Predictive Species Habitat Distribution Modelling of Indian Sandalwood Tree Using GIS

Nambi Manavalan Rajan[1]*, Shanmugam Jayalakshmi[2]

[1]Institute of Remote Sensing, Department of Civil Engineering, Anna University Chennai,
Tamil Nadu, India
[2]Institute of Remote Sensing, Department of Civil Engineering, Anna University Chennai,
Tamil Nadu, India

Abstract

The Indian sandalwood tree is a highly valuable aromatic wood resource. But its population is declining at an alarming rate that cannot be controlled through manually employed management techniques. Also, the limitations of the automated management techniques in practice, such as the drawbacks of models like MAXENT, GARP, etc., can dealt with the usage of the techniques used in this research. 'Strategy-oriented prescriptive process modelling' has been employed in this research so that the management techniques are brought under automation for quicker processing of algorithms. The outcome of this research is a toolbox that contains the 'machine-learned algorithms' for performing ecological niche modelling of *Santalum album* for any given study area, ensuring that all the necessary data (concerned with the study area) are fed as inputs so that the final site suitability map is obtained as the output. The results have been well-discussed in the form of tables in this manuscript. Upon validation (through ground truthing) of the final output we observed that the toolbox designed in this research is able to give 90.625% accuracy in the resulting maps it produces.

Keywords: ecological niche modelling, GIS, strategy-oriented prescriptive process modelling

Introduction

There is widespread agreement that global biodiversity is declining at an accelerated rate. The main threat to tropical forest biodiversity is habitat loss, particularly loss of forest cover. Rapidly changing climate, habitat fragmentation, invasion of alien species and pathogens, pollution, over-exploitation, and escalating human population are the most important factors responsible for ecosystem degradation worldwide that alter the structural and functional integrity of the ecosystems. Such alterations have brought approximately one fifth of the plant species to the brink of extinction.

Indian Sandalwood Tree

In this research, the Indian Sandalwood tree is considered as the study species. Its botanical name is *Santalum album*.

Sandal is a nationally protected resource in India. Despite the protection status the natural resources of sandal (particularly in India) are being indiscriminately exploited, perhaps because of its extremely high export

*e-mail: rnambimanavalan@gmail.com

value. Increasing demand for timber by a rapidly growing population has put high pressure on limited forest resources of sandalwood. The high value of the species has caused its past exploitation to the point making this tree's population vulnerable to extinction. Indian sandalwood still commands high prices for its essential oil, but due to the lack of sizable trees it is no longer used for fine woodworking.

Nearly 85% of the supply of Indian sandalwood comes from the southern states of Tamil Nadu, Karnataka, and Kerala [1]. The Indian sandalwood industry has been in the grip of significant crisis of sandalwood, in terms of its procurement, for decades. Excessive harvesting without replenishment of this invaluable resource has substantially reduced the sandalwood industry, resulting in a global shortage and the soaring of market prices.

The species is threatened by over-exploitation and degradation of habitat through altered land use; fire, agriculture, and land-clearing are the factors of greatest concern. In 1998 this species was recognized as "vulnerable" by the International Union for the Conservation of Nature's (IUCN) Red List, due to threats like the rapid spread of disease, fire, grazing by cattle, deer, and other animals (to which the sandalwood trees are extremely sensitive), and exploitation through illegal activity. Today the sandalwood tree still faces many threats from not only humans, in the form of exploitation and illegal harvesting, but also from natural forces. According to forestry consultant Dr. Anantha Padmanabha, "the tree's growth under natural forest conditions is very slow due to reasons like fire, grazing, and human intervention." It is feared that being associated with the severe exploitation of the natural populations of sandal, the genetic resources of the species may have also been affected. Such a loss of genetic resources may have far-reaching consequences on the improvement of sandal for its heartwood and oil quality.

To preserve this tree from over-exploitation, suitable cultivation practices at other places are necessary. Prediction and mapping of potential suitable habitats for this tree is critical for monitoring and restoring its declining native population. If suitable sites are identified, we can use the germplasm of this tree and after tissue culture, and cultivation can be done at modeled habitats enabling the expansion of Sandalwood tree plantations [1]. The scope of this research reaches only to the identification of suitable sites in the study area.

Predictive Species Habitat Distribution Modelling (PSHDM)

Predictive species habitat distribution modelling (PSHDM), otherwise called environmental niche modelling, refers to the process of using computer algorithms to predict the distribution of species in geographic space on the basis of a mathematical representation of their known distribution in environmental space (i.e., realized ecological niche). The environment is in most cases represented by climate data (such as temperature and precipitation), but other variables such as soil type, water depth, and land cover can also be used. These models get applied in several research areas such as conservation biology, ecology, and evolution. According to these models, the presence of a species is considered a physiological or mechanistic logic of favorable site factors, or in other words a causal function of the explanatory variables based on the niche requirements of a species.

Agro-Forestry Database

According to the agro-forestry database (Version 4.0) of the World Agroforestry Centre (Nairobi, Kenya), altitude, annual rainfall, mean annual temperature, pH, and soil type are the required factors for *Santalum album's* survival. These are the abiotic factors, and a suitable host plant presence is the biotic factor that is required for the absorption of water and nutrients (especially nitrogenous compounds). The host plants require the abiotic factors as required by *Santalum album*, and it also needs soil moisture content as its additional abiotic factor [2].

The agro-forestry database provides a comprehensive list of variables that favor the growth of *Santalum album*. For finding their respective value ranges, dependency analysis is performed. Dependency analysis is used for finding the percentage of influence of each factor which is computed in this research by an algorithm different from those of the currently used niche models. This research uses a previous map that depicts the tree's distribution (obtained in shapefile format) in 2014 (Fig. 1), the source of which is the ENVIS Center on Floral diversity.
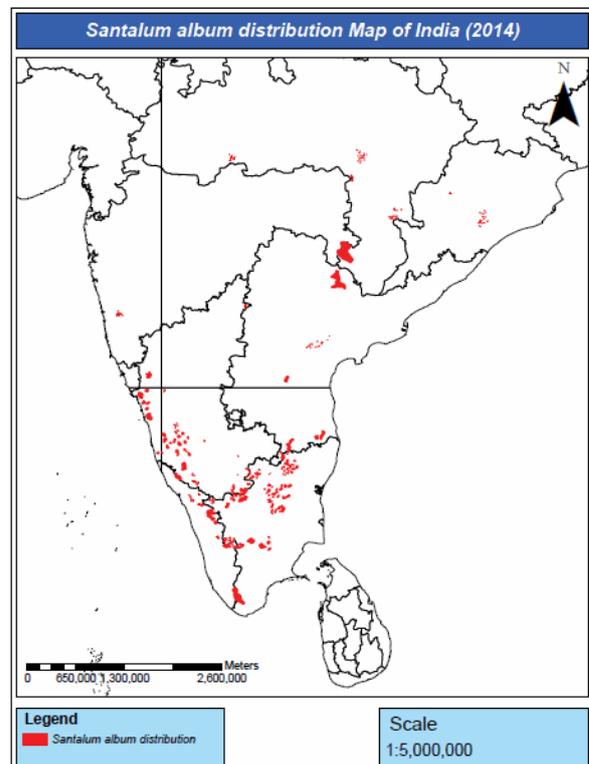

Fig. 1. *Santalum album* distribution map in 2014.

Table 1. Sources of the obtained data.

Factor	Type of data	Year of data	Data source	Reference
Mean annual temperature		2014	http://www.worldclim.org/current	[3]
Altitude		2014	http://srtm.csi.cgiar.org	[4]
Mean annual rainfall	Parametric	2014	http://www.worldclim.org/current	[3]
Soil moisture		2014	http://rda.ucar.edu/	[5]
Soil pH		2014	http://www.isric.org/content/soilgrids	[6]
Soil type	Non-parametric	2014	http://www.isric.org/content/soilgrids	[6]

The identified variables are checked for their values (pertaining to 2014) in the regions present within the delineated boundaries of this map. Clustering of classes has been done using fuzzy logic to maximize accuracy. Then site suitability analysis has been done in the study area for the present conditions. The site suitability maps for the study area thus prepared are validated by direct field visits. In order to make these algorithms as 'machine-learning algorithms,' process modelling is done. The GIS software used for all operations and analyses in this research is ArcGIS 9.3.

Material and Methods

Ecological Niche Modelling

This research perceives ecological niche modelling of *Santalum album* in any study-area as the combination of:
1. Dependency analysis.
2. Membership value estimation.
3. Site suitability analysis.
4. Strategy-oriented prescriptive process modelling.

Table 1 lists the sources of both parametric and non-parametric data used in this research. These data were downloaded from the websites mentioned in the table.

