

RESEARCH ARTICLE

Desertification vulnerability index—an effective approach to assess desertification processes: A case study in Anantapur District, Andhra Pradesh, India

Subramanian Dharumarajan¹  | Thomas F.A. Bishop³ | Rajendra Hegde¹ | Surendra Kumar Singh²

¹ ICAR-National Bureau of Soil Survey and Land Use Planning, Hebbal, Bangalore 560024, India

² ICAR-National Bureau of Soil Survey and Land Use Planning, Amaravati Road, Nagpur 440033, India

³ School of Life and Environmental Sciences, Sydney Institute of Agriculture, The University of Sydney, Sydney 2006, Australia

Correspondence

S. Dharumarajan, ICAR-National Bureau of Soil Survey and Land Use Planning, Hebbal, Bangalore 560024, India.

Email: sdharmag@gmail.com

Abstract

There is a need for the up-to-date assessment of desertification/land degradation maps that are dynamic in nature at different scales for comprehensive planning and preparation of action plans. This paper aims to develop the desertification vulnerability index (DVI) and predict the different desertification processes operating in Anantapur District, India, based on machine language techniques. Climate, land use, soil, and socioeconomic parameters were used to prepare DVI by a multivariate index model. The computed DVI along with climate, terrain, and soil properties was used as explanatory variable to predict the desertification processes by using a random forest model. About 14.2% of the area was created as a training dataset in 9 places for modeling and remaining area was tested for prediction of desertification processes. We used desertification status map (DSM) of Anantapur District prepared under Desertification status mapping of India-2nd cycle as a reference dataset for calculation of accuracy indices. Kappa and classification accuracy index were calculated for training and validation datasets. We recorded overall accuracy rate and kappa index of 85.5% and 75.8% for training datasets and 71.0% and 51.8% for testing datasets. The results of variable importance analysis of random forest model showed that DVI was the most important predictor followed by potential evapotranspiration and Normalized Difference Vegetation Index for prediction of desertification processes. The results from this work given new insight into using the existing knowledge on prediction of desertification in unvisited areas and also quick update of DSM maps.

KEYWORDS

desertification, desertification vulnerability indices, prediction, random forest model, variable importance

1 | INTRODUCTION

Desertification is a process of land degradation in arid, semiarid, and dry subhumid regions resulting from various factors, including climatic variation and improper human activities (UNEP, 1992). It affects one-third of the world's land area making this area extremely vulnerable to over exploitation and inappropriate land use. United Nations estimates show that about 3.6 billion hectares of the world's land surface are already desertified (UNCCD, 2014), and every year, 5,300 million tonnes of fertile soil is lost along with 8 million tonnes of plant nutrients due to different desertification processes. The vegetal degradation, erosional processes, waterlogging and salinization, and loss of

soil fertility are the major processes responsible for desertification worldwide (Dregne & Chou, 1992).

In the case of India, an increasing human as well as livestock population exerting huge pressure on land resources, which have led to very significant land degradation. Out of India's total geographical area (TGA) of 328 Mha, about 228 Mha (69%) is under dry conditions (arid or semiarid or dry subhumid), and these lands are highly populated and more vulnerable to environmental stress. Sehgal & Abrol (1994) reported that 57% of land area in India has been degraded in one way or the other, and lands under cultivation are the most degraded followed by grazing land and forests. Singh (2009) reported that the salinity and sodicity problem has degraded 6.73 Mha in the Country.

As a signatory to the United Nations Convention for Combating Desertification, the Indian government has conducted various desertification status assessment and monitoring projects at different scales (Ajai, Dhinwa, Pathan, & Ganesh, 2009; Singh, 2009). Very recently, the Space Application Centre (ISRO), Ahmedabad in collaboration with 19 partner institutes completed the task of making an inventory of the desertification status of India (SAC, 2016). The report revealed that during 2011–2013, about 82.64 Mha of arid, semiarid, and dry subhumid regions of the country were affected by different desertification processes that is an increase of 1.16 Mha as compared to 2003–2005. The most significant process of desertification in arid regions is wind erosion whereas vegetation degradation and water erosion dominates in semiarid and dry subhumid regions of the Country.

Identification and delineation of vulnerable areas help the policy makers to prepare strategies to reverse or slow down desertification processes. Desertification is a product of the combined action of biophysical and socioeconomic factors including climate, terrain, soil, and human pressure (Dharumarajan et al., 2017a). A multidisciplinary approach based on biophysical and socioeconomic factors is required to identify areas vulnerable to desertification. Several mathematical models have been developed across the world to delineate areas vulnerable to desertification (Dasgupta, Sastry, Dhinwa, Rathore, & Nathawat, 2013; Salvati & Zitti, 2009; Symeonakis, Karathanasis, Koukoulas, & Panagopoulos, 2016). The Environmental Sensitivity Index is assessed by computing the Soil Quality Index, Climate Quality Index, Vegetation Quality Index, and Management Quality Index in the MEDALUS methodology (Kosmas, Ferrara, Briasouli, & Imeson, 1999). Though the model has its own weakness of equal weight for all the indices (Salvati, Zitti, Ceccarelli, & Perini, 2009), it is a useful index particularly in Mediterranean climates due to its simplicity in model building and flexibility in selection of indicators (Trott, Menegoni, Manfredi Frattarelli, & Iannetta, 2015), and the model requires an expert judgment for assigning the scores to the different classes (Giordano et al., 2003). Salvati and Zitti (2008) proposed an index based on both biophysical and socioeconomic factors and obtained the results that are comparable with the MEDALUS results. In another example, Karamesouti et al. (2015) used the PESERA and TERON models to assess the soil erosion rates, and MEDALUS methodology was used to assess the overall desertification risk in a traditional Mediterranean cropland ecosystem. Fuzzy inference rules have also been used for delineation of environmentally vulnerable areas by Dasgupta et al. (2013). Jafari and Bakhshandehmehr (2013) combined fuzzy logic and the Environmental Sensitivity Index to map the desertification sensitivity index in Central Iran. Dutta and Chaudhuri (2015) studied environmentally sensitive areas in the Jhunjhun and Sikar districts in Rajasthan, India, and the results showed that 13% of the study area was highly sensitive to desertification.

The mapping and inventory of desertification processes provide baseline data for prioritizing areas and preparing action plans for combating desertification. Since desertification is a dynamic process, we need to map the status of desertification at regular time intervals for monitoring purposes. But mapping at the required scale at regular intervals with high accuracy is always a challenging, time consuming, and costly affair (Vågen, Winowiecki, Tondoh, Desta, & Gumbricht, 2016). This is largely due to large amount of expert

knowledge, for example, image interpretation, and fieldwork required to create the maps of desertification processes. Recently, machine learning algorithms have emerged as more efficient alternatives to conventional parametric algorithms in large scale mapping (Rogan et al., 2003). A variety of machine-language techniques such as classification and regression tree (Lawrence, Bunn, Powell, & Zambon, 2004), artificial neural network (Priori, Bianconi, & Constantini, 2014), k-nearest neighbor (Mansuy et al., 2014), multinomial logistic regression (Kempen, Brus, Heuvelink, & Stoorvogel, 2009), support vector machine (Kovačević, Bajat, & Gajić, 2010; Priori et al., 2014), and random forest model (RFM; Kim, Grunwald, Rivero, & Robbins, 2012; Vågen et al., 2016) were tested by different authors in natural resources mapping.

