ABSTRACT

Plant diversity plays a dominant role in affecting the key ecological processes i.e., primary productivity, regulation of climate and hydrological cycle, maintenance of atmosphere, water and soil fertility, and the ecosystem health. However, the impact of contemporary climate and other threats (warming, deforestation, invasion etc.) on species has led to serious biodiversity crisis. India, one amongst the 12 mega-biodiversity countries of the world, harbours most unique species rich areas within ten biogeographically distinct zones. Understanding the distribution pattern of species and ability to predict them in space and time are thus imperative to meet the convention on biological diversity (CBD), Aichi biodiversity targets for 2020.

The main objective of this study is to analyse the distribution of plant species in present and future perspective and the putative determinants in different biogeographic zones of Indian mainland. A comprehensive national database on plant species from the project *'Biodiversity Characterization at Landscape Level'* was utilized to generate the maiden plant richness map at 1˚ grid level for the first time. The sampling sufficiency was evaluated at the 1˚ and 2˚ spatial grid levels at acceptable threshold limits (60% to 80%). Plant species richness was then related with a widely used set of climate, edaphic, topographic and disturbance variables via different statistical modelling approaches. The performance of individual model e.g. generalized linear model (GLM), generalized boosted model (GBM), random forest (RF) and support vector machine (SVM) was compared against their ensembles. The species distribution was predicted for future scenarios of climate using two different representative concentration pathways (RCP4.5 & RCP8.6).

Clench model was significantly fitted ($p < 0.001$) at both the grid levels, with a very high coefficient of determination ($r^2>0.95$). At an acceptable threshold limit of 70%, almost all the grids at the 2˚ level and > 80% of the grids at the 1˚ level were found to be sufficiently sampled. Sampling sufficiency was scale dependent as greater number of 2˚ grids attained asymptotic behaviour following the species–area curve.

This study revealed that majority of the selected variables were related to water and energy (r>40%), which is consistent with the widely documented trend of plant species driven by water-energy dynamics. Individually, water and temperature variables could show an agreement between 10% and 90% for different zones, which enhanced for their combination (40% to 95%). The ensemble model improved the results with a significant reduction in the RMSE and increase in correlation

coefficient. Elevation, vegetation indices, soil organic carbon (SOC) and human influence index, at individual level, could show their significant influence (r>25%) in deriving species distribution for different zones.

Continuous loss of species richness was predicted for both the future climate scenarios during 2050 and 2070. The semi-arid and arid zone showed maximum probability (69% and 52.5% grids) of species gain while, north-east zone showed maximum probability of species loss (80% and 93% grids) for these scenarios.

The present study proposes to lay out quadrats at 1° grid level on a priority basis for the under-sampled grids (46;16.6%), on the basis of heterogeneity and area of vegetation. In addition, this study proposes to set up conservation priority at the 1° grid level, on the basis of available area of vegetation, species richness, human influence index, protected area and probable loss and gains in species richness in the future. The database generated from this study will be helpful in linking and harmonizing various themes of raster data and can be aggregated at multiple resolutions to use them in relational database management systems.