Dependency Analysis

Dependency analysis is performed using the previous distribution map (Fig. 1) of *Santalum album*. It is performed with two types of factors (parametric and non-parametric) – each having different algorithms developed in this research.

Case I: For Parametric Factors

Dependency analysis in the case of parametric factors first includes the following three operations:
1. Intra-layer zonal weightage calculation.
2. Clustering using fuzzy logic.
3. Contributive weightage calculation.

Intra-layer Zonal Weightage Calculation

Using the boundaries in the map, analysis has been done to find the parametric range of values for each factor that are likely suitable for the tree's survival.

In the first step the values of each layer (such as altitude, mean annual temperature, mean annual rainfall, soil pH, and soil moisture) are normalized using equations 4.1, 4.2, and 4.3.

$$\mu = \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)$$

Eq. 4.1

…where 'x' is a parametric value of each layer, 'N' is the number of parametric values in the concerned layer, and 'μ' is the mean of all 'x' values which is obtained as the result of this equation. Symbol 'i' ranges from the 1st to Nth value of 'x'.
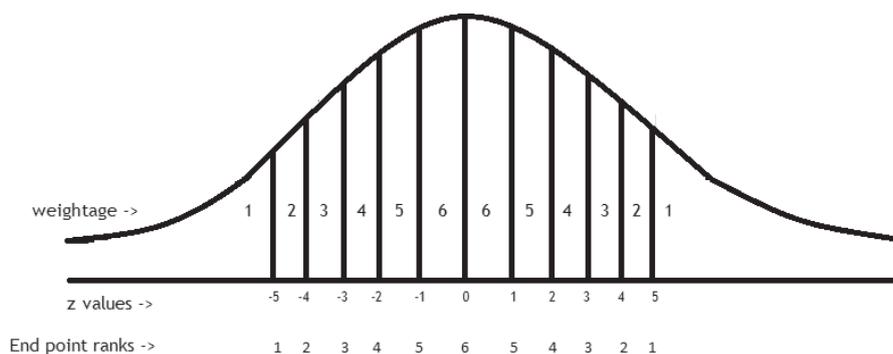

Fig. 2. Normal curve showing 'z' values and their corresponding weightages.

$$\sigma = \sqrt{\frac{1}{N} \left(\sum_{i=1}^{N} (x_i - \mu)^2 \right)}$$
Eq. 4.2

...where 'σ' is the standard deviation of 'x' values from the mean 'μ', and the value of 'μ' is obtained from equation 4.1.

$$z = \frac{(x-\mu)}{\sigma}$$
Eq.4.3

...where 'z' is the normalized value. Values of 'μ' and 'σ' are obtained from equations 4.1 and 4.2. respectively.

Ranking the 'z' value ranges is then done (as shown in Fig. 2) by allotting suitable weightages. This enables the calculation of intra-layer zonal weightages which can be substituted in equation 4.8. Highly frequent values tend to get crowded near the central region of the bell-shaped curve and less frequent values form the tail of the bell-shaped curve at both sides (as shown in Fig. 2). More priority in assigning higher weightages has been given to the zones with values that are present nearer the central region of 'z' value range. This is because *Santalum album* tends to get located more in areas that have parametric factor-values of minimal difference between the values themselves [7]. This means that higher weighted parametric zones within each factor layer are more preferable for the survival of *Santalum album*.

Clustering with Fuzzy Logic

Now it can be noted that for all the parametric values within a 'z' value zone, a similar weightage has been provided throughout that zone. This indicates that if these 'crisp' weightage clusters are proceeded with further, the ecological niche modelling performed will have discrete zonations of site suitability indices, making the final map less accurate. So in order to get an accurate niche modelling of *Santalum album*, 'fuzzy logic' is applied. The Fuzzy logic is implemented using equations 4.4, 4.5, 4.6, and 4.7.

Algorithm

Let (z_1, z_2) denote the extreme 'z' values at both ends of a zone. Let (e_1, e_2) denote the corresponding discrete end-point ranks of z_1 and z_2. Fig. 2 shows the corresponding end-point ranks of all zones end values. Let 'c' indicate the finally computed contributive index of 'z' in the zone of concern. Let 'w' be the weightage of the zone. Let 'c_{min}' denote the minimum contributive index. This means that the value of 'c' should always be greater than the 'c_{min}' value. This Algorithm as a whole helps in maximizing the accuracy of the final output map.

Case I

When $z \leq 0$, the 'c_{min}' value is calculated from equation 4.4.

$$c_{min} = \frac{e_1}{w}$$
Eq. 4.4

...where e_1 is the end-point rank of z_1, and 'w' is the weightage of the zone of concern. The value of c_{min} obtained from equation 4.4 is substituted in equation 4.5:

$$c = [(z - z_1) \times (1 - c_{min})] + c_{min}$$
Eq. 4.5

...where 'z' is the normalized value calculated from equation 4.3. Adding 'c_{min}' value in equation 4.5 ensures that the value of 'c' is always greater than the 'c_{min}' value. Thus Equation 4.5 gives the value of 'c' if $z \leq 0$.

Case II

When $z > 0$, the 'c_{min}' value is calculated from equation 4.6.

$$c_{min} = \frac{e_2}{w}$$
Eq. 4.6

...where 'e_2' is the end-point rank of 'z_2' and 'w' is the weightage of the zone of concern. The value of 'c_{min}' obtained from equation 4.6 is substituted in equation 4.7.

$$c = [(z_2 - z) \times (1 - c_{min})] + c_{min}$$
Eq. 4.7

...where 'z' is the normalized value calculated from equation 4.3. Adding 'c_{min}' value in equation 4.7 ensures that the value of 'c' is always greater than the 'c_{min}' value. Thus Equation 4.7 gives the value of 'c' if $z > 0$.

Case II: For Non-Parametric Factor

Soil-type factor values are non-parametric. The soil type data are based on the taxonomic soil type classification based on the World Reference Base (WRB) system. The areal extents of all the soil types in the *Santalum album* distribution were computed using ArcGIS 9.3. The soil type that spreads to the maximum extent is found to be Luvisol.

The areal extent of Luvisol is considered equivalent to the maximum Contributive Index of 6. Now the 'c' values for all other soil types are computed using the 'direct proportionality' concept as the areal extent of a soil type is directly proportional to the membership value. The 'c' values thus computed are listed in Table 2.

Contributive Weightage Calculation

This procedure is followed for five parametric-factors: altitude, annual rainfall, mean annual temperature, pH, and soil moisture content. The maps thus prepared for each factor have undergone 'overlay analysis,' and by using 'single output map algebra expressions,' all the 'contributive indices' of the layers are added together.

Table 2. Soil types and their respective contributive indices.

Soil number	Soil type	Area coverage (km²)	Contributive index ('c')
1.	Luvisol	0.398429	6.0000
2.	Alisol	0.030232	5.0871
3.	Regosol	0.354129	3.6272
4.	Acrisol	0.316433	2.8067
5.	Vertisol	0.087412	2.6727
6.	Nitisol	0.048922	1.8862
7.	Cambisol	0.437767	1.3052
8.	Lixisol	0.023634	0.5866
9.	Andosol	0.035325	0.1226
10.	Podzol	0.296963	0.0226
11.	Ferralsol	0.186047	0.0132
12.	Kastanozems	0.018305	0.0132
13.	Leptosol	0.576661	0.0113
14.	Solonchaks	0.049144	0.0057
15.	Fluvisol	0.089761	0.0019
16.	Histosol	0.006738	0.0019

Now each layer's percent contribution to this sum is determined, which in turn is the contributive weightage (w_c). The factors and their corresponding contributive weightages have been tabulated in Table 3, from which it can be concluded that mean annual temperature is the most important factor for the survival of *Santalum album*. Thus all the operations in 'dependency analysis' have been accomplished. The analyses and operations that needed to be undertaken in the study-area of interest are discussed in detail in the next two headings.

Membership Value Estimation

Membership values are estimated for each pixel in the study-area considering each factor. Two cases arise in the estimation process. The first case deals with the parametric

Table 3. Factors and their corresponding contributive weightages.

Soil number	Factor	Factor's contributive weightage (w_c)
1.	Mean annual temperature	0.263392239019
2.	Altitude	0.242469531299
3.	Mean annual rainfall	0.013959986711
4.	Soil moisture	0.064255526186
5.	Soil pH	0.203901440089
6.	Soil type	0.166666666762

factor and the second case deals with the non-parametric factor.

Case I: For Parametric Factor

In this case, two operations are performed. They are:
1. Normalization of values
2. Clustering using fuzzy logic

Normalization of Values

The Factor-data values are normalized using equations 4.1, 4.2, and 4.3.