These machine learning languages are mostly used in digital soil mapping (Dobos, Carré, Hengl, Reuter, & Tóth, 2006; McBratney, Mendonça, & Minasny, 2003), land use land cover classification (Ghosh, Sharma, & Joshi, 2014), and different ecological studies (Lawrence et al., 2004). The application of RFM on digital soil mapping was proved by several authors (Dharumarajan et al., 2017b; Grimm, Behrens, Märker, & Elsenbeer, 2008; Vågen et al., 2016), and the soil property at unvisited locations is predicted using environment covariates such as slope, temperature, and rainfall (Akpa, Odeh, Bishop, & Hartemink, 2014; Lagacherie, Legros, & Burroughs, 1995). RFM can be applied for prediction of both continuous properties such as soil organic matter content, clay, sand, silt, pH, and electrical conductivity using regression algorithm (Akpa et al., 2014; Vågen et al., 2016) and categorical properties such as soil taxonomic units using a classification algorithm (Hastie, Tibshirani, & Friedman, 2009; Kidd, Malone, McBratney, & Minasny, 2014; McBratney et al., 2003).

But there are no studies available on use of machine language classifiers in land degradation or desertification mapping. In this context, the objective of this study is to develop desertification vulnerability indices using the multivariate index model and also to evaluate the RFM for prediction of different desertification processes for unvisited locations in the Anantapur District, Andhra Pradesh, India.

In particular, we explore the possibility of assessing small regions for desertification using traditional methods, and then extrapolating the results to a larger region using machine learning models.

2 | MATERIALS AND METHODS

2.1 | Description of the study area

Anantapur, the driest District in India next to Rajasthan is located between 13°40' and 15°15' N latitudes and 76°50' and 78°30' E longitudes in Rayalaseema region of Andhra Pradesh (Figure 1). The TGA of the district is 19,130 km². The climate of Anantapur is warm all around the year and mainly classified as a hot arid bioclimatic condition. The mean temperature is always above 23 °C, and April and May are the hottest months with mean temperatures between 32 and 35 °C. The mean average rainfall is 552 mm. The annual average potential evapotranspiration is 1,857 mm, which is more than three times as much as the annual rainfall. Inceptisols are the dominant soil covering 31.04% of total TGA followed by Alfisols (30.1%), Entisols (14.8%), and Vertisols (8.84%) (NBSS&LUP, 2008).

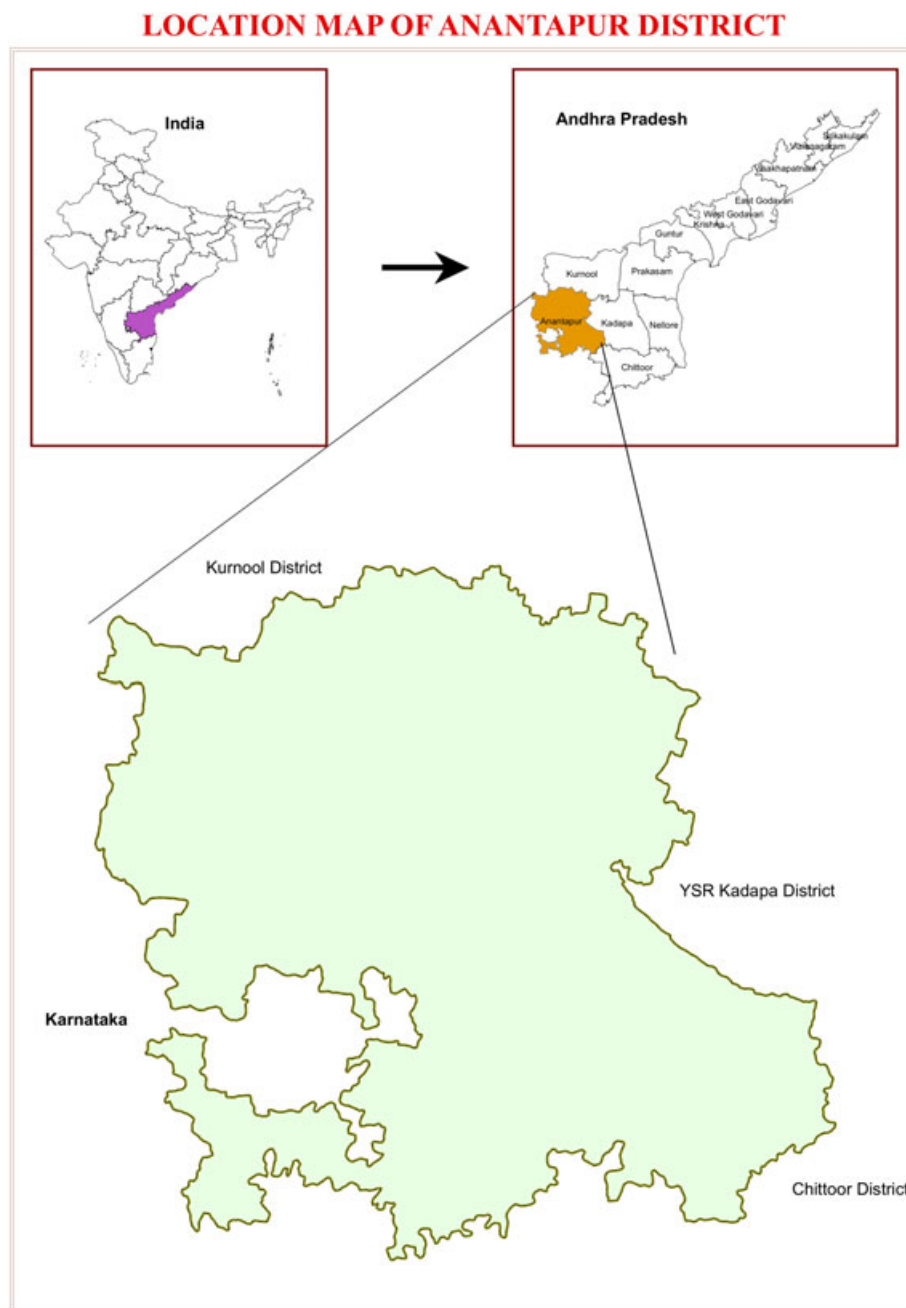


FIGURE 1 Location map [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | Derivation of desertification vulnerability index

Desertification vulnerability assessment is important to prepare proper conservation measures for its prevention (Dasgupta et al., 2013). We used a multivariate index model to calculate desertification vulnerability index. In contrast to the conventional way of classifying land degradation where each parameter is classified into certain upper and lower limit values, we used a modified weighted index for classifying the different parameters. Climate, soil, and land use indices are generated separately in the form of spatial layers and integrated in a Geographic Information System (GIS) environment to delineate environmentally sensitive areas for desertification, in what is termed a biophysical index (BPI). More than natural parameters, human intervention leads to significant changes in the environment; therefore,

the socioeconomic index (SEI) is also considered to determine the desertification vulnerability. So the final index of desertification vulnerability is calculated based on integration of SEI and the BPI (Figure 2) by expert knowledge inference rules (Dasgupta et al., 2013).

2.3 | Socioeconomic index

Society and its economic conditions directly or indirectly influence every piece of land and are responsible for desertification. Population pressure, unemployment, and illiteracy pose a serious threat to the environment in conjunction with poverty (Vu, Le, Frossard, & Vlek, 2014). Population growth and population density were used to develop a land vulnerability index by Salvati et al. (2009) in Italy.

