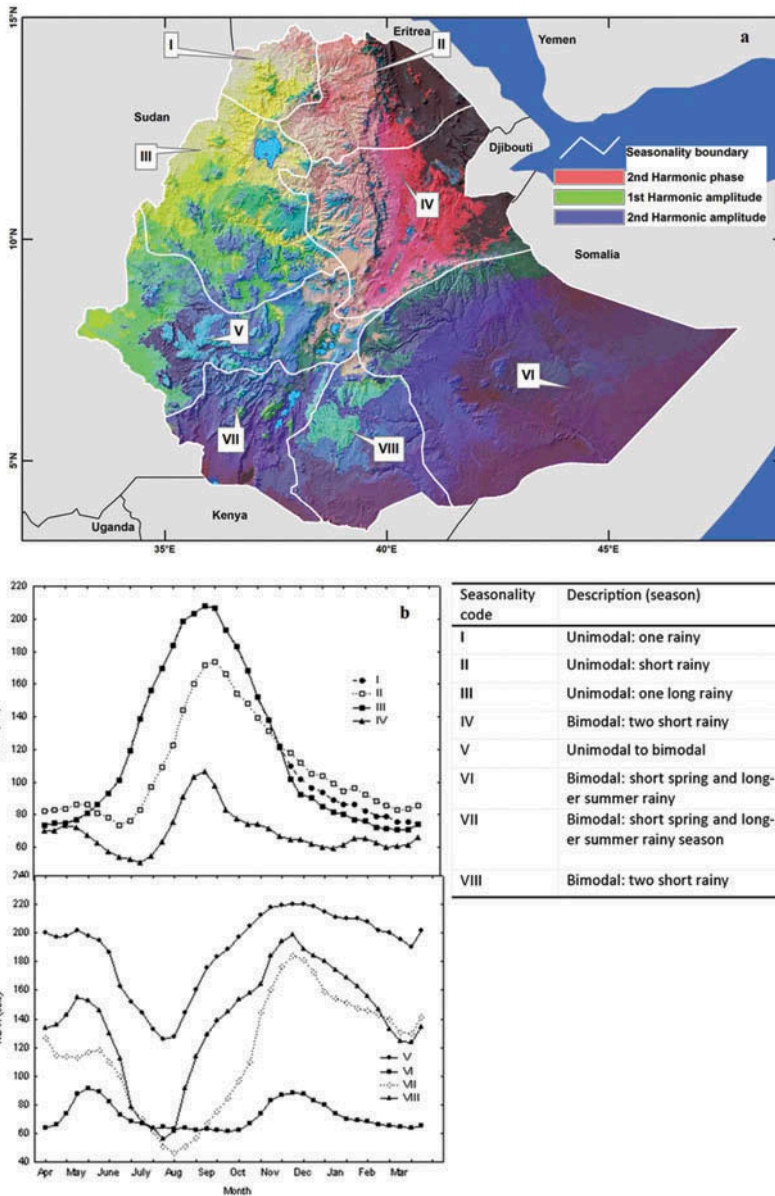


Figure 9. Seasonality map (adapted from Hurni 1998) overlaid on RGB map of the top three harmonic predictors (red = 2nd harmonic phase, green = amplitude of 1st harmonic, and blue = amplitude of 2nd harmonic) from the MAXENT model (a). The NDVI signatures extracted from the exact locations indicated in (a) that correspond to seasonality (b). It is shown that *B. papyrifera* predominantly occurs in the unimodal medium/long rainy season units shown in bold (units I, II, and III).

of occurrence. Moreover, comparison of the conventional (Figure 10) and the NDVI-clustered map (Figure 4) with the distribution map (Figure 7) shows moderate and substantial (Kappa) agreement, respectively. One of the limitations of the conventional