Clustering Using Fuzzy Logic

Let (z_1, z_2) denote the extreme 'z' values at both ends of a zone. Let (e_1, e_2) denote the corresponding discrete end-point ranks of z_1 and z_2. Fig. 2 shows the corresponding end-point ranks of all zone end values. Let 'm' indicate the finally computed membership value of 'z' in the zone of concern. Let 'w' be the weightage of the zone. Let 'm_{min}' denote the minimum membership value. This means that the value of 'm' should always be greater than the "m_{min}' value. This algorithm as a whole helps in maximizing the accuracy of the final output map.

Case I

When $z \leq 0$, the 'm_{min}' value is calculated from equation 4.8.

$$m_{min} = \frac{e_1}{w} \qquad \text{Eq. 4.8}$$

…where 'e_1' is the end-point rank of 'z_1' and 'w' is the weightage of the zone of concern. The value of 'm_{min}' obtained from equation 4.8 is substituted in equation 4.9.

$$m = [(z - z_1) \times (1 - m_{min})] + m_{min} \qquad \text{Eq. 4.9}$$

…where 'z' is the normalized value calculated from equation 4.3. Adding 'm_{min}' value in equation 4.9 ensures that the value of 'm' is always greater than the 'm_{min}' value. Thus Equation 4.9 gives the value of 'm' if $z \leq 0$.

Case II

When $z > 0$, the 'm_{min}' value is calculated from equation 4.10.

$$m_{min} = \frac{e_2}{w} \qquad \text{Eq. 4.10}$$

…where 'e_2' is the end-point rank of 'z_2' and 'w' is the weightage of the zone of concern. The value of 'm_{min}' obtained from equation 4.10 is substituted in equation 4.11.

$$m = [(z_2 - z) \times (1 - m_{min})] + m_{min}$$

Eq. 4.11

…where 'z' is the normalized value calculated from equation 4.3. Adding "m_{min}' value in equation 4.11 ensures that the value of 'm' is always greater than the "m_{min}' value. Thus Equation 4.11 gives the value of 'm' if $z>0$.

Case II: For Non-Parametric Factor

From Table 2, the respective membership values of soil types present in the study-area are found out. Thus membership values are obtained for each factor (Case I or Case II) concerned with the study-area.

Site Suitability Analysis

The term 'Site Suitability Index' is used to denote the degree of suitability of an area in a map which can be computed from equation 4.12. The values of 'm,' after being calculated from equations 4.9 or 4.11 (for parametric factors), are substituted in equation 4.12, where they are multiplied with their respective contributive weightages (w_c) to obtain the required '*Site Suitability Index*'.

$$[Site\ Suitability\ Index] = (m \times w_c)$$

Eq. 4.12

Now the prepared site suitability maps for each factor-layer undergo 'overlay analysis,' in which all the site suitability indices computed for each factor in a pixel are added together to obtain the final site suitability index for the concerned pixel. Likewise, this procedure is repeated for all the pixels of the raster image to obtain the final site suitability map of the study-area.

Strategy-Oriented Prescriptive Process Modelling

Strategy-oriented prescriptive process modelling is defined as the simulation of procedures involved in an operation with the goal to establish guidelines and behaviour patterns which, if followed, would lead to the desired process performance in other similar cases of study. The patterns may range from strict enforcement to flexible guidance. This technique may prove to be useful in applying tactics and methods used in a situation to any other similar case studies, those of which also have the same type of desired target outputs [8].
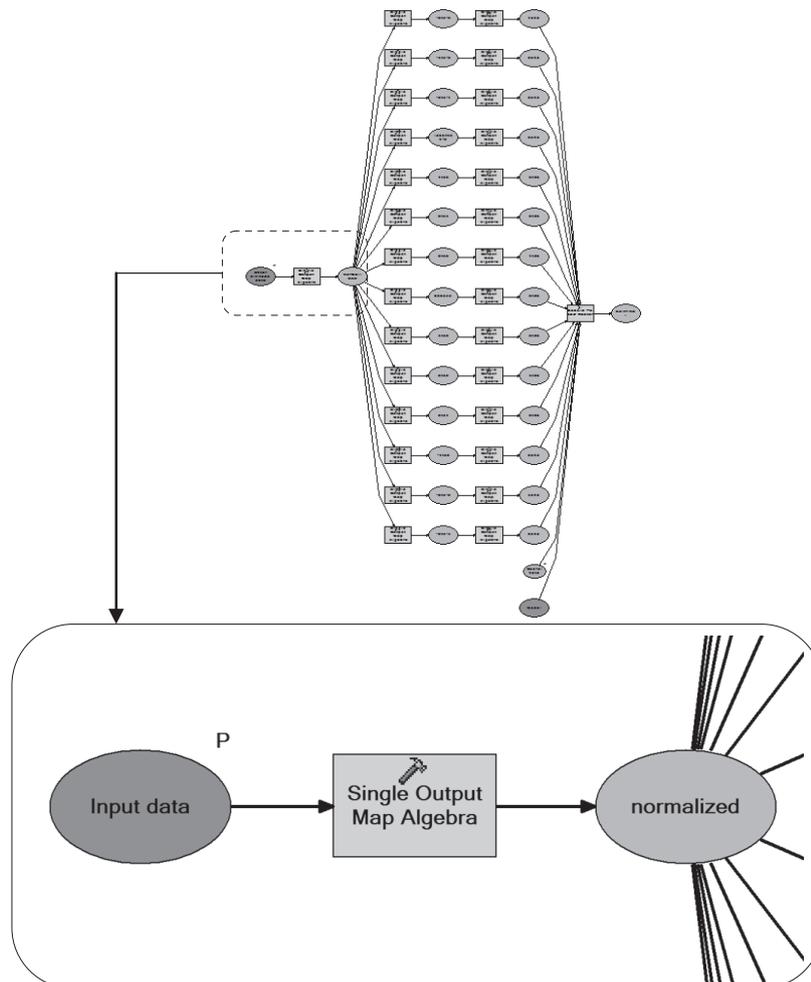

Fig. 3. The model structure and the zoomed view of the first module of the model.
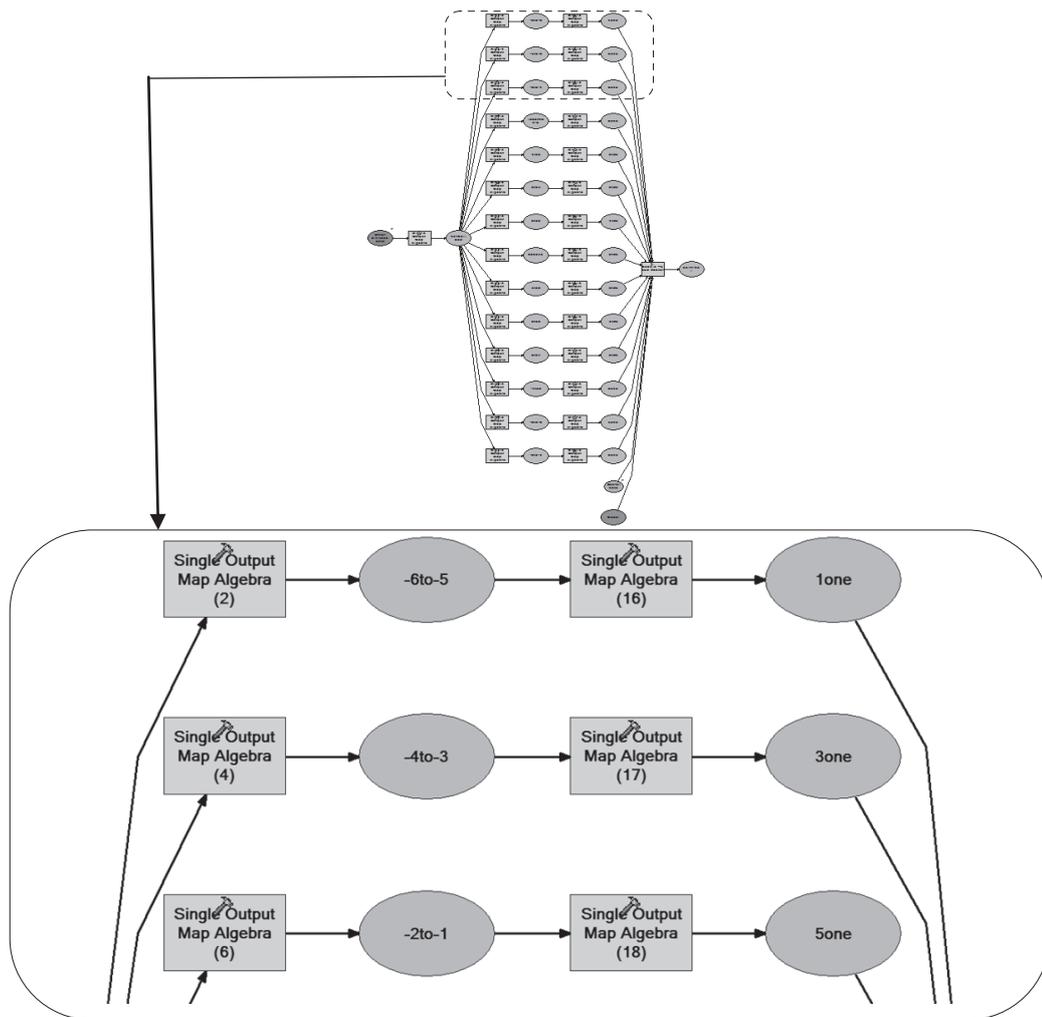
Fig. 4. The model structure and the zoomed view of the 2ⁿᵈ, 3ʳᵈ, and 4ᵗʰ modules of the model.