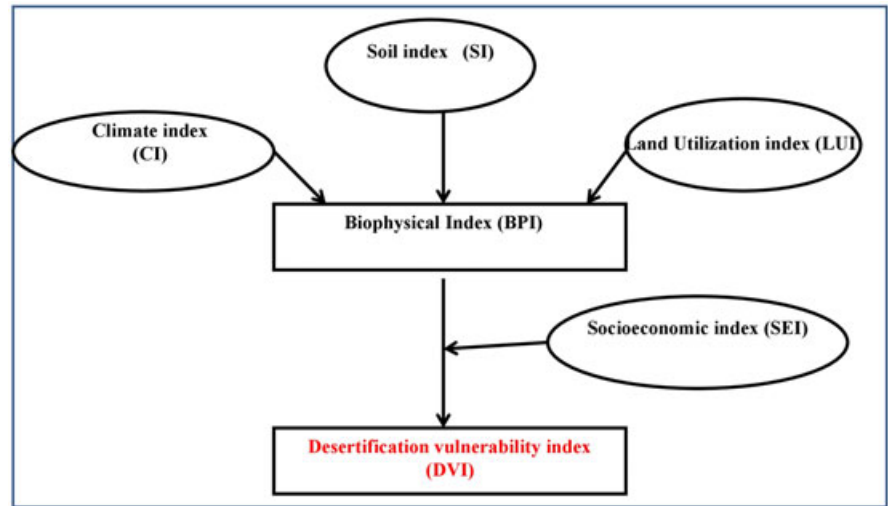


FIGURE 2 Components of desertification vulnerability index [Colour figure can be viewed at wileyonlinelibrary.com]

Symeonakis et al. (2016) included an old age index and education level along with population growth and population density for assessment of desertification vulnerability in island of Lesvos (Greece). For this study, we used two indices, that is, amenities index and economic development index for calculation of the SEI.

2.3.1 | Amenities index

Amenities are considered as an important indicator for social development. The status of each settlement was rated based on availability of all the amenities for the particular settlement. We considered four amenities: education, medical, transport, and communication for calculation of the amenities index. Since the availability or unavailability of these amenities are indirectly influencing the sustainability of land, we derived the amenities index based on following formulae.

$$I_a = \frac{\sum (A_i \times W_i)}{\sum W_i} \quad i = 1 \text{ to } n, \quad (1)$$

where I_a is index for amenities of particular settlement; n is the number of amenities in a category; A_i is 1 or 2 (1 = Available, 2 = Not available); W_i is the weight of the amenity within category/class of facility and it is defined as

$$W_i = \frac{(N - f_i)}{N} \times 100, \quad (2)$$

where N is the total number of settlements; f_i is the number of settlements having amenity i .

Cumulative amenities index was calculated based on

$$CAI = \sum I_a, \quad (3)$$

where a is 1 to m ; m is the number of amenity categories (medical, educational, transport, communication); CAI is cumulative amenities index for a particular settlement.

The presence or absence of each amenity was extracted from the 2011 National Census (Census of India, 2011) for 987 settlements of the Anantapur District.

2.3.2 | Economic development index

Economic development is one of the major drivers for land degradation (Salvati, Mancini, Roberta Gemmiti, & Carlucci, 2011), and the economic development of particular settlements is related to the population of the settlement. We considered population density of the settlement and proportion of employed population to total population for calculation of an economic development index.

$$EDI = \sqrt{D \cdot W (W - A)}, \quad (4)$$

where EDI is economic development index; D is population density; W is the proportion of employed population; A is the proportion of unskilled workers (i.e., unemployed + agricultural laborers + marginal workers/total population).

The cumulative amenities index (CAI) and economic development index was classified into five classes based on its mean (μ) and standard deviation (σ).

$< \mu - 2\sigma$ = very low,
 $\mu - 2\sigma$ to $\mu - \sigma$ = low,
 $\mu - \sigma$ to $\mu + \sigma$ = moderate,
 $\mu + \sigma$ to $\mu + 2\sigma$ = high,
 $> \mu + 2\sigma$ = very high.

The socioeconomic development index was calculated based on CAI and economic development index by applying expert knowledge inference rules, and the final maps of five classes (very low, low, moderate, high, and very high) were developed in GIS environment.

2.4 | Biophysical index

Climatic index calculated based on an aridity index. Annual average rainfall data of 30 years (1981 to 2011; IMD, 2013) and potential evapotranspiration was collected from Indian Meteorological Department (IMD) and used for calculation of aridity index. The aridity index (P/PET) was calculated and classified into two classes (arid < 0.2 ; semi-arid 0.2–0.5). The soil map of the Anantapur District (NBSS&LUP, 2008) was used for preparation of a soil index. This was based on soil depth, drainage, and soil texture, available water holding capacity

(AWC) and rock fragments, and classified into five classes based on the range of data, and ranks were assigned (Kosmas et al., 1999). The criteria used for evaluation of the soil index are provided in Table 1.

$$\text{Soil Index} = (\text{soil depth} \times \text{drainage} \times \text{soil texture} \times \text{AWC} \times \text{rock fragments})^{1/5}$$

The land utilization index was calculated from present land use and slope and had three classes; under utilization, optimal utilization, and over utilization. The land utilization index indicates the status of present land use and the risk toward desertification processes (Sastry, 2011).

Climate, soil, and land utilization index were integrated in GIS environment, and composite index (BPI) was developed by using expert knowledge in GIS environment (Dasgupta et al., 2013). Since land use have more impact on land degradation that can be altered radically by human in a short time compared to soil and climate, the order of significance in calculation of BPI is Land Utilization index > soil index > climatic index (Sastry et al., 2009). Accordingly, due weightage was assigned to each indices for calculation of BPI. SEI and BPI are integrated, and final desertification vulnerability index were developed and classified into five classes, that is, very low (nonthreatened areas for desertification), low (potential areas and subjected to very slight degradation), moderate (fragile areas and subjected to slight degradation), high (critical areas and subjected to moderate degradation), and very high (critical areas and subjected to very high erosion). SEI was given more weight compared to BPI because demographic and economic variables are more important for land degradation than natural variables (Vu et al., 2014)

$$\text{DVI} = \text{BPI} \times \text{SEI}$$

2.5 | Environment covariates used for modeling

We used the final desertification vulnerability index along with other environmental covariates to predict the major desertification processes operating in Anantapur District. Preprocessing of environment variables was carried out using ArcGIS. Climatic variables

(temperature, rainfall, and potential evapotranspiration), terrain variables (slope and aspect), soil variables (depth, sand, clay, bulk density, pH, and organic carbon (OC)), and vegetation variables (Normalized Difference Vegetation Index) were used as covariates for prediction. The mean annual temperature and annual rainfall of 55 locations (30 years) were interpolated by ordinary kriging before being used as covariates. Slope and aspect were derived from Shuttle Radar Topography Mission (SRTM) at a resolution of 30 m using the ArcGIS. Both variables represent the shape of the terrain (Grinand, Arrouays, Laroche, & Martin, 2008). Soil physical and chemical properties were extracted at a resolution 250 m from a global soil map (www.soilgrid.org). Normalized Difference Vegetation Index (NDVI) was derived from the MODIS NDVI product (MOD13Q1), and the mean value for 2011–2015 was used. The final database was composed of 14 environmental variables and one target variable representing four desertification units (Table 2).

2.6 | Spatial prediction of desertification processes using random forest model

We used Random Forest 4.6 package (Liaw & Wiener, 2002) in the R environment for prediction of the major desertification processes in the Anantapur District. RFM is based on assemblage of a number of classification or regression trees (depending on the response variable) by using two levels of randomization for every tree in the forest (Breiman, 2001). In comparison to other prediction models, RFM not only improves the prediction accuracy (Breiman, 2001; Liaw & Wiener, 2002) and also reduces model overfitting. The advantage of the RFM is that it is non-sensitive to missing data and it has capacity to handle large number of quantitative and categorical data (Grinand et al., 2008). Since it selects the most important predictor at each node split, it is considered to be an insensitive to noisy or weak predict or variables (Okun & Priisalu, 2007). The internal out-of-bag (OOB) prediction generated through bootstrapping provides an estimate of accuracy across the decision trees that were used for initial assessment of performance of model (Vågen et al., 2016). Relatively

TABLE 1 Assigned scores of soil quality indices for calculation of DVI

Soil depth		Soil texture	
Class	Score	Class	Score
Deep (>100 cm)	4	Sandy loam, Sandy clay loam, clay loam	3
Moderately deep (75–100 cm)	3	Sandy clay, clay	2
Moderately shallow (50–100 cm)	2	Sand, loamy sand	1
Shallow (<50 cm)	1		
Gravelliness		Drainage	
Slightly gravelly (<15%)	1	Well-drained	3
Moderate (15–35%)	2	Mod. well drained	2
Strong (>35%)	3	Imperfectly drained	1
AWC		Soil quality indices	
Very high (>200 mm/m)	3	Very low	1–125
High (150–200 mm/m)	2.5	Low	1.25–1.5
Medium (100–150 mm/m)	2	Moderate	1.50–1.75
Low (50–100 mm/m)	1.5	High	1.75–2.0
Very low (<50 mm/m)	1	Very high	>2.0

Note. DVI = desertification vulnerability index; AWC = water holding capacity; SL = sandy loam; SCL = sandy clay loam; CL = clay loam.

