Joint ICTP-IAEA Advanced School/Workshop on Machine Learning in Citizen Science

27 February 2023 - 03 March 2023

Projected Distribution of 218 Invasive Plant Species in India under climate change: Identification of Hotspots and Management Implications

JYOTI JAGDISH PRAJAPATI

DEPARTMENT OF MATHEMATICS,

INSTITUTE OF CHEMICAL TECHNOLOGY (ICT), MUMBAI

AND

AGRICULTURAL AND ECOLOGICAL RESEARCH UNIT,

INDIAN STATISTICAL INSTITUTE (ISI), GIRIDIH

Agenda

- Problem Statement
- Materials and Methods
- Results
- Future Direction

Invasive Alien Plants

Lantana Camara

Mikania micrantha

Jatropha Curcas

Catharanthus roseus

Why is this study important?

Alien plants growing together threatening tiger habitats: Study
20 January 2023

Think twice before you eat *Ponnaganti kura*
24 November 2017

- **India has numerous Invasive Alien species pose an unprecedented threat to biodiversity and ecosystems at different spatial scales.**
- **It is crucial to understand there current distribution and also under climate change.**

Species Distribution Modelling

- The Probability of Occurrence of each species in a particular location is modelled as a function of climatic Variables.
- This is typically obtained by fitting logistic regression model, F is the logistic function.

Materials and Methods

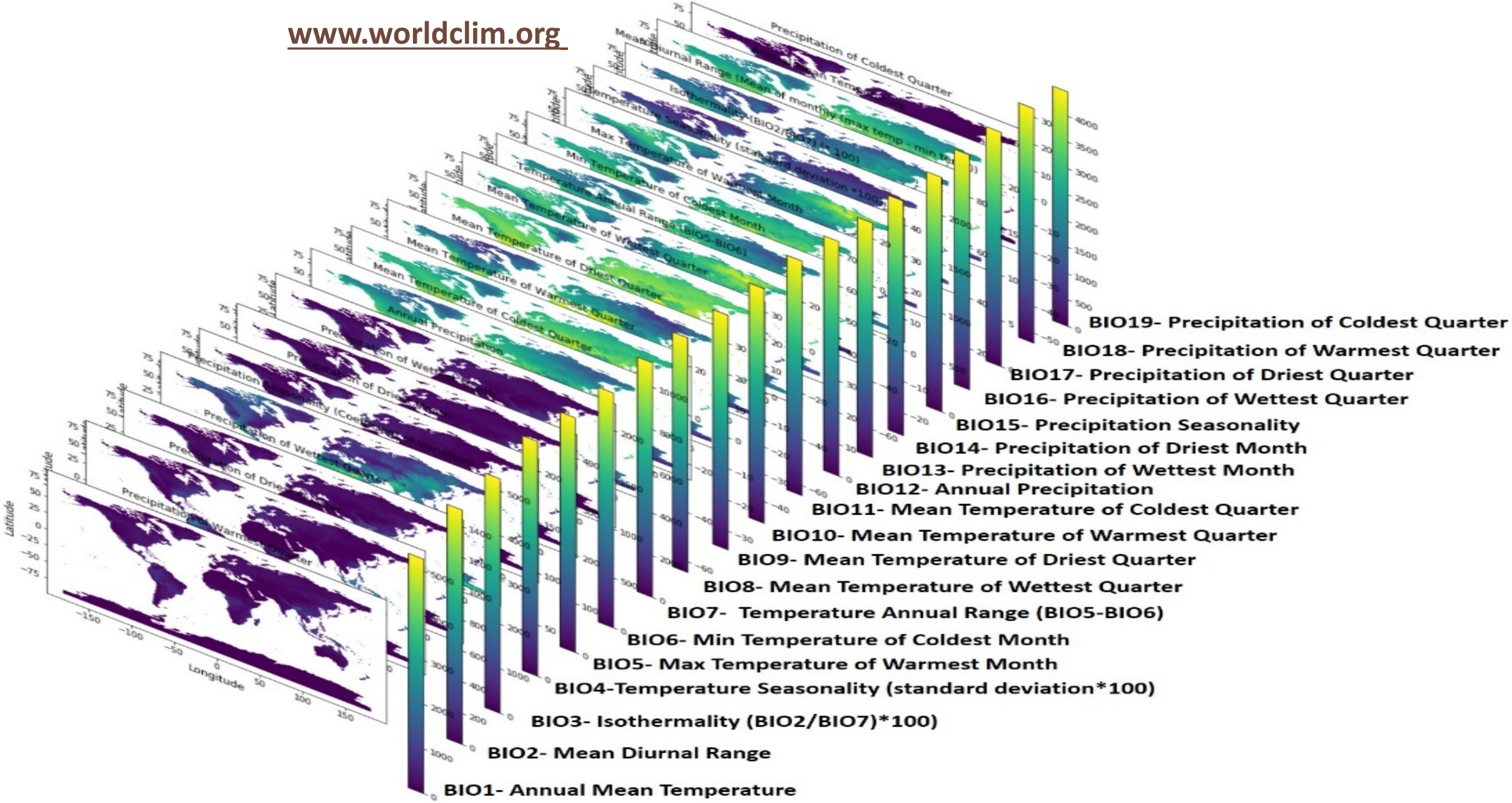
- Collection of occurrence records and Climate Data
- Processing of data
- Variable Selection Method
- Model building (Ensemble)
- Method Selection
- Modelling and Projections
- Categorization of species
- Post Modelling Analysis
- Results