It is helpful in making the so-far discussed algorithms a combined 'machine-learned algorithm.' This enables automatic execution of all the operations for the six factors concerned with any study-area. This modelling operation is done using the ArcGIS 9.3 'model builder.'

The model developed for each factor is shown in Fig. 3. This model is responsible for the computation of membership values for all pixels in the study-area pertaining to the parametric factors alone. As shown in the figure, the model has 16 modules. The figure also shows the zoomed view of the first module. Module 1 is designed such that it gets the parametric factor data values of any study area as the input from the user and normalizes the data using equations 4.1, 4.2, and 4.3 using the 'single output map algebra expression'.

There are two sub-modules for each of the modules from the 2ⁿᵈ module to the 15ᵗʰ module. These modules (i.e., from module 2 to module 15) are concerned with the extraction of the 'z' values in each zone separately and are also concerned with the calculation of membership values for the 'z' values (computed by module 1).

Thus these modules prepare dismantled rasters, where each pixel denotes the degree of membership (i.e., the membership value) in the final site suitability index to be computed. These dismantled rasters are then assembled by module 16, which is the final operation of this model. The final outputs of this model are the membership value maps concerned with all the parametric factors.

Fig. 10 shows the model that is responsible for the computation of membership values concerned with the non-parametric factor (i.e., soil type). Thus these two models perform the function of 'membership value estimation.' The outputs are then multiplied with their corresponding contributive weightages (w_c) to obtain site suitability maps for each factor (both parametric and non-parametric). These maps are then added by overlay operation to obtain the final site suitability map of the study-area for the survival of *Santalum album*. The model that performs this operation is shown in Fig. 10.

Fig. 4 shows the enlarged view of modules 2, 3, and 4 present in the model. The functions of modules 2 and 3 have been elucidated in Table 4. Table 5 also elucidates the functions of module 4.

Table 4 enlists modules 2, 3, 5, and 15 and their respective sub-modules that extract the negative numbers from the normalized values and compute their respective 'm' values. Figures showing the concerned modules are also given.

Table 4. Functions of some modules that deal with negative normalized values.

Module	Sub-modules	Function	Equation or 'map algebra expression' used	Relevant figure
Module 5	Sub-module 1	Extraction of normalized values below -6	FLOAT (SELECT (normalized, "Value < -6"))	Fig. 5
	Sub-module 2	Calculation of the membership value	Equations 4.8 and 4.9	
Module 2	Sub-module 1	Extraction of normalized values between -6 and -5	FLOAT (SELECT (normalized, "Value > -6 AND Value ≤ -5"))	Fig. 4
	Sub-module 2	Calculation of the membership value	Equations 4.8 and 4.9	
Module 15	Sub-module 1	Extraction of normalized values between -5 and -4	FLOAT (SELECT (normalized, "Value > -5 AND Value ≤ -4"))	Fig. 8
	Sub-module 2	Calculation of the membership value	Equations 4.8 and 4.9	
Module 3	Sub-module 1	Extraction of normalized values between -4 and -3	FLOAT (SELECT (normalized, "Value > -4 AND Value ≤ -3"))	Fig. 4
	Sub-module 2	Calculation of the membership value	Equations 4.8 and 4.9	

Table 5. Functions of the remaining modules that deal with negative normalized values.

Module	Sub-modules	Function	Equation or 'map algebra expression' used	Relevant figure
Module 14	Sub-module 1	Extraction of Normalized values between -3 and -2	FLOAT (SELECT (normalized, "Value > -3 AND Value ≤ -2"))	Fig. 8
	Sub-module 2	Calculation of the membership value	Equations 4.8 and 4.9	
Module 4	Sub-module 1	Extraction of Normalized values between -2 and -1	FLOAT (SELECT (normalized, "Value > -2 AND Value ≤ -1"))	Fig. 4
	Sub-module 2	Calculation of the membership value	Equations 4.8 and 4.9	
Module 13	Sub-module 1	Extraction of normalized values between -1 and 0	FLOAT (SELECT (normalized, "Value > -1 AND Value ≤ 0"))	Fig. 7
	Sub-module 2	Calculation of the membership value	Equations 4.8 and 4.9	

Table 6. Functions of some modules that deal with positive normalized values.

Module	Sub-modules	Function	Equation or 'map algebra expression' used	Relevant figure
Module 12	Sub-module 1	Extraction of normalized values between 0 and 1	FLOAT (SELECT (normalized, "Value > 0 AND Value ≤ 1"))	Fig. 7
	Sub-module 2	Calculation of the membership value	Equations 4.10 and 4.11	
Module 6	Sub-module 1	Extraction of normalized values between 1 and 2	FLOAT (SELECT (normalized, "Value > 1 AND Value ≤ 2"))	Fig. 5
	Sub-module 2	Calculation of the membership value	Equations 4.10 and 4.11	
Module 11	Sub-module 1	Extraction of normalized values between 2 and 3	FLOAT (SELECT (normalized, "Value > 2 AND Value ≤ 3"))	Fig. 7
	Sub-module 2	Calculation of the membership value	Equations 4.10 and 4.11	
Module 7	Sub-module 1	Extraction of normalized values between 3 and 4	FLOAT (SELECT (normalized, "Value > 3 AND Value ≤ 4"))	Fig. 5
	Sub-module 2	Calculation of the membership value	Equations 4.10 and 4.11	
	Sub-module 2	Calculation of the membership value	Equations 4.10 and 4.11	

Table 7. Functions of the remaining modules that deal with positive normalized values.

Module	Sub-modules	Function	Equation or 'map algebra expression' used	Relevant figure
Module 10	Sub-module 1	Extraction of normalized values between 4 and 5	FLOAT (SELECT (normalized, "Value > 4 AND Value ≤5"))	Fig. 6
	Sub-module 2	Calculation of the membership value	Equations 4.10 and 4.11	
Module 8	Sub-module 1	Extraction of normalized values between 5 and 6	FLOAT (SELECT (normalized, "Value > 5 AND Value ≤ 6"))	Fig. 6
	Sub-module 2	Calculation of the membership value	Equations 4.10 and 4.11	
Module 9	Sub-module 1	Extraction of normalized values above 6	FLOAT (SELECT (normalized, "Value > 6"))	Fig. 6
	Sub-module 2	Calculation of the membership value	Equations 4.10 and 4.11	

Table 5 enlists modules 4, 13, and 14 and their respective sub-modules that extract the negative numbers from the normalized values and compute their respective 'm' values. The Figures showing the concerned modules are also given.