TABLE 2 Datasets used for modeling

Predictor	Source	Resolution	
Bulk density (mg/m ³)	Soil grid.org (global soil map)	250 m	1.37–1.72
Depth (cm)			69–200
Coarse fragments (%)			3–45
Sand (%)			28–61
pH			6.5–7.8
Clay (%)			19–48
OC (%)			0.4–6.4
Slope (%)	SRTM	30 m	0–59.4%
Aspect			0–359
NDVI	MOD13Q1	250 m _16 days	–0.06–0.69
Temperature (max) °C	IMD	Selected locations	32.0–34.7
Rainfall (mm)	IMD	Selected locations	381–693
Potential evapotranspiration (mm)	IMD	Selected locations	1610–2043
DVI	Climatic index soil index Land utilization index Socioeconomic index	IMD Soil map of Anantapur Resourcesat LISS III data, soil map of Anantapur Census of India, 2011	5 classes

Note. NDVI = Normalized Difference Vegetation Index; DVI = desertification vulnerability index.

important variables identified based on number of times variable was used in nodes. The final prediction error is calculated as the classification error rate and reported as mean percent decrease in classification rate. The internal OOB prediction generated through bootstrapping provides an estimate of accuracy across the decision trees that was used for initial assessment of performance of model.

2.7 | Model accuracy assessment

To evaluate the prediction performance of the model, we used datasets developed from desertification status mapping project as a reference for model building and validation. Desertification status mapping project is a gigantic task carried out by Space Application Centre, Ahmedabad in collaboration with ICAR-National Bureau of Soil Survey and Land Use Planning, Bangalore. Three seasons (*kharif*, *rabi*, summer) of georeferenced Indian Remote sensing (Resourcesat-2) LISS III data were analysed using on-screen visual interpretation techniques along with ancillary information to interpret different desertification and land degradation classes, and a desertification status map was finalized after intensive ground truthing. The final output of the project was used for a reference dataset for validation of RFM. Training datasets and testing datasets are divided manually so that all desertification processes were present in both areas. The number of pixels used for training and testing data are 7,373 and 51,898, respectively, and approximately 14% of the area was used for training (Figure 3). We tried to predict three major desertification classes; vegetal degradation, soil erosion, and salinisation along with no desertification class. To predict the desertification status map of the entire district, 250 m grid values of covariates were extracted and used for mapping. This accuracy assessment is summarized in the form of confusion matrix that was used to derive the classification accuracy indexes; overall accuracy rate and kappa index. Overall accuracy rate is based on the ratio between the correctly predicted (confusion matrix diagonal) and the total number of classified pixels. The kappa index is a robust index calculated based on expected accuracy and observed accuracy. Spatial autocorrelation and representativeness of sampling are the two issues that affect the statistically

valid independent estimate of accuracy (Muchoney & Strahler, 2002). Training and testing samples within the same site cannot be considered as independent because those pixels are auto-correlated. We achieved sample independency by choosing the testing and training samples separated by certain distance (Grinand et al., 2008).

3 | RESULTS AND DISCUSSION

3.1 | Desertification vulnerability index

3.1.1 | Socioeconomic index

Statistics of socioeconomic development index are presented in Table 3. Education index is calculated based on 28 amenities listed (e.g., government primary school, government middle school, and government arts college) in the census 2011 (see Equations 1, 2, and 3). The calculated education index ranged from 0 to 0.57 with the mean of 0.04 and standard deviation of 0.06. Out of 987 settlements, 378 settlements recorded an education index more than the mean value. The medical index is calculated based on 15 medical related amenities (primary health subcenter, maternity and child welfare center, dispensary and veterinary hospital, etc.) listed in the census. The calculated medical index ranged from 0 to 0.96 with the mean and standard deviation of 0.09 and 0.12, respectively. Transport and communication indices were calculated based on eight (post office, post and telegraph office, telephone facilities, etc.) and 21 (bus services, railway stations, taxi, etc.) amenities and the calculated index ranged from 0 to 0.69 and 0 to 1.00, respectively. The cumulative amenities index of education, medicine, transport, and communication ranged from 0 to 3.23 with the mean (μ) and standard deviation (σ) of 0.56 and 0.38. Based on the mean and standard deviation, the settlements are classified into five classes: very low, low, moderate, high, and very high. The results showed that only limited settlements having very high (37) and high (58) CAI. The area statistics revealed that 44.5% of the area classified under moderate CAI category followed by low CAI (34.9%) and high CAI (8.87%).

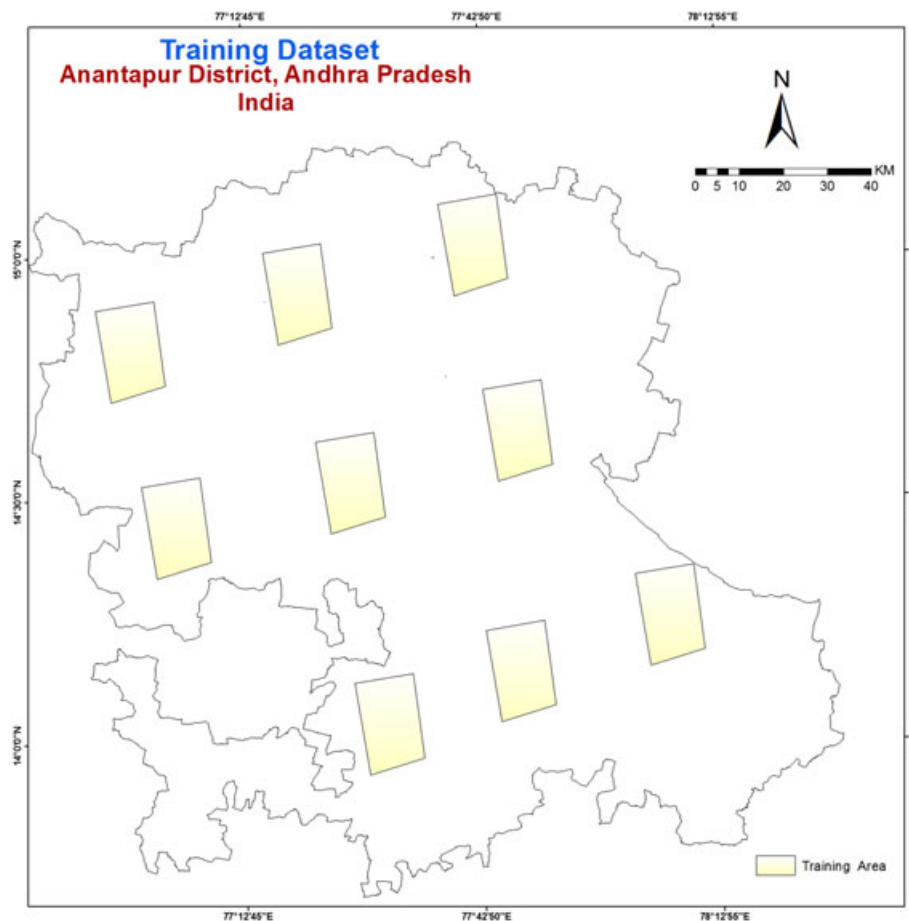


FIGURE 3 Training areas used in the environmental modeling [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Statistics of components of socioeconomic development index