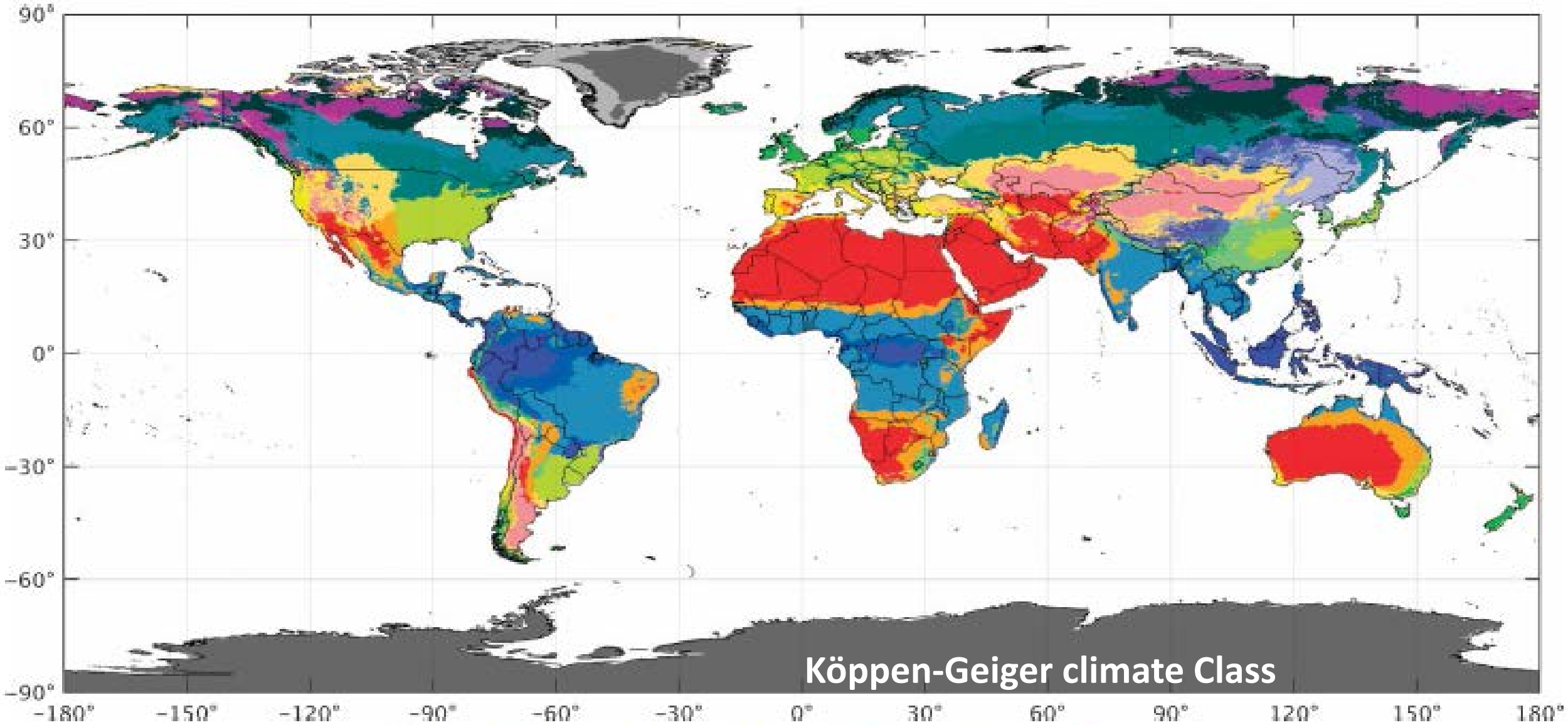
What is GBIF?