Table 6 enlists modules 6, 7, 11, and 12 and their respective sub-modules that extract the positive numbers from the normalized values and compute their respective 'm' values. The Figures showing the concerned modules are also given.
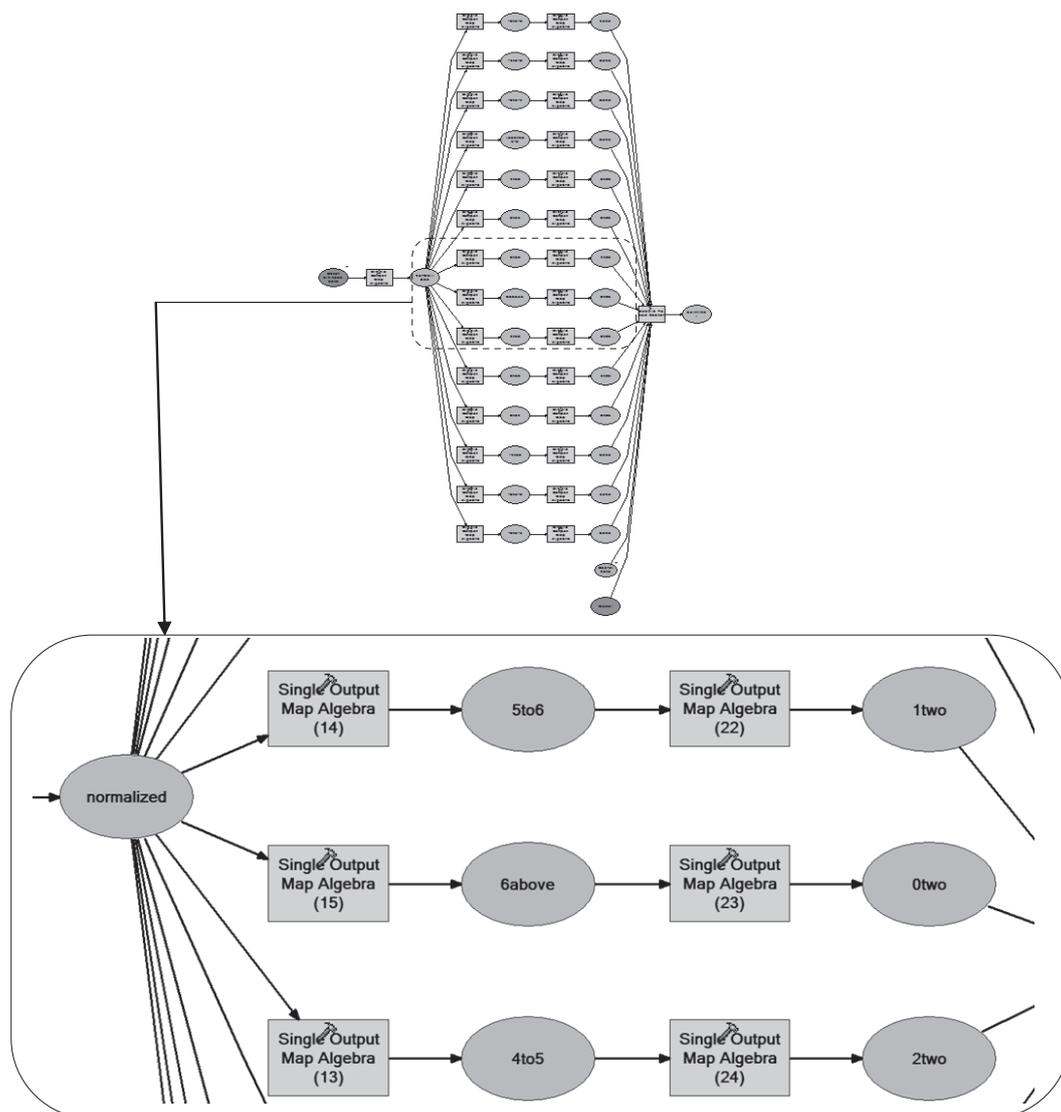

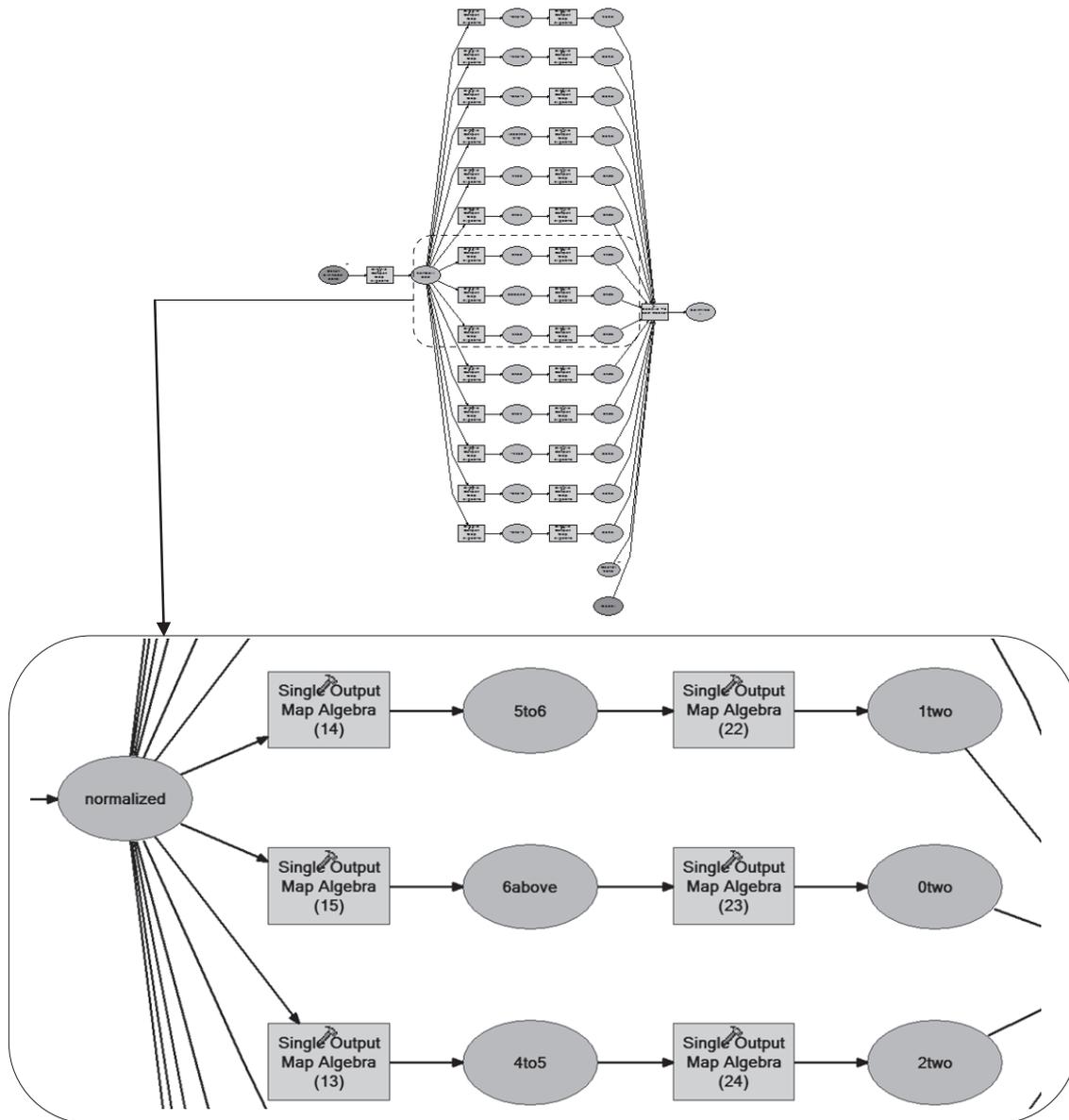Fig. 5. Model structure and the zoomed view of modules 5, 6, and 7 of the model.

Fig. 6. Model structure and zoomed view of modules 8, 9, and 10 of the model.

Table 7 enlists modules 8, 9, and 10 and their respective sub-modules that extract the positive numbers from the normalized values and compute their respective 'm' values. Figures showing the concerned modules are also given.

Fig. 5 shows the enlarged view of modules 5, 6, and 7 present in the model. The functions of modules 5 and 6 have been elucidated in Table 4 and Table 6, respectively. Table 6 also elucidates the functions of module 7.

Fig. 6 shows the enlarged view of modules 8, 9, and 10 present in the model. The functions of these modules have been elucidated in Table 7.

Fig. 7 shows the enlarged view of modules 11, 12, and 13 present in the model. The functions of modules 11 and 12 have been elucidated in Table 6. Table 5 elucidates the functions of module 13.

Fig. 8 shows the enlarged view of modules 14 and 15 present in the model. The functions of modules 14 and 15

have been elucidated in Table 5 and Table 4, respectively.

Module 16 (shown in Fig. 9) receives the output rasters from modules 2 to 15 (discussed so far) and does 'mosaic' operation by assembling all the individual output rasters and gives the final membership value map (i.e., output) for each individual factor-layer.

The membership values of soil type data are obtained by joining the columns of the attribute tables of both the study-area data and the soil type data (Table 2) of the *Santalum album* distribution using the 'join field' tool. Using the 'lookup' tool, the site suitability map concerned with the soil type presence in the study-area is prepared. The model thus developed using this methodology is shown in Fig. 10.