Parameter	Communication index	Transport index	Education index	Medical index	Cumulative amenities index	Economic development index
Mean	0.23	0.21	0.04	0.09	0.56	2.73
Standard deviation	0.18	0.11	0.06	0.12	0.38	2.04

Economic development index: Economic development describes the well-being of society as a whole. The positive correlation between land vulnerability and low income regions is reported by many researchers across the world (Salvati et al., 2011; Vu et al., 2014). The main reason for being that a lack of financial resources and alternatives can lead to increasing intensification and over utilization of natural resources (Kelly et al., 2015). We considered population density of the settlement and proportion of employed population in the total population for calculating economic development index for Anantapur District. The district has a population of 4,081,148 with 968,160 households. The population density is 213 per Km² (census, 2011). Population growth has an always strong positive effect on the extent of land degradation (Vu et al., 2014) especially in rural areas. The total working population of the district is 49.89% that includes main (41.16%) and marginal workers (8.74%), and 50.1% are classified as nonworkers. Economic development index was calculated (see Equation 4) based on population density and proportion of employed population of the settlement ranged from 0 to 15.28 with the mean and standard deviation of 2.72 and 2.04. The results of the economic

index classification showed that 24 and 98 settlements have very high and high economic index, respectively. Area statistics showed that 49.7% was classified as moderate economic development index followed by very low (21.5%) and low (19.7%) category.

Cumulative amenities index and economic development index are integrated in the GIS environment, and the output was classified into five classes based on expert knowledge. Very low economic development and cumulative amenities index resulted in very high degree of desertification and vice versa (Kelly et al., 2015). Well-developed economic, social, institutional, and cultural domains are crucial for communities being resilient against desertification vulnerability. The final results showed that 38.8% of the area has low socioeconomic development followed by moderate SEI (34.9%).

3.2 | Biophysical index (BPI)

3.2.1 | Soil index

Soil is an important factor for evaluating the environmental vulnerability of any ecosystem. Soil properties related to desertification by

affecting water retention capacity and erosion resistance. The soil quality index was classified into five classes based on its natural ranges. The results reveal that 29.39% of area is classified as high category followed by very low (21.14%) and moderate quality soils (20.7%).

3.2.2 | Climatic index

A climate index was calculated based on the level of aridity. Average annual rainfall of the Anantapur District ranged from 381 to 693 mm with a mean (μ) and standard deviation (σ) of 552 and 67 mm. The aridity index was developed based on annual rainfall and potential evapotranspiration that is ranged from 0.07 to 0.32 with the mean (μ) of 0.25 and standard deviation of 0.02 (σ). Based on aridity index, the district divided into two climatic classes. A semiarid region covers 99% of Anantapur District and arid regions occupies 1% of TGA of Anantapur.

3.2.3 | Land utilization index

A Land utilization index was used for assessment of desertification vulnerability in place of Land management Quality index adopted in MEDALUS model (Kosmas et al., 1999). Land utilization index was calculated based on land capability classification and slope that gives better picture on how the lands are allotted to different land uses. Slope, land use, and land capability classification were integrated in a GIS environment and classified into three Land Utilization Index classes, that is, under-utilization, over-utilization, and optimal-utilization. The results revealed that 43.2% of lands are underutilized followed by 26.6% lands are optimal utilized and 26.6% of lands are over utilized.

3.3 | Desertification vulnerability index

The desertification vulnerability indices results (Table 4 and Figure 4) showed that 10.2% lands are classified as having a very high index. These lands are major priority lands of Anantapur District that needs immediate action to combat the desertification. About 18.7% of lands are classified as high, and 34.4% of lands are classified as moderate lands for desertification vulnerability.

3.4 | Prediction and extrapolation of desertification process using random forest model

We used an RFM for prediction of desertification/land degradation processes operating in Anantapur District. Desertification status map of Anantapur prepared under the Project Desertification status mapping of India-2nd cycle was used as a reference dataset for assessment

of prediction quality. A total of 7,373 pixels were selected in nine places as training dataset. Number of trees (n_{tree}) for model was finalized based on OOB error estimates. Since the OOB error estimates stabilized around 500 trees, parameter n_{tree} was fixed as 500 for final model. The number of variables tried at each split is three. The OOB classification error rate of the model is 14.46%. The confusion matrix results (Table 5) showed that out of 1,055 pixels of soil erosion class, 825 pixels are correctly predicted (user accuracy 78.1% and producer accuracy 82.7%). User and producer accuracy of vegetal degradation class is 87.5% and 88.3%, respectively. Out of 4,010 pixels of no desertification class, 3,743 pixels were correctly predicted (User accuracy-93.3%). Salinisation process was under predicted with user and producer accuracy of 38.7% and 85.6%, respectively, compared to other processes due to less number of pixels selected for training data. Overall accuracy of training datasets is 85.54% with the kappa index of 75.77% (Table 6).

RFM developed from training datasets was used to predict the desertification process for the entire district other than training regions. The overall accuracy rate and kappa index was calculated based on confusion matrix. The overall accuracy rate for the extrapolated areas is found to be good (71.01%) with the kappa index (51.8%). The difference between overall accuracy rate and kappa index indicates that all the classes are not predicted well. High user accuracy of 75.8% was recorded for vegetal degradation class followed by no desertification (71.0%), soil erosion (66.19%), and salinisation (54.0%) classes. We recorded a gap of 14.53% of overall accuracy rate between the predicting quality using OOB observations and testing datasets that is due to training and testing samples are located closely in training datasets compared to testing data (Muchoney & Strahler, 2002). The results obtained from this work presents a new approach for the prediction of desertification process for unvisited location using the available knowledge in training areas (Grinand et al., 2008; Vågen et al., 2016). The analysis of predicted map showed that 19.2%, 2.12%, and 19.77% of the district is affected by soil erosion, salinisation, and vegetal degradation processes, respectively (Table 7 and Figure 5) against 19.6%, 6.2%, and 22.5% of area actually mapped for soil erosion, salinisation, and vegetal degradation processes, respectively. This model well performed for soil erosion, vegetal degradation, and no desertification class and under predicted salinisation process. It may be due to the nature of the environmental variables used in the model (Adhikari et al., 2013) and variation in number of pixels selected for training. The prediction for salinisation could also be improved by using additional covariates such as soil electrical conductivity, parent material, and drainage. RFM model helps prediction of land degradation such as alkalinity from the predicted soil maps (Vågen et al., 2016), and knowledge of the desertification process at a particular location is effectively captured, we can able to extrapolate to surrounding areas based on spatial correlation of available predictors. The training site relies on surveyor experience and local landscape knowledge (Kidd et al., 2014). The information needed to train classification algorithms can be obtained for training areas through ground truth verification for quickly updating DSM maps at frequent time intervals by using this methodology. A another advantage of the RFM model that this model can be applied to any legacy data having desertification/land degradation processes to derive either new maps or to improve existing mapping (Kidd et al., 2014).

TABLE 4 Area under different vulnerability classes

DVI	Area (ha)	Area (%)
Very low	95929.95	5.0
Low	564356.9	29.4
Medium	659957.9	34.4
High	359347.4	18.7
Very high	196411.1	10.2
Settlement	11930.25	0.6
Waterbody	28781.87	1.5

Note. DVI = desertification vulnerability index.