GBIF—the Global Biodiversity Information Facility—is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.

Bioclimatic Variable

www.worldclim.org





www.gloh2o.org/koppen

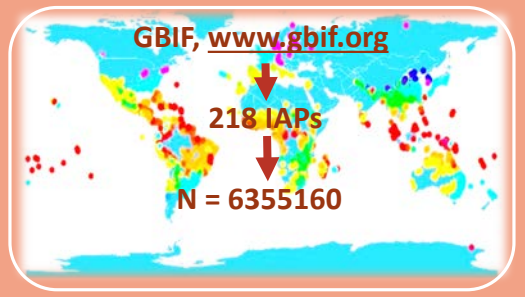
Data Collection

Species Selection

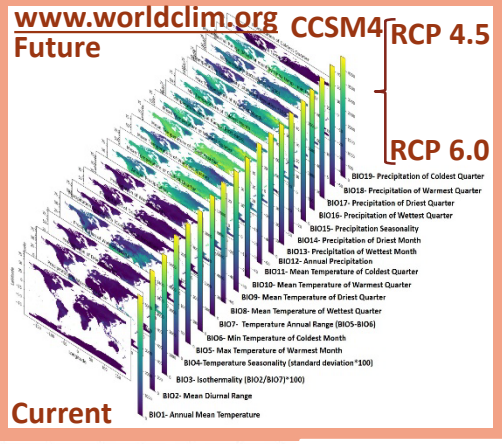
https://ilora2020.wixsite.com/ilora2020

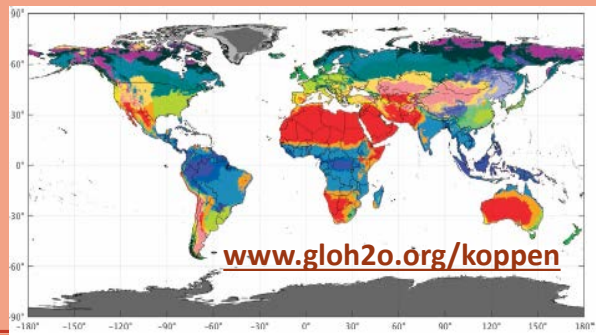
218 IAPs

Occurrence collection



Climate information





Köppen-Geiger climate Class

Data Cleaning

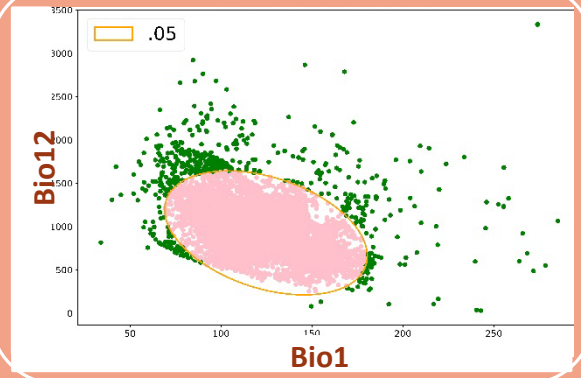
Removing duplicates ← Removing missing coordinates ← Cleaning of occurrence data In 4 steps