Fig. 11 shows the model that multiplies the membership values of all layers with their respective contributive weightages (w_c) to obtain the final site suitability map of any study area. The model does this operation using 'single
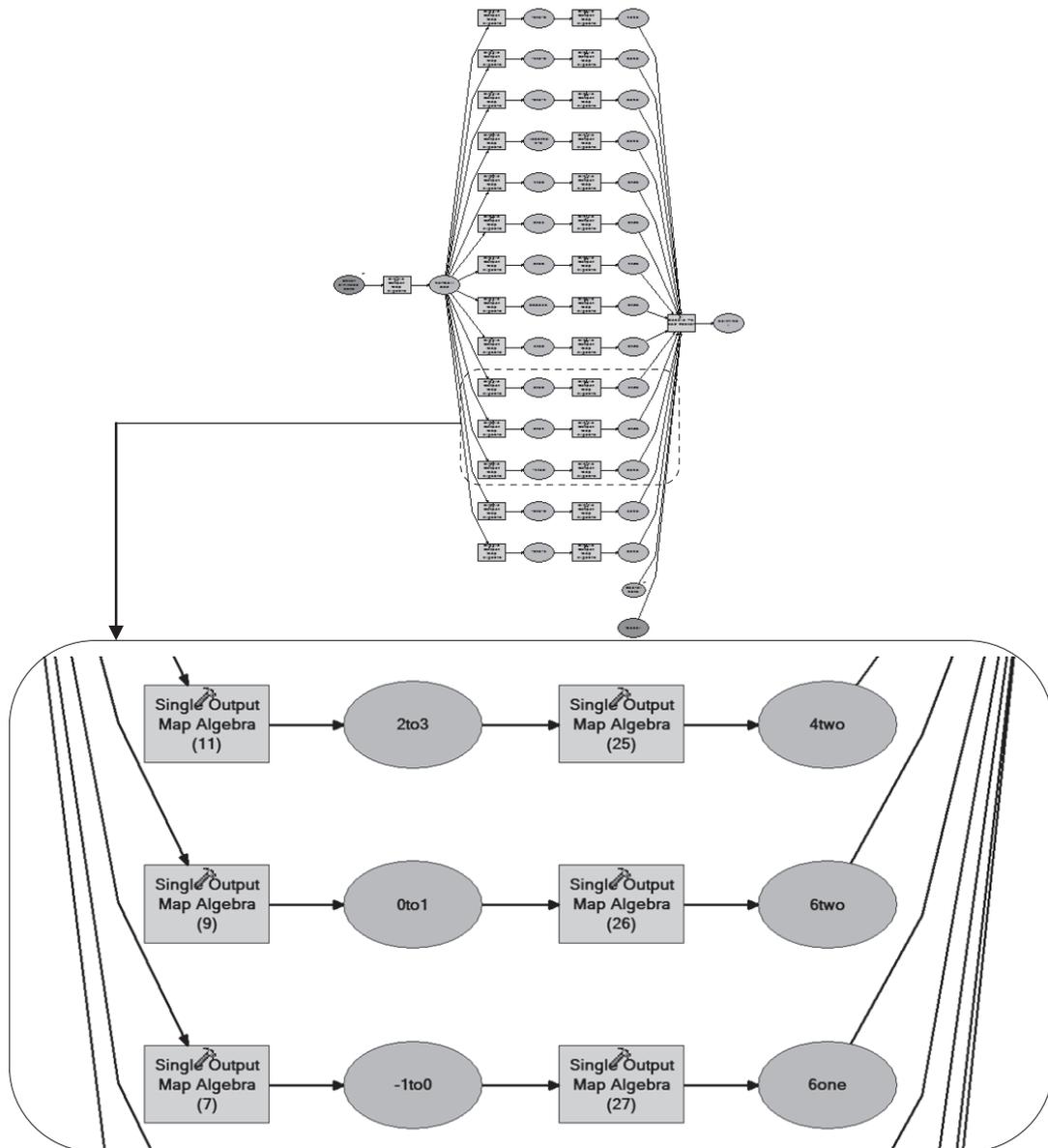
Fig. 7. Model structure and the zoomed view of modules 11, 12, and 13 of the model.

output map algebra expression,' which uses equation 4.8 to get the ultimate result.

Fig. 12 shows the 'ecological niche modelling toolbox' designed and developed in this research using the algorithms discussed in this chapter. This toolbox has two toolsets (one for dependency analysis and the other for membership value estimation) and one model. This toolbox helps in preparing the final site suitability map for any study area. Each toolset of this toolbox consists of six models, each one of which is responsible for the preparation of raster images having contributive indices and membership values concerned with all the respective individual factor-layers.

Advantages of the 'Ecological Niche Modelling Toolbox'

Fig. 13 shows the dialog box which takes the required inputs from the two toolsets discussed before, and provides the output which is the 'Site Suitability Map' of the study area of concern. Thus the toolbox developed in this research contains the 'machine-learned algorithms' bringing higher accuracy and speed in the attainment of the final site suitability map for any study-area. This toolbox also helps in making the job of *Santalum album* management easier, when compared with the manual
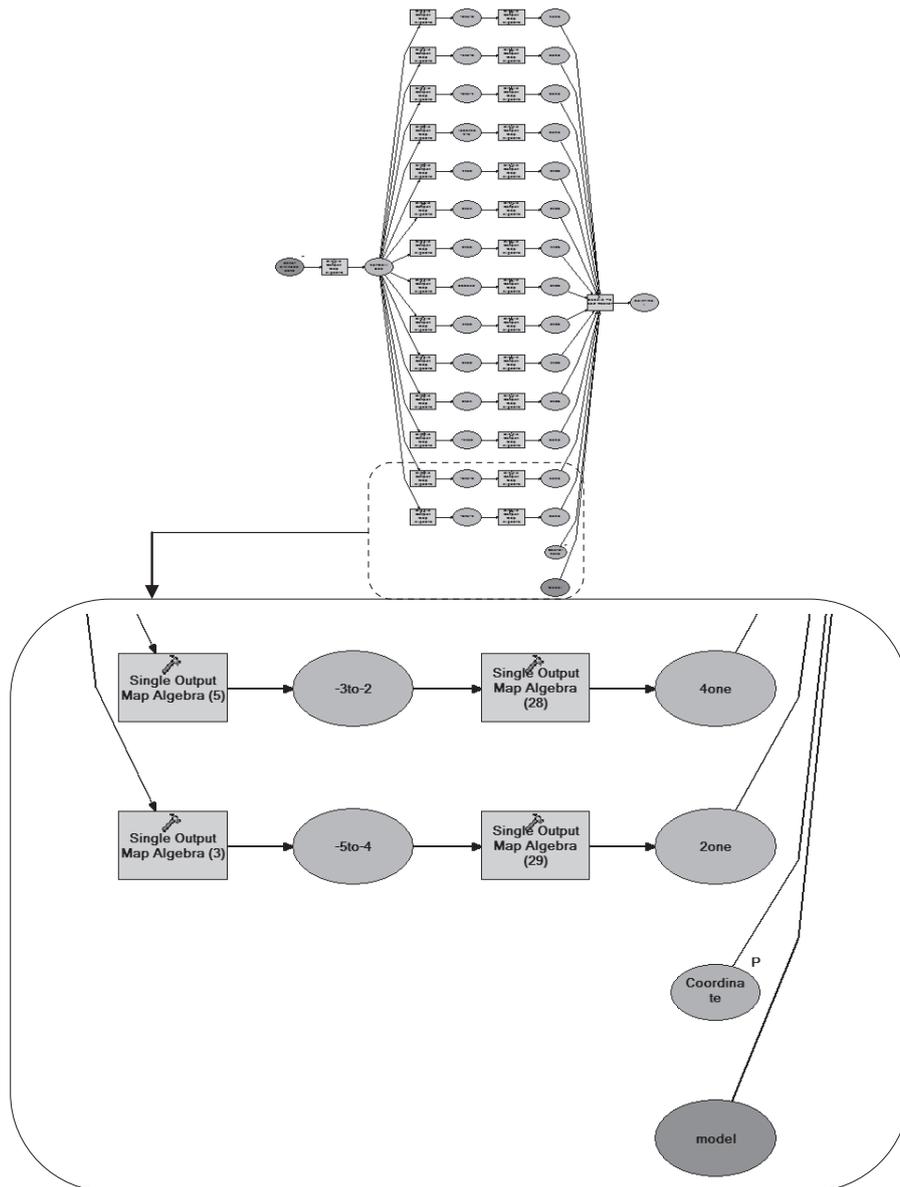
Fig. 8. Model structure and the zoomed view of the modules 14 and 15 of the model.

employment of the algorithms (developed in this research) every time a new study-area needs to be investigated.

Case Study

The study area of this research is Tamil Nadu, the 11th-largest state in India by area (1,30,394.39 km²) and the sixth most populous (7,21,47,030 people). If the appropriate data concerning Tamil Nadu are fed as inputs to the models (discussed so-far) present in the 'dependency analysis' and 'membership value estimation' toolsets, the map (shown in Fig. 14) is obtained as the output.

Results and Discussion

The membership value and site suitability index ranges are the same (i.e., 0 to 6) for all maps prepared in

this research. This entire value range has been divided into three zones, namely high, medium, and low membership value zones. These zones have their zonal value ranges as 4 to 6, 2 to 4, and 0 to 2, respectively.