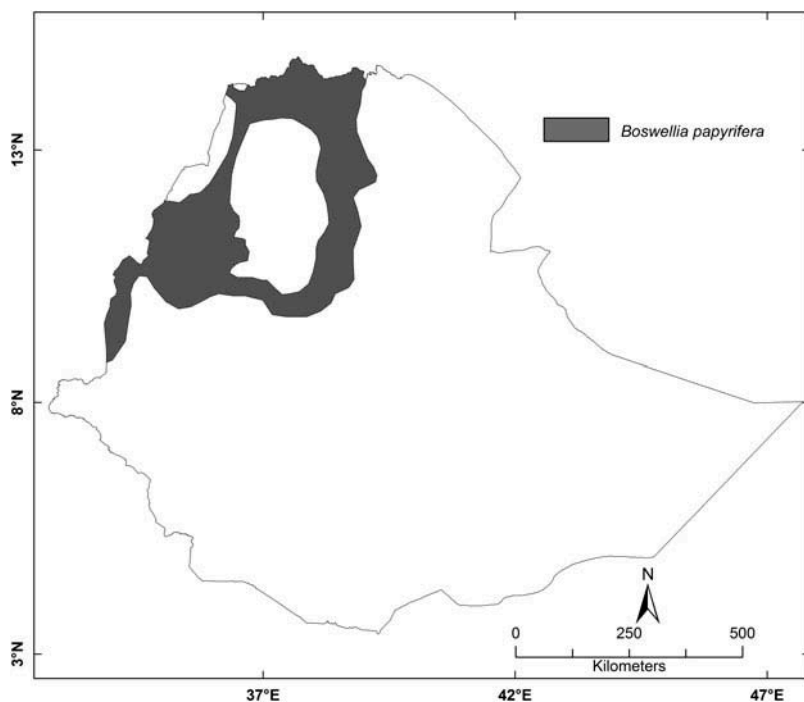


Figure 10. Distribution map of *B. papyrifera* in Ethiopia (digitized from Forestry and Wildlife Conservation DAUTH; Forest Inventory Surveying and Management Planning, October 1984, scale 1:9,500,000).

and the NDVI-clustered map units is that it assumes the species to occur uniformly in the map unit. The distribution map, however, does not assume uniformity. Instead, the probability of *B. papyrifera* occurrence is computed for each pixel.

Thus, classifying hyper-temporal NDVI images and extracting signature parameters using HANTS captures the PAV greenness behaviour of a species and its environment, consequently improving mapping of the species distribution. The overall accuracy of the species map for *B. papyrifera* is very high (90%). For the first time an operational (management) quality map of this remarkable and commercially important species is available. Such a map allows natural resource managers concerned with the sustainable utilization and conservation of *B. papyrifera* to plan for the harvesting and regeneration of this important species.

### Acknowledgement

We would like to thank Wouter Verhoef for his technical support on HANTS software.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

We are grateful for the financial support rendered by the WOTRO, The Netherlands fellowship program, The Netherlands.

## ORCID

Atkilt Girma  <http://orcid.org/0000-0001-7081-9503>

## References

- Adamson, P.B., 1969. Personal hygiene and cosmetics in the ancient Near East. *Soap, Perfumery & Cosmetics*, 42, 6.
- Ali, A., et al., 2013. Mapping land cover gradients through analysis of hyper-temporal NDVI imagery. *International Journal of Applied Earth Observation and Geoinformation*, 23, 301–312. doi:10.1016/j.jag.2012.10.001
- Atta-ur-Rahman, et al., 2005. Bioactive constituents from *Boswellia papyrifera*. *Journal of Natural Products*, 68, 5.
- Azene, B.-T., 2007. *Useful trees and shrubs for Ethiopia: identification, propagation and management for 17 agroclimatic zones*. Nairobi: RELMA in ICRAF project, World Agroforestry Centre-Eastern Africa Region Programme.
- BoANR, 2003. *Assessment of spatial distribution, regenerating capacity, degradation and development prospects of Boswellia papyrifera in Kafta-Humera*. Mekelle: Tigray Bureau of Agriculture and Natural Resources.
- BoANR, 2004. *Assessment of spatial distribution, regenerating capacity, degradation and development prospects of Boswellia papyrifera in Tahitay Adiabo*. Mekelle: Bureau of Agriculture and Natural Resources.
- Botkin, D.B., 1990. *Discordant harmonies: a new ecology for the twenty-first century*. Oxford: Oxford University Press.
- Congalton, R.G. and Green, K., eds., 2009. *Assessing the accuracy of remotely sensed data: principles and practices*. 2nd ed. New York, NY: CRC Press.
- Convertino, M., et al., 2014. Untangling drivers of species distributions: global sensitivity and uncertainty analyses of MaxEnt. *Environmental Modelling & Software*, 51, 296–309. doi:10.1016/j.envsoft.2013.10.001
- Cooley, J.W. and Tukey, J.W., 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19, 297–301. doi:10.1090/S0025-5718-1965-0178586-1
- CSA, 2000. *Statistical abstracts*. Addis Ababa: Ethiopian Central Statistical Authority (CSA).
- de Bie, C.A.J.M., Venus, V., and Skidmore, A.K., 2011. Improved mapping and monitoring with hyper-temporal imagery. In *Proceedings of the international workshop on advanced use of satellite- and geo-information for agricultural and environmental intelligence*, 2–4 March 2011 Tsukuba. Tsukuba: National Institute for Agro-Environmental Sciences (NIAES), 122–133.
- de Bie, C.A., et al., 2008. Hyper-temporal image analysis for crop mapping and change detection. In *The international archives of the photogrammetry, remote sensing and spatial information sciences*, Beijing, Vol. XXXVII, Part B7.
- de Bie, C.A.J.M., et al., 2012. LaHMa: a landscape heterogeneity mapping method using hyper-temporal datasets. *International Journal of Geographical Information Science*, 26 (11), 2177–2192. doi:10.1080/13658816.2012.712126
- Edwards, S., Jonquil, W.A., and Sandford, A.E., 1976. *Some wild flowering plants of Ethiopia*. Addis Ababa: Addis Ababa University.
- Elith, J., et al., 2011. A statistical explanation of MAXENT for ecologists. *Diversity and Distributions*, 17 (1), 43–57. doi:10.1111/j.1472-4642.2010.00725.x
- Fernandez, M., et al., 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. *Biodiversity Informatics*, 6. doi:10.17161/bi.v6i1.3314
- Fichtl, R. and Admasu, A., 1994. *Honey bee flora of Ethiopia*. Weikersheim: DED/Margraf Verlag.
- Fung, T. and Siu, W., 2000. Environmental quality and its changes, an analysis using NDVI. *International Journal of Remote Sensing*, 21 (5), 1011–1024. doi:10.1080/014311600210407
- FWC, 1984. *Forest inventory surveying and management planning. Forestry and wildlife conservation D-AUTH: gum and resin producing species in Ethiopia*. Addis Ababa: Forestry and Wildlife Conservation, 1:9,500,000.
- Gebrehiwot, K., 2003. *Ecology and management of Boswellia papyrifera (Del.) Hochst. dry forests in Tigray, northern Ethiopia*. Doctoral Dissertation. Georg-August-University of Göttingen.