Removing points falling in ocean → Removing Sampling Bias

N = 1394752

Data Processing

Delimiting geographic range



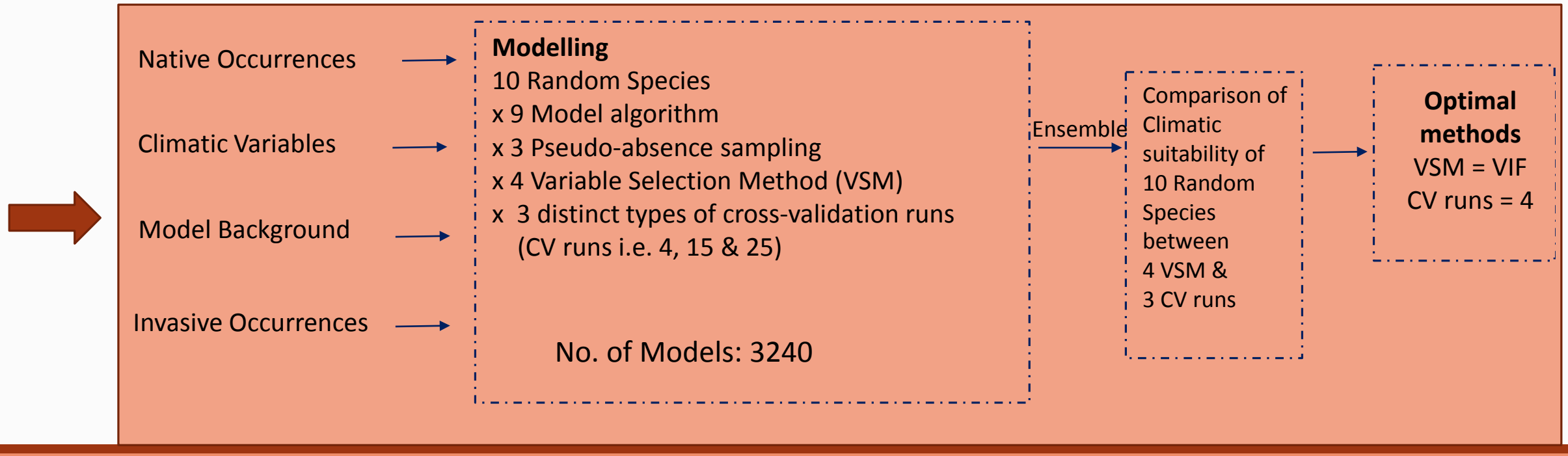
Background selection

Model Description

Consider, a collection of m independent covariates denoted by $\mathbf{X_0}$ **' = (**$\mathbf{x_1}$ **,** $\mathbf{x_2}$ **,…,** $\mathbf{x_{19}}$**), where** $\mathbf{x_1}$ **= BIO1,** $\mathbf{x_2}$**=BIO2,…,** $\mathbf{x_{19}}$ **=BIO19** There are total 19 covariates. m = 19. **Y=0/1, where 1 represent presence (occurrences) and 0 represent pseudo-absence**

- **Logistic Regression**
  - The logistic regression function for estimating probability of occurrence data on given values of environmental factors, **Pr(Y= 1|X) = π(X)** is given by

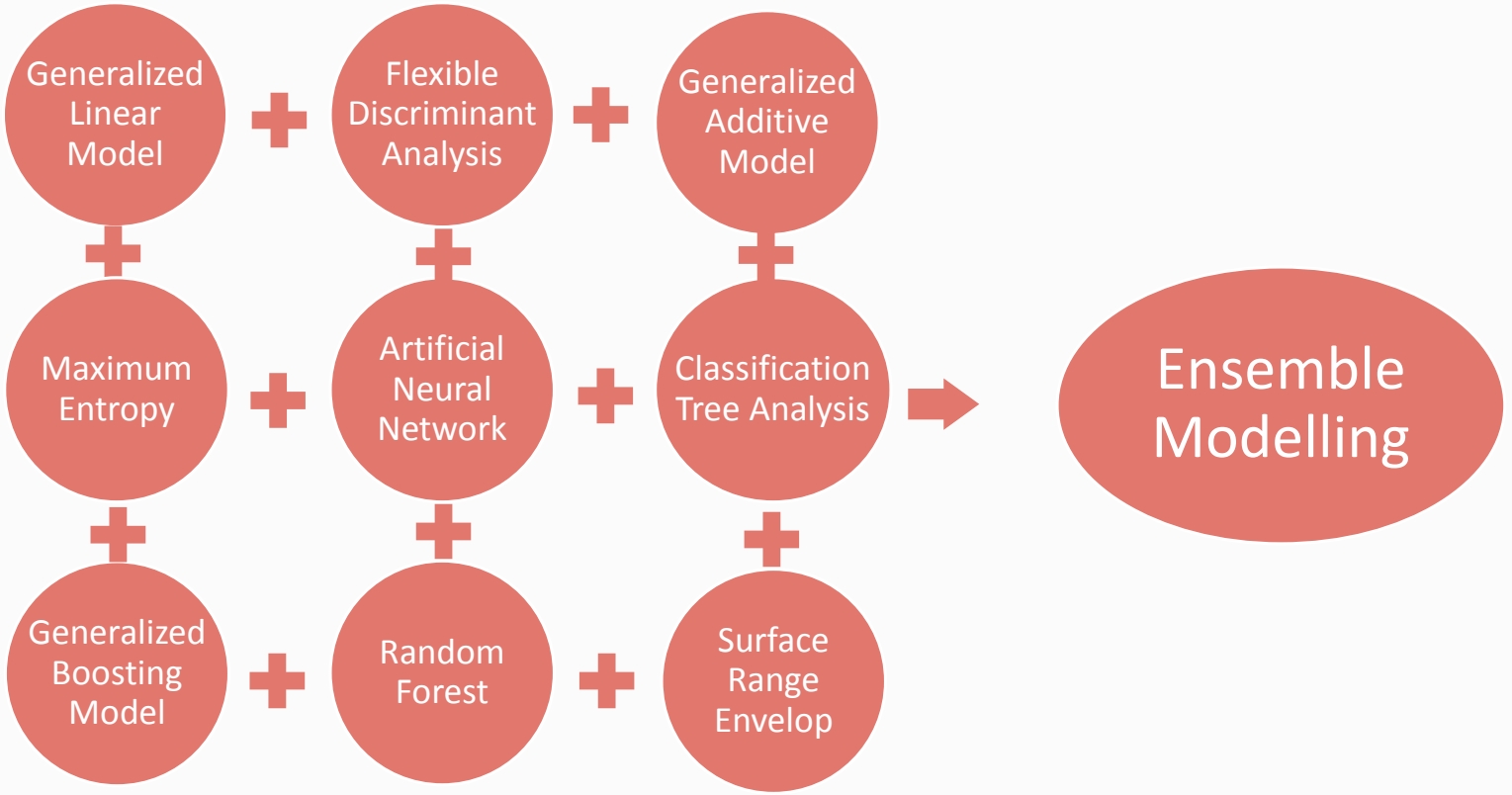
$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}}$ where $g(x) = \beta_0 + \sum_{i=1}^{19} \beta_i X_i$, $\beta_i' s$ are the coefficients (James et al 2013)