Table 8, 9, and 10 discuss the outputs (which are intermediate results in this research) obtained using the 'dependency analysis' toolbox. Table 8 shows the actual zonal value ranges present in the respective zones for each parametric factor in 2014. Table 9 shows the soil types that contribute highly, moderately, and lowly in 2014. Table 10 shows the most preferred soil type and the most favored parametric factor value by *Santalum album* with each factor layer in 2014.

Table 11 shows the areal coverages of the high, medium, and low membership value zones in the membership value map, concerned within each factor layer pertaining to Tamil Nadu. As shown in this table, the entire Tamil Nadu comes under the high membership value zone when 'mean
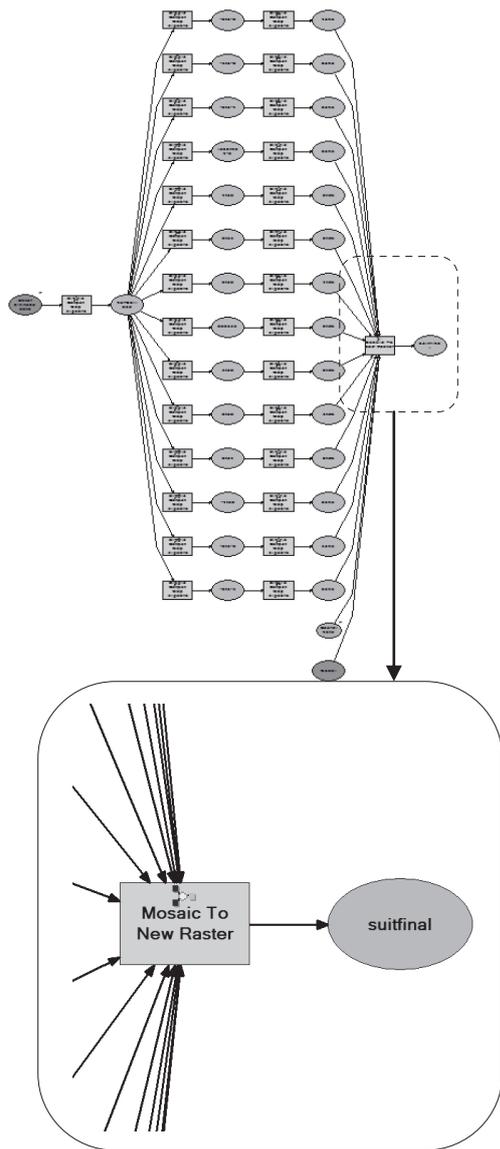
Fig. 9. Model structure and zoomed view of the module 16 of the model.

annual temperature' and 'Soil pH' are considered. Thus Table 11 discusses the final site suitability map, which has been obtained as the ultimate output of this research. It can be noted from Fig. 14 that almost the entire study area (Tamil Nadu) is highly suitable for the survival of *Santalum album,* as the minimum suitability index within the study
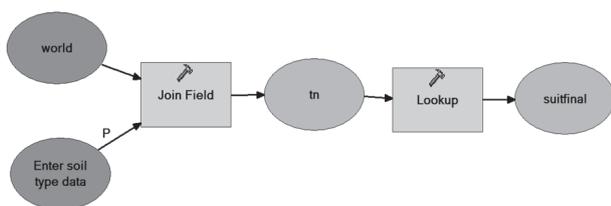

Fig. 10. The model concerned with soil type membership value extraction.
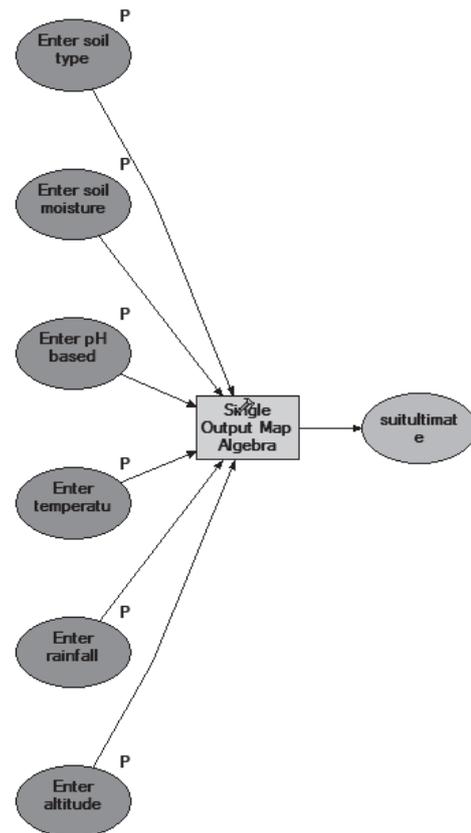

Fig. 11. Model producing the ultimate result.

area is 5.609 (which falls within the high suitability range of 4 to 6). If all the pixels of the final output raster shown in Fig. 14 are statistically analyzed, it can be seen that the scope of survival of *Santalum album* in the study-area as a whol**e** is 83.42%.

Ground Truthing

Accuracy assessment of the prepared map is done by checking and overlaying the same over the areas where currently there is an acceptable greater distribution of the tree. Such areas in Tamil Nadu are the Kalrayan Hills, Chitteri Hills, Shevaroys, Javadi Hills, and Pachamalai Hills. If the suitability index over these areas (with more number of Sandalwood trees) is high, then it indirectly means that the prepared map has high accuracy. Table 12 shows the accuracies attained within each hill range.
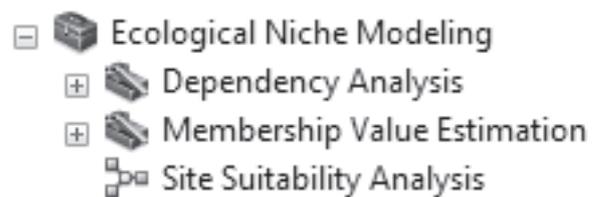

Fig. 12. The 'ecological niche modelling toolbox' developed in this research.
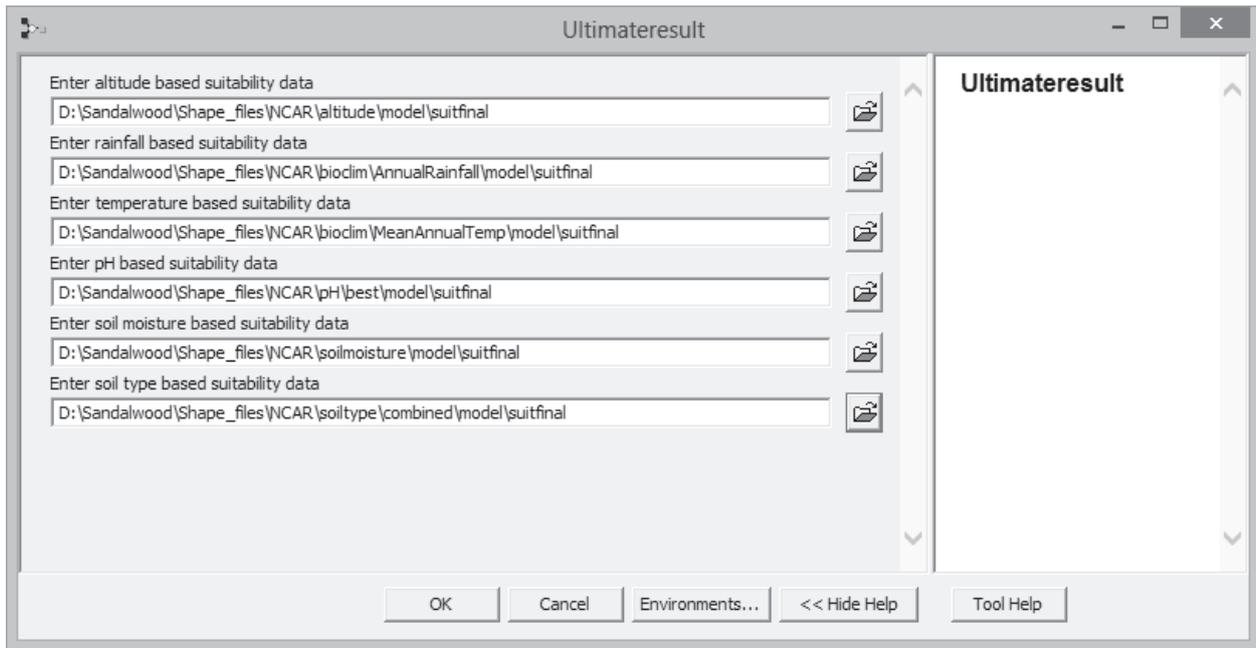
Fig. 13. Input dialog box of site suitability analysis for *Santalum album*.

Table 8. The actual zonal value ranges present in the respective zones for each parametric factor in 2014.