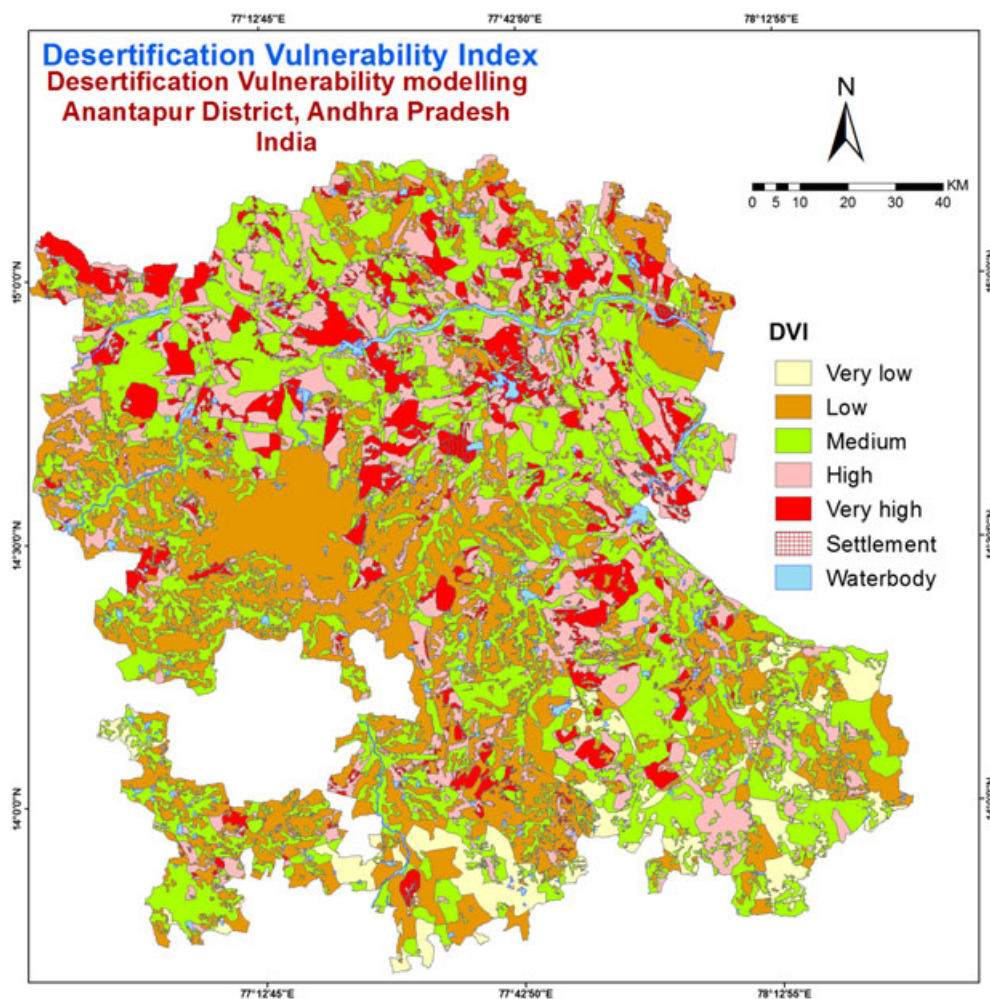


FIGURE 4 Desertification vulnerability index of Anantapur district [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

TABLE 5 Confusion matrix of training and testing datasets

Class	Training				Testing			
	No desertification	Salinisation	Vegetal degradation	Erosion	No desertification	Salinisation	Vegetal degradation	Erosion
No desertification	3743	33	124	110	22781	2632	1970	4709
Salinisation	314	220	18	20	92	208	11	25
Vegetal degradation	173	1	1519	43	1741	180	9048	971
Erosion	168	3	59	825	1926	114	675	4815

TABLE 6 Accuracy parameters of random forest models

Accuracy parameter	Training	Testing
Overall accuracy	85.54	71.01
Kappa index	75.77	51.80

TABLE 7 Actual and predicted area of desertification in Anantapur

S.No	Class	Percentage of total geographical area	
		Actual	Predicted
1	Soil erosion	19.6	19.24
2	Salinisation	6.2	2.12
3	Vegetal degradation	22.5	19.77
4	No desertification	51.7	58.99

3.5 | Importance of predictor variables

Importance of predictor variables on prediction of desertification processes in Anantapur District was assessed using variable importance function. We found desertification vulnerability index developed from climate, soil, land use, and socioeconomic condition emerged as most important predictor for prediction of desertification processes (Figure 6). These results confirm that the desertification vulnerability index developed from the present methodology helped to identify more vulnerable areas for preparation of combating plan.

Besides desertification vulnerability index, potential evapotranspiration is the next most important variable in predicting desertification processes in Anantapur District based on the calculated mean decrease in accuracy. Vegetation index, NDVI is the next most important predictor in predicting the desertification followed by coarse fragments and

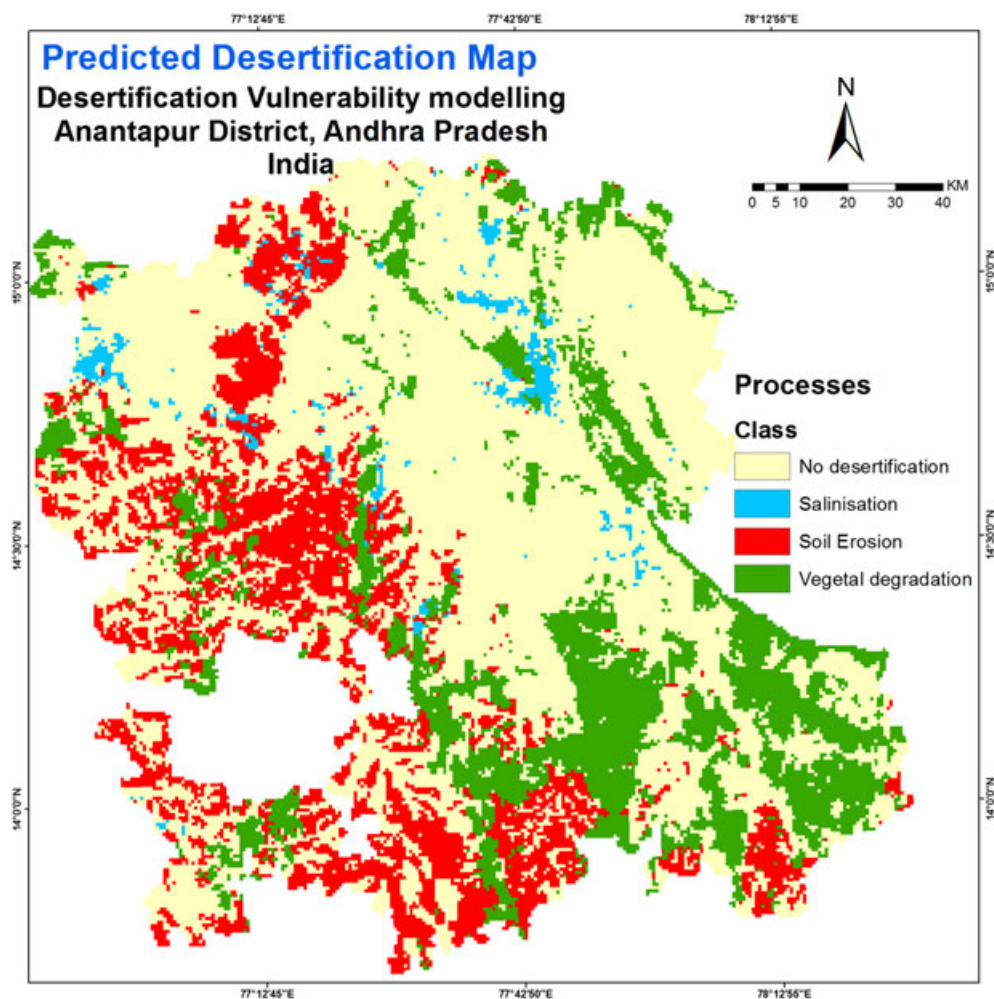


FIGURE 5 Predicted desertification process map of Anantapur district [Colour figure can be viewed at wileyonlinelibrary.com]