Method Selection

Models using Biomod2 package in R

1. Generalized Linear Model (Searle and McCulloch, 2001)
2. Flexible Discriminant Analysis (Hastie et al., 2009)
3. Generalized Additive Model (Guisan et al., 2002)
4. Maximum Entropy (Favretti, 2017)
5. Artificial Neural Network (Zhang, 2010)
6. Generalized Boosting Model (Einziger et al., 2019)
7. Random Forest (Nordhausen, 2014)
8. Surface Range Envelop (Hannah, 2012)
9. Classification Tree Analysis (Breiman et al., 2017)



Variable Selection method

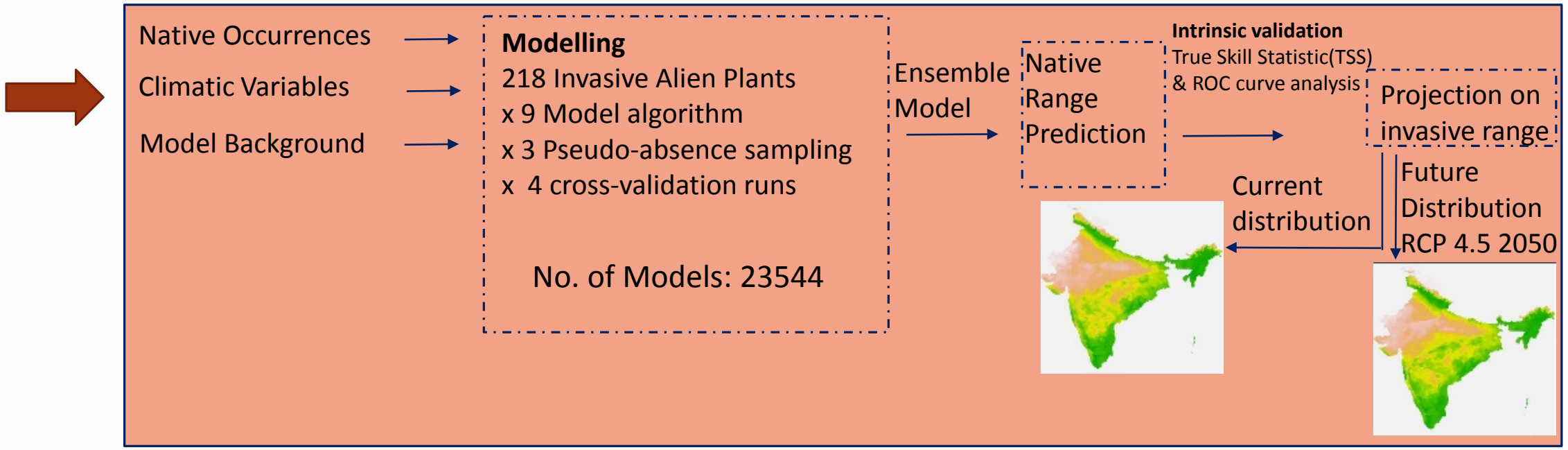
1) First four bioclimatic variables from the first two principal with highest loadings in PC1 & PC2 using PCA package in R (Guisan et al., 2017).