Soil number	Parametric factor layer	Actual parametric factor's zonal value ranges occurring in		
		High contributive index zone (i.e., 4 to 6)	Medium contributive index zone (i.e., 2 to 4)	Low contributive index zone (i.e., 0 to 2)
1.	Mean annual temperature	19.05-28.90ºC	13.3-19.05ºC	-
2.	Altitude	451-1,951.36 m	1,951.36-2,852.77 m	2,852.77-2,924 m
3.	Mean annual rainfall	981.85-1,983.95 mm	1,983.95-2,045.72 mm	-
4.	Soil moisture	3.70-114.86 Kg m^{-2}	114.86-117.71 Kg m^{-2}	-
5.	Soil pH	3.11-7.30	1.8-3.11 AND 7.30-7.60	-

Table 9. Soil types that contribute highly, moderately, and lowly in 2014.

Non-parametric factor layer	Soil types		
	Having high contribution	Having medium contribution	Having low contribution
Soil type	Luvisols and Alisols	Regosols, Acrisols, and Vertisols	Nitisols, Cambisols, Lixisols, Phaeozems, Andosols, Podzols, Kastanozems, Ferralsols, Leptosols, Solonchaks, Histosols, and Fluvisols

Table 10. Optimal value for all the factors in 2014.

Soil number	Factor	Optimal value for the factor in 2014
1.	Mean annual temperature	24.82ºC
2.	Altitude	1,049.94 m
3.	Mean annual rainfall	1,443.46 mm
4.	Soil moisture	59.27 Kg m^{-2}
5.	Soil pH	5.20
6.	Soil type	Luvisol
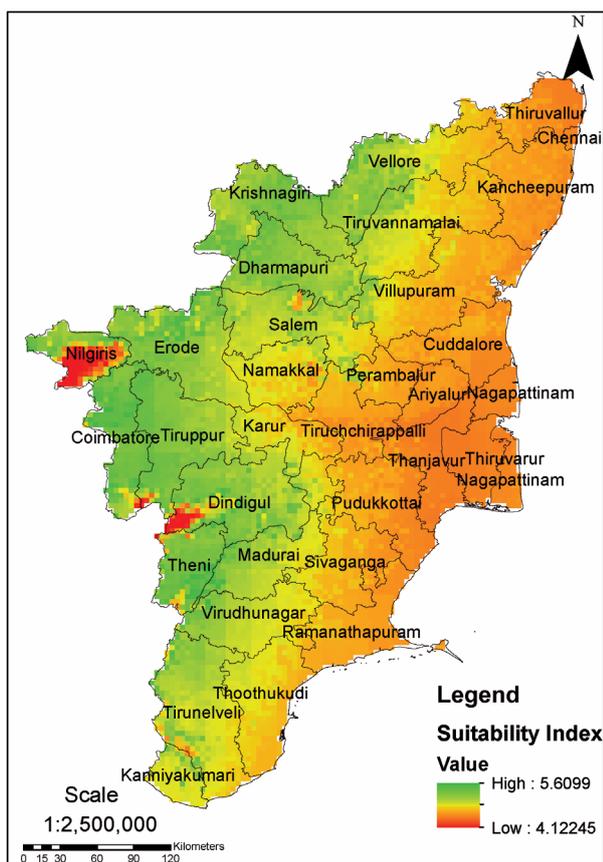
Fig. 14. Site suitability index map of Tamil Nadu for the growth of *Santalum album* in 2014.

Fig. 15. Photo taken while conducting a field visit in the Shevaroy Hills for ground truthing.

Direct field visit (Fig. 15) has been conducted for each of the above-mentioned hills for verifying whether sandalwood trees are actually present at their locations, which was done with the help of hand-held GPS by noting the respective latitude and longitude and verifying the same with respect to the high suitability zones in the site suitability map prepared. A total of 32 points has been visited in all hills together and these points to give 90.625% accuracy.
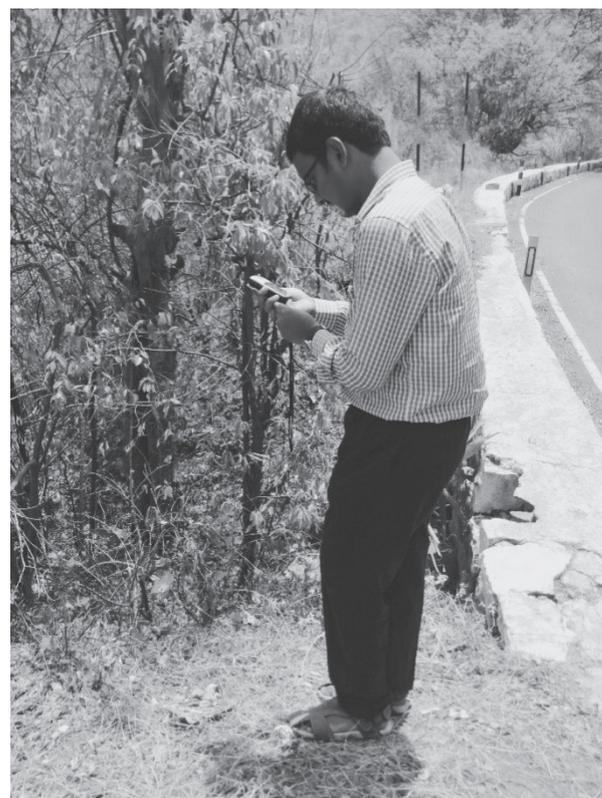
Table 12. Accuracies attained within hill ranges.

Hill range	Percentage of accuracy of 'site suitability map' attained over these hill ranges
Kalrayan Hills	86.75%
Pachamalai Hills	90.59%
Shevaroy Hills	87.22%
Chitteri Hills	86.95%
Javadi Hills	87.10%

Table 11. Areas covered under each membership value zone in 2014.

Soil number	Factor layer	Area (km²)		
		High membership value zone	Medium membership value zone	Low membership value zone
1.	Mean annual temperature	1,30,394.39	0	0
2.	Altitude	1,28,356.32	1,998.94	36.90
3.	Mean annual rainfall	126,482.56	3,911.83	0
4.	Soil moisture	129,090.45	1,303.94	0
5.	Soil pH	1,30,394.39	0	0
6.	Soil type	13,039.44	46,941.98	70,412.97

Conclusion

The 'ecological niche modelling' toolbox designed in this research can be helpful in obtaining 'site suitability maps' for any study area quickly and effectively if the required data are fed as inputs. The Indian state of Tamil Nadu was taken as the study area and it was observed that this Indian state provided 83.42% scope of survival of *Santalum album* in 2014. The output maps are validated using overlay operations in GIS and also through ground truthing. The Accuracy achieved through direct field visits was 90.625%, but through GIS overlay operations we attained variable percentages of accuracies (within each of the hill ranges) for the same map. This was because there is a variance in the quantum of distribution of sandalwood trees among these hill ranges.

Probable reasons for the non-achievement of 100% accuracy of the 'site suitability map' are:

1. There cannot be a strict condition that if a hill poses a high suitability for *Santalum album*'s survival, then that hill should be completely covered with sandalwood trees.

2. Also, it can be noted that there are still some areas with non-favorable situations for *Santalum album*'s survival if the 'site suitability map' of the hills is carefully analyzed.

References

1. ARUN KUMAR A., GEETA JOSHI, MOHAN RAM H. Sandalwood: history, uses, present status and the future. Current Science, 103, **2012**.

2. ORWA, C., MUTUA, A., KINDT, R., JAMNADASS, R., & SIMONS, A. Agro-forestry database 4.0: a tree reference and selection guide, **2009**.

3. HIJMANS R.J., CAMERON S.E, PARRAN J.L., JONES P.G., JARVIS A. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, **2005**.

4. JARVIS A., REUTER H.I., NELSON A., GUEVARA E. CGIAR-CSI SRTM 90 m Database. Available online: http:// srtm.csi.cgiar.org (accessed on 09 July **2014**).

5. MARIANA V., TONY C., TOM H., SYLVIA M., GEORGE R., CARR J.R., NANCY N. Community Climate System Model 3.0. National Center for Atmospheric Research, Boulder, Colorado, USA, **2004**.

6. TOMISLAV H., JORGE M.D.J, GERARD B.M.H., MARIA R.G., MILAN K., ALEKSANDAR B., WEI S., MARVIN N.W., XIAOYUAN G., BERNHARD B.M., MARIO A.G., RODRIGO V.R, BAS K. SoilGrids 250 m: global gridded soil information based on Machine Learning, 1, **2016**.

7. PEARSON R.G. Species Distribution Modeling for Conservation Educators and Practitioners, Center for Biodiversity and Conservation & Department of Herpetology, American Museum of Natural History, USA, 40, **2008**.

8. ROLLAND C., THANOS P. A Comprehensive View of Process Engineering. Proceedings of the 10th International Conference, CAiSE'98. 1410, **1998**.