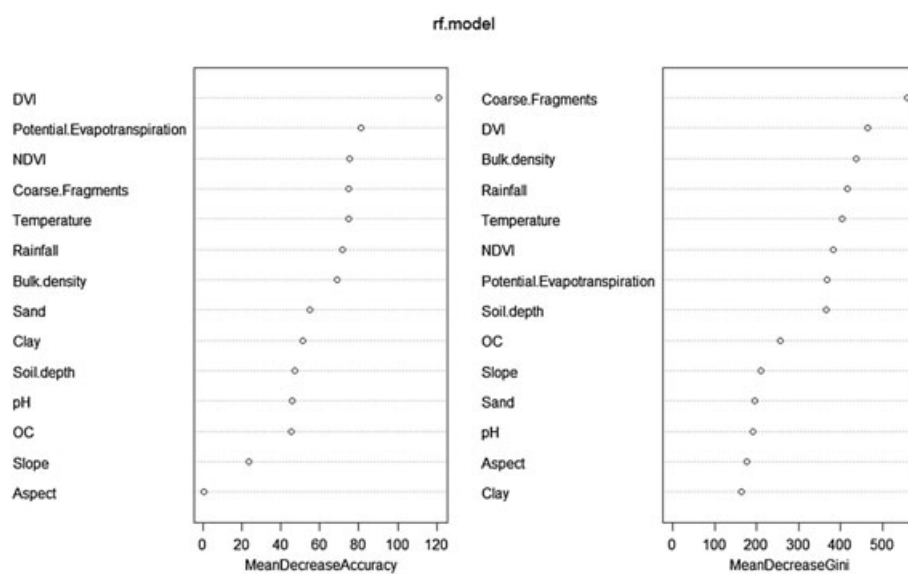


FIGURE 6 Variable importance ranking of random forest model. DVI = desertification vulnerability index; NDVI = Normalized Difference Vegetation Index

temperature. Many studies reported precipitation and temperature as the most important predictors in predictive soil mapping process (Akpa et al., 2014; Vågen et al., 2016). Coarse fragments emerged as most important variable among the seven soil parameters used for

prediction followed by bulk density and sand. Presence of coarse fragments act as a mulch and reduces the desertification process as reported by many researchers (Chow & Rees, 1998; Chow, Rees, Monteith, Toner, & Lavoie, 2007). The terrain parameters slope and

aspect are least important predictor as per mean decrease in accuracy score. This could be due to more than 80% of lands in Anantapur District covered by very gently to gently sloping lands.

4 | CONCLUSION

We prepared a desertification vulnerability index based on climate, soil, vegetation, and SEI combined in a multivariate index model. About 10.2% lands are identified as a very highly vulnerable area, and these areas need an immediate combating plan to arrest the desertification. We also tried to predict the desertification map based on training set data and compared with already prepared desertification status map of Anantapur. We used a recent machine language technique RFM for prediction of different desertification processes operating in Anantapur District. We recorded overall accuracy rate of 85.54% and 71.01% for OOB and test sites, respectively. The RFM model predicted well for vegetal degradation, soil erosion, and no desertification process and salinisation process was under predicted. We observed desertification vulnerability index emerged as top most predictor followed by potential evapotranspiration and NDVI. This technique can be extrapolated to similar regions/countries to prepare the desertification/land degradation maps and also helps to quick update maps in every 5 or 10 years for preparation of comprehensive planning to combat desertification.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Education and Training, Government of Australia for the 2016 Endeavour Research Fellowship and Sydney Institute of Agriculture, The University of Sydney for providing their support to conduct this study. The authors also would like to thank anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

ORCID

Subramanian Dharumarajan  <http://orcid.org/0000-0002-0614-2101>

REFERENCES

- Adhikari, K., Kheir, R. B., Greve, M. B., Bøcher, P. K., Malone, B. P., Minasny, B., ... Greve, M. H. (2013). High-resolution 3-D mapping of soil texture in Denmark. *Soil Science Society of America Journal*, 77, 860–876. <https://doi.org/10.2136/sssaj2012.0275>
- Ajai, A. A. S., Dhinwa, P. S., Pathan, S. K., & Ganesh, R. K. (2009). Desertification/land degradation status mapping of India. *Current Science*, 97, 1478–1483.
- Akpa, S. I. C., Odeh, I. O. A., Bishop, T. F. A., & Hartemink, A. E. (2014). Digital mapping of soil particle-size fractions for Nigeria. *Soil Science Society of America Journal*, 78, 1953–1966. <https://doi.org/10.2136/sssaj2014.05.0202>
- Breiman, L. (2001). Random forests. *Machine Learning*. <https://doi.org/10.1023/A:1010933404324>
- Census of India. (2011). House listing and housing census schedule. Government of India, Anantapur
- Chow, T. L., & Rees, H. W. (1998). Effects of coarse-fragment content and size on soil erosion under simulated rainfall. *Canadian Journal Soil Science*, 75, 227–232. <https://doi.org/10.4141/cjss95-031>
- Chow, T. L., Rees, H. W., Monteith, J. O., Toner, P., & Lavoie, J. (2007). Effects of coarse fragment content on soil physical properties, soil erosion and potato production. *Canadian Journal Soil Science*, 87, 565–577. <https://doi.org/10.4141/CJSS07006>
- Dasgupta, A., Sastry, K. L. N., Dhinwa, P. S., Rathore, V. S., & Nathawat, M. S. (2013). Identifying desertification risk areas using fuzzy membership and geospatial technique—A case study, Kota District, Rajasthan. *Journal of Earth System Science*, 122, 1107–1124.
- Dharumarajan, S., Lalitha, M., Natarajan, A., Naidu, L. G. K., Balasubramanian, R., Hegde, R., ... Singh, S. K. (2017a). Biophysical and socio-economic causes for increasing fallow land in Tamil Nadu. *Soil use and management*. <https://doi.org/10.1111/sum.12361>
- Dharumarajan, S., Hegde, R., & Singh, S. K. (2017b). Spatial prediction of major soil properties using random forest techniques—A case study in semi-arid tropics of South India. *Geoderma Regional*, 10, 154–162. <https://doi.org/10.1016/j.geodrs.2017.07.005>
- Dobos, E., Carré, F., Hengl, T., Reuter, H. I., & Tóth, G. (2006). Digital soil mapping as a support to production of functional maps. Office for official publications of the European Communities, Luxembourg. EUR, 22123, 68
- Dregne, H. E., & Chou, N. T. (1992). Global desertification dimensions and costs. In *Land Degradation and restoration of arid lands*. Lubbock: Texas Tech. University.
- Dutta, S., & Chaudhuri, G. (2015). Evaluating environmental sensitivity of arid and semiarid regions in northeastern Rajasthan, India. *Geographical Review*, 105, 441–461. <https://doi.org/10.1111/j.1931-0846.2015.12093.x>
- Ghosh, A., Sharma, R., & Joshi, P. K. (2014). Random forest classification of urban landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion. *Applied Geography*, 48, 31–41. <https://doi.org/10.1016/j.apgeog.2014.01.003>
- Giordano, L., Giordano, F., Grauso, S., Iannetta, M., Sciortino, M., Rossi, L., & Bonati, G. (2003). *Identification of areas sensitive to desertification in Sicily Region ENEA (eds)*. Italy: Rome.
- Grimm, R., Behrens, T., Märker, M., & Elsenbeer, H. (2008). Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using random forests analysis. *Geoderma*, 146, 102–113. <https://doi.org/10.1016/j.geoderma.2008.05.008>
- Grinand, C., Arrouays, D., Laroche, B., & Martin, M. P. (2008). Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143, 180–190. <https://doi.org/10.1016/j.geoderma.2007.11.004>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. Data mining, inference, and prediction*. 2nd ed. New York: Springer.
- IMD (2013). *Climate of Andhra Pradesh*. Govt of India: India Meteorological Department
- Jafari, R., & Bakhshandehmehr, L. (2013). Quantitative mapping and assessment of environmentally sensitive areas to desertification in central Iran. *Land Degradation & Development*. <https://doi.org/10.1002/ldr.2227>
- Karamesouti, M., Detsis, V., Kounalaki, A., Vasiliou, P., Salvati, L., & Kosmas, C. (2015). Land-use and land degradation processes affecting soil resources: Evidence from a traditional Mediterranean cropland (Greece). *Catena*, 132, 45–55. <https://doi.org/10.1016/j.catena.2015.04.010>
- Kelly, C., Ferrara, A., Wilson, G. A., Ripullone, F., Nollé, A., Harmer, N., & Salvati, L. (2015). Community resilience and land degradation in forest and shrubland socio-ecological systems: Evidence in Gorgoglione, Basilicata, Italy. *Land Use Policy*, 46, 11–20. <https://doi.org/10.1016/j.landusepol.2015.01.026>
- Kempen, B., Brus, D. J., Heuvelink, G. B. M., & Stoorvogel, J. J. (2009). Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151, 311–326. <https://doi.org/10.1016/j.geoderma.2009.04.023>
- Kidd, D. B., Malone, B. P., McBratney, A. B., & Minasny, B. W. M. A. (2014). Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia. *Soil Research*, 52, 107–119. <https://doi.org/10.1071/SR13100>