2) Variables with a correlation value less than 0.7 using stats package (Braunisch et al., 2013).

3) Principal components developed globally using bioclimatic variables using the kuenm package in R and used first 5 principal components as predictors (Cobos et al., 2019).

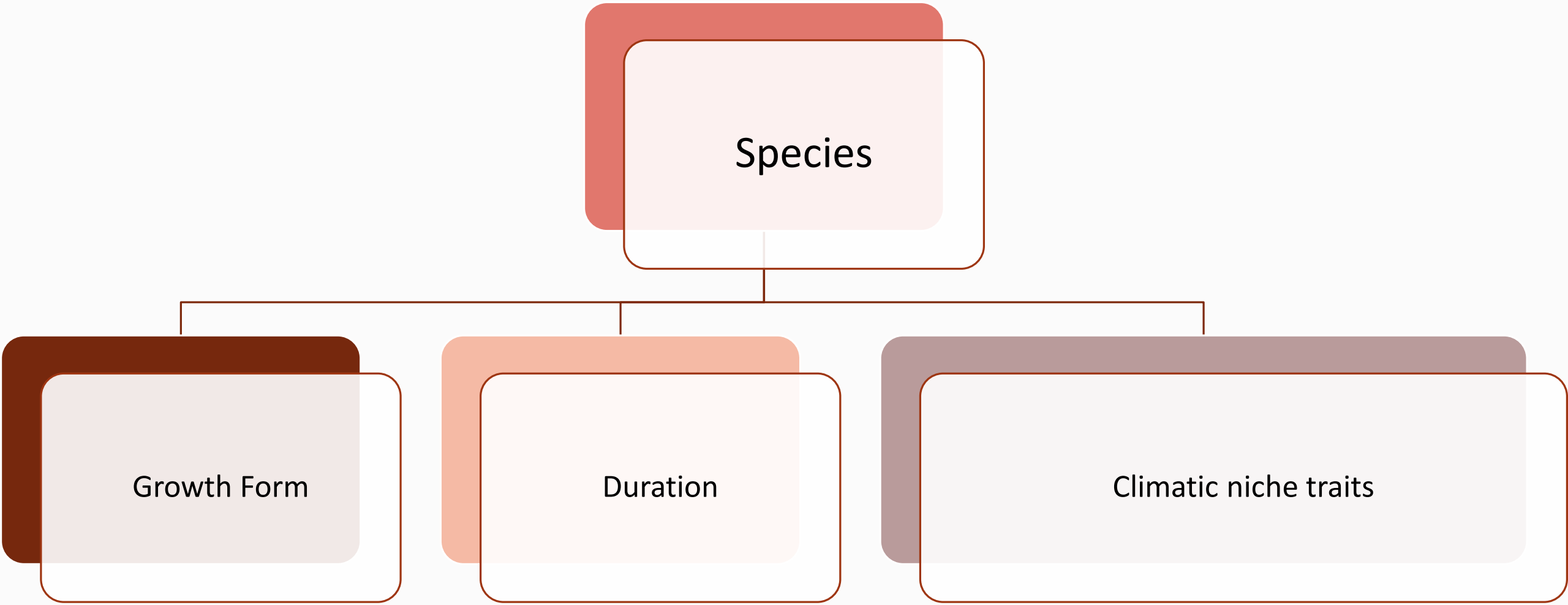
4) Variation Inflation Factor with a threshold as 5 using regclass package in R (Mpakairi et al., 2017; Rodríguez-Rey et al., 2019).