- Gebrehiwot, K., et al., 2003. Introducing *Boswellia Papyrifera* (Del.) Hochst and its non-timber forest product, frankincense. *International Forestry Review*, 5 (4), 6. doi:10.1505/IFOR.5.4.348.22661
- Gillespie, T.W., et al., 2008. Measuring and modelling biodiversity from space. *Progress in Physical Geography*, 32 (2), 203–221. doi:10.1177/0309133308093606
- Girma, A., et al., 2013. Photosynthetic bark: use of chlorophyll absorption continuum index to estimate *Boswellia papyrifera* bark chlorophyll content. *International Journal of Applied Earth Observation and Geoinformation*, 23, 71–80. doi:10.1016/j.jag.2012.10.013
- Girma, A., et al., 2015. Understanding *Boswellia papyrifera* tree secondary metabolites through bark spectral analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 30–37. doi:10.1016/j.isprsjprs.2015.03.010
- Girma, D., 1998. *Data collection and analysis for sustainable forest management in ACP countries – linking national and international efforts: non-wood forest products in Ethiopia*. Addis Ababa: EC-FAO PARTNERSHIP PROGRAMME (1998-2000).
- Guisan, A. and Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135 (2–3), 147–186. doi:10.1016/S0304-3800(00)00354-9
- Hadgu, K.M., 2008. *Temporal and spatial changes in land use patterns and biodiversity in relation to farm productivity at multiple scales in Tigray, Ethiopia*. PhD thesis. Wageningen University.
- He, K.S., Zhang, J., and Zhang, Q., 2009. Linking variability in species composition and MODIS NDVI based on beta diversity measurements. *Acta Oecologica*, 35 (1), 14–21. doi:10.1016/j.actao.2008.07.006
- Hurni, H., 1998. *Soil conservation research programme Ethiopia: agroecological belts of Ethiopia, explanatory notes on three maps at a scale of 1:1,000,000*. Ethiopia: Centre for Development and Environment University of Bern, Switzerland in association with The Ministry of Agriculture, 43.
- Immerzeel, W.W., Quiroz, R.A., and De Jong, S.M., 2005. Understanding precipitation patterns and land use interaction in Tibet using harmonic analysis of SPOT VGT-S10 NDVI time series. *International Journal of Remote Sensing*, 26 (11), 2281–2296. doi:10.1080/01431160512331326611
- Jakubauskas, M.E., Legates, D.R., and Kastens, J.H., 2001. Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering and Remote Sensing*, 67 (4), 461–470.
- Jakubauskas, M.E., Legates, D.R., and Kastens, J.H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture*, 37 (1–3), 127–139. doi:10.1016/S0168-1699(02)00116-3
- Khan, M.R., et al., 2010. Disaggregating and mapping crop statistics using hyper-temporal remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 12 (1), 36–46. doi:10.1016/j.jag.2009.09.010
- Khan, M.R., 2011. *Crops from space: improved earth observation capacity to map crop areas and to quantify production*. PhD thesis. University of Twente.
- Lasaponara, R., 2006. On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series. *Ecological Modelling*, 194 (4), 429–434. doi:10.1016/j.ecolmodel.2005.10.035
- Lemenih, M. and Teketay, D., 2003. Frankincense and Myrrh resources of Ethiopia: II medicinal and industrial uses. *Ethiopian Journal of Science*, 26 (2), 12.
- Marmion, M., et al., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15 (1), 59–69. doi:10.1111/j.1472-4642.2008.00491.x
- MoARD, 2007. *Revised agro-ecological zonations of Ethiopia*. Addis Ababa: Ministry of Agriculture and Rural Development (MoARD), Natural Resources Management Directorate.
- Nguyen, T.T.H., et al., 2012. Mapping the irrigated rice cropping patterns of the Mekong delta, Vietnam, through hyper-temporal SPOT NDVI image analysis. *International Journal of Remote Sensing*, 33 (2), 415–434. doi:10.1080/01431161.2010.532826
- Pearce, J.L. and Boyce, M.S., 2006. Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43 (3), 405–412. doi:10.1111/jpe.2006.43.issue-3
- Pearson, R.G., 2007. *Species' distribution modeling for conservation educators and practitioners*. Synthesis. American Museum of Natural History. Available from: <http://ncep.amnh.org>