- Kim, J., Grunwald, S., Rivero, R. G., & Robbins, R. (2012). Multi-scale modeling of soil series using remote sensing in a wetland ecosystem. *Soil Science Society of America Journal*, 76, 2327. <https://doi.org/10.2136/sssaj2012.0043>.
- Kosmas, C., Ferrara, A., Briassoulis, H., & Imeson, A. (1999). Methodology for mapping Environmentally Sensitive Areas (ESAs) to Desertification. In C. Kosmas, M. Kirkby, & N. Geeson (Eds.), *The Medals project: Mediterranean desertification and land use Manual on key indicators of desertification and mapping environmentally sensitive areas to desertification* (pp. 31–47). Brussels: European Union 18882. ISBN: 92-828-6349-2.
- Kovačević, M., Bajat, B., & Gajić, B. (2010). Soil type classification and estimation of soil properties using support vector machines. *Geoderma*, 154, 340–347. <https://doi.org/10.1016/j.geoderma.2009.11.005>
- Lagacherie, P., Legros, J., & Burrough, P. (1995). A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. *Geoderma*, 65, 83–301. [https://doi.org/10.1016/0016-7061\(94\)00040-H](https://doi.org/10.1016/0016-7061(94)00040-H)
- Lawrence, R. L., Bunn, A., Powell, S., & Zambon, M. (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing and Environment*, 90, 331–336. <https://doi.org/10.1016/j.rse.2004.01.007>
- Liaw, A., & Wiener, M. (2002). Classification and regression by RandomForest. *R News*, 2, 18–21.
- Mansuy, N., Thiffault, E., Paré, D., Bernier, P., Guindon, V. P., Poirier, V., & Beaudoin, A. (2014). Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbor method. *Geoderma*, 235–236, 59–73. <https://doi.org/10.1016/j.geoderma.2014.06.032>
- McBratney, A. B., Mendonça, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117, 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Muchoney, D. M., & Strahler, A. H. (2002). Pixel-and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data. *Remote Sensing and Environment*, 84, 290–299. [https://doi.org/10.1016/S0034-4257\(02\)00006-8](https://doi.org/10.1016/S0034-4257(02)00006-8)
- NBSS&LUP. (2008). Soil resources of Anantapur District, Andhra Pradesh. NBSS&LUP publication No. 1017. NBSS&LUP, Nagpur, India
- Okun, O., & Priisalu, H. (2007). Random forest for gene expression based cancer classification: Overlooked issues. In J. Martí, J. M. Benedí, A. M. Mendonça & J. Serrat (Eds.), *Pattern recognition and image analysis IbPRIA 2007. Lecture Notes in Computer Science*, vol 4478. Berlin, Heidelberg: Springer.
- Priori, S., Bianconi, N., & Constantini, E. A. C. (2014). Can γ -radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine learning methods. *Geoderma*, 226–227, 354–364. <https://doi.org/10.1016/j.geoderma.2014.03.012>
- Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., & Fischer, C. (2003). Land-cover change monitoring with classification trees using Landsat TM and ancillary data. *Photogrammetric Engineering and Remote Sensing*, 69, 784–793. <https://doi.org/10.14358/PERS.69.7.793>
- SAC. (2016). Desertification and land degradation atlas of India (based on IRS AWIFS data of 2011–13 and 2003–05), Space Applications Centre, ISRO, Ahmedabad, India, 219
- Salvati, L., & Zitti, M. (2008). Assessing the impact of ecological and economic factors on land degradation vulnerability through multiway analysis. *Ecological Indicators*, 9, 357–363. <https://doi.org/10.1016/j.ecolind.2008.04.001>
- Salvati, L., & Zitti, M. (2009). Convergence or divergence in desertification risk? Scale-based assessment and policy implications in a Mediterranean country. *Journal of Environmental Planning and Management*, 52, 957–970. <https://doi.org/10.1080/09640560903181220>
- Salvati, L., Zitti, M., Ceccarelli, T., & Perini, L. (2009). Developing a synthetic index of land vulnerability to drought and desertification. *Geographical Research*, 47, 280–291. <https://doi.org/10.1111/j.1745-5871.2009.00590.x>
- Salvati, L., Mancini, A., Roberta Gemmiti, B. S., & Carlucci, M. (2011). Socio-economic development and vulnerability to land degradation in Italy. *Regional Environment Change*, 11, 767. <https://doi.org/10.1007/s10113-011-0209-x>
- Sastry, G. S. (2011). Desertification vulnerability index information system to develop a combat plan of action. A Case Study of Bellary District. Publication No. CEENR/50. Centre for Ecological Economics and Natural Resources (CEENR), Bangalore
- Sehgal, J., & Abrol, I. P. (1994). *Soil Degradation in India: Status and Impact*. (p. 80). New Delhi, India: Oxford and IBH.
- Singh, G. (2009). Salinity-related desertification and management strategies: Indian experience. *Land Degradation & Development*, 20, 367–385. <https://doi.org/10.1002/ldr.933>
- Symeonakis, E., Karathanasis, N., Koukoulas, S., & Panagopoulos, G. (2016). Monitoring sensitivity to land degradation and desertification with the environmentally sensitive area index: The case of lesvos island. *Land Degradation & Development*, 27, 1562–1573. <https://doi.org/10.1002/ldr.2285>
- Trott, C., Menegoni, P., Manfredi Frattarelli, F. M., & Iannetta, M. (2015). Assessing desertification vulnerability on a local scale: The Castelporziano study case (central Italy). *Rendiconti Lincei*, 26, 421. <https://doi.org/10.1007/s12210-014-0362-5>
- UNCCD. (2014). Desertification the invisible frontline. <http://www.unccd.int/Lists/SiteDocumentLibrary/publications>
- UNEP (1992). *World Atlas of Desertification* (1st ed.). London: Edward Arnold Publication.
- Vågen, T. G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T., & Gumbrecht, T. (2016). Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma*, 263, 216–225. <https://doi.org/10.1016/j.geoderma.2015.06.023>
- Vu, Q. M., Le, Q. B., Frossard, E., & Vlek, P. L. G. (2014). Socio-economic and biophysical determinants of land degradation in Vietnam: An integrated causal analysis at the national level. *Land Use Policy*, 36, 605–617. <https://doi.org/10.1016/j.landusepol.2013.10.012>

How to cite this article: Dharumarajan S, Bishop TFA, Hegde R, Singh SK. Desertification vulnerability index—an effective approach to assess desertification processes: A case study in Anantapur District, Andhra Pradesh, India. *Land Degrad Dev*. 2018;29:150–161. <https://doi.org/10.1002/ldr.2850>