VARIATION INFLATION FACTOR WITH A THRESHOLD AS 5

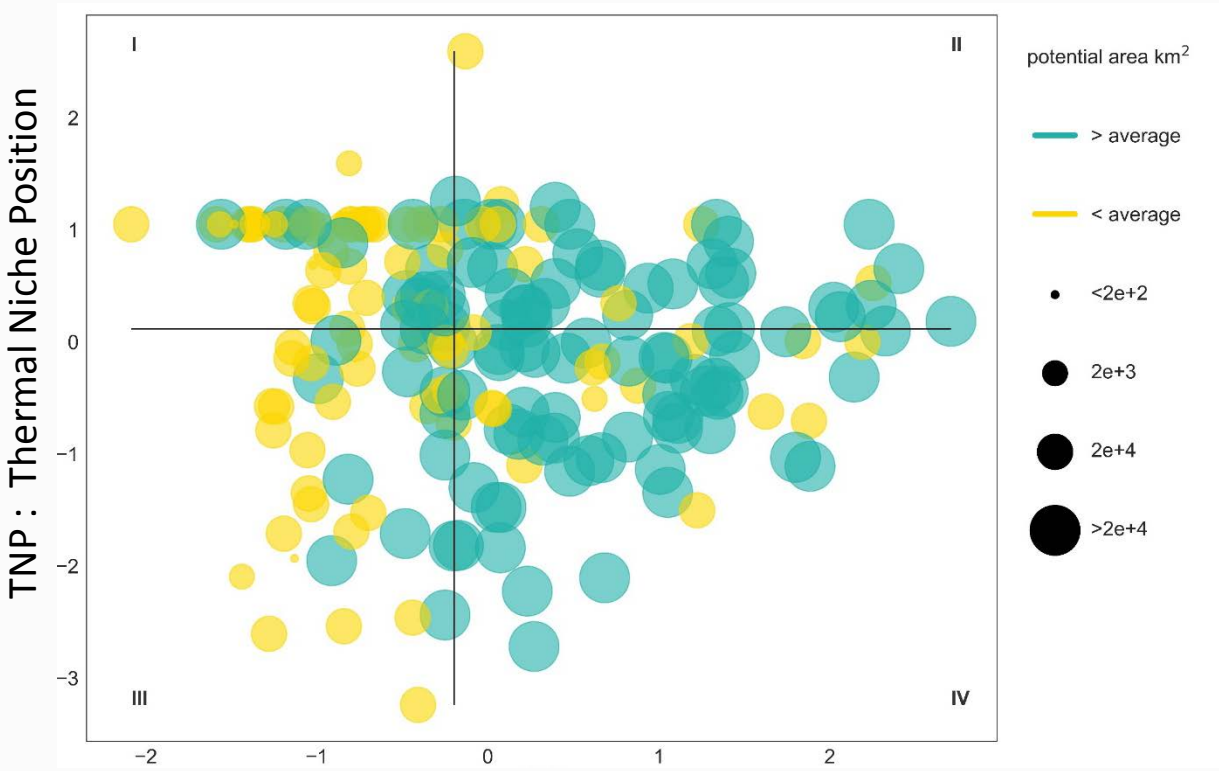


Modelling and Projections

Categorization of Species



Climatic niche traits

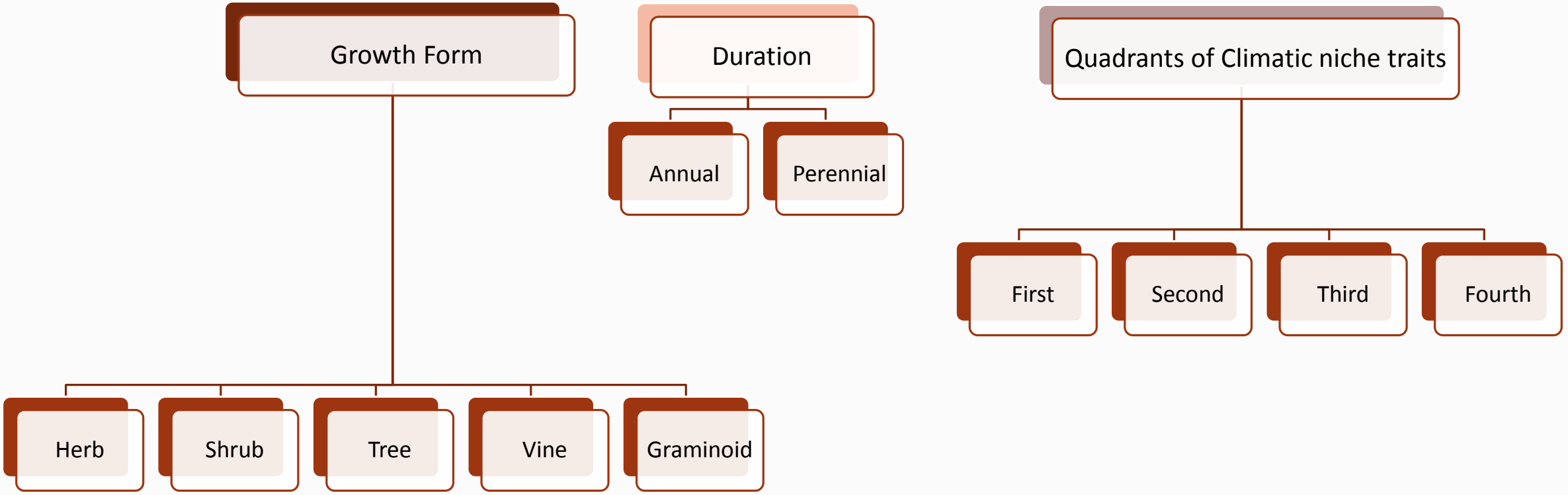


$$\mathrm{TNP} = \frac{\max(\mathrm{Bio5}) + \min(\mathrm{Bio6})}{2} - \mathrm{India}\left(\frac{\max(\mathrm{Bio5}) + \min(\mathrm{Bio6})}{2}\right)$$