- Phillips, S.J., *et al.*, 2004. A maximum entropy approach to species distribution modeling. In *Proceedings of the twenty-first international conference on machine learning*. Banff, Alberta, Canada: ACM, 83.
- Phillips, S.J., Anderson, R.P., and Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259. doi:10.1016/j.ecolmodel.2005.03.026
- Prasad, V.K., Anuradha, E., and Badarinath, K.V., 2005. Climatic controls of vegetation vigor in four contrasting forest types of India—evaluation from National Oceanic and Atmospheric Administration's Advanced Very High Resolution Radiometer datasets (1990–2000). *International Journal of Biometeorology*, 50 (1), 6–16. doi:10.1007/s00484-005-0268-0
- Roerink, G.J., Menenti, M., and Verhoef, W., 2000. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, 21 (9), 1911–1917. doi:10.1080/014311600209814
- Rulinda, C.M., Bijker, W., and Stein, A., 2010. Image mining for drought monitoring in eastern Africa using Meteosat SEVIRI data. *International Journal of Applied Earth Observation and Geoinformation*, 12, S63–S68. doi:10.1016/j.jag.2009.10.008
- Saatchi, S., *et al.*, 2008. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sensing of Environment*, 112 (5), 2000–2017. doi:10.1016/j.rse.2008.01.008
- Scott, J.M., *et al.*, 2002. *Predicting species occurrences: issues of accuracy and scale*. Washington, DC: Island Press.
- Sequist, J.W., Olsson, L., and Ardö, J., 2003. A remote sensing-based primary production model for grassland biomes. *Ecological Modelling*, 169 (1), 131–155. doi:10.1016/S0304-3800(03)00267-9
- Tadesse, W., Desalegn, G., and Alia, R., 2007. Natural gum and resin bearing species of Ethiopia and their potential applications. *Investigación Agraria: Sistemas y Recursos Forestales*, 167 (3), 11.
- USGS, 2004. *Shuttle Radar Topography Mission. 1 Arc Second scene SRTM\_u03\_n008e004*. Unfilled Unfinished 2.0. College Park, MD: Global Land Cover Facility, University of Maryland, February 2000. Available from: <http://www.landcover.org/data/srtm/>
- Verhoef, W., Menenti, M., and Azzali, S., 1996. Cover a colour composite of NOAA-AVHRR-NDVI based on time series analysis (1981-1992). *International Journal of Remote Sensing*, 17, 231–235. doi:10.1080/01431169608949001
- Verhoef, W., van der Kamp, A., and Koelemeijer, R., 2005. *Climate indicators from time series of NDVI images (CITISEN, Final Report)*. Delft: Netherlands Agency for Aerospace Programmes (NIVR), 28 p.
- Wang, Q., *et al.*, 2005. On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sensing of Environment*, 94 (2), 244–255. doi:10.1016/j.rse.2004.10.006
- WBISP, 2004. *Forest resources of Ethiopia*. Addis Ababa: Woody biomass inventory and strategic plan project (WBISP).
- Wiens, J.A., *et al.* 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences*, 106 (Supplement 2), 19729–19736.