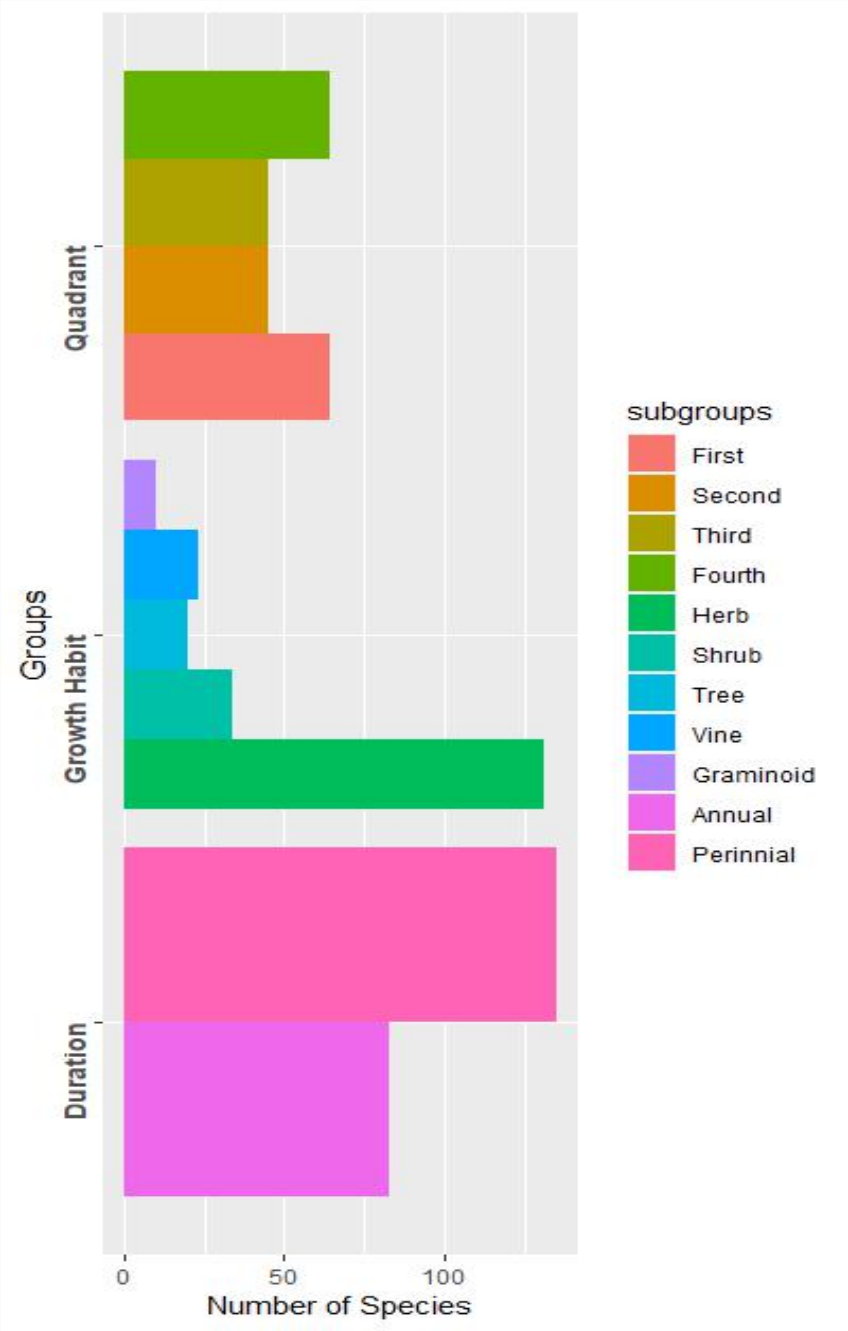
$$\mathrm{HNB} = \max(\mathrm{Bio13}) - \min(\mathrm{Bio14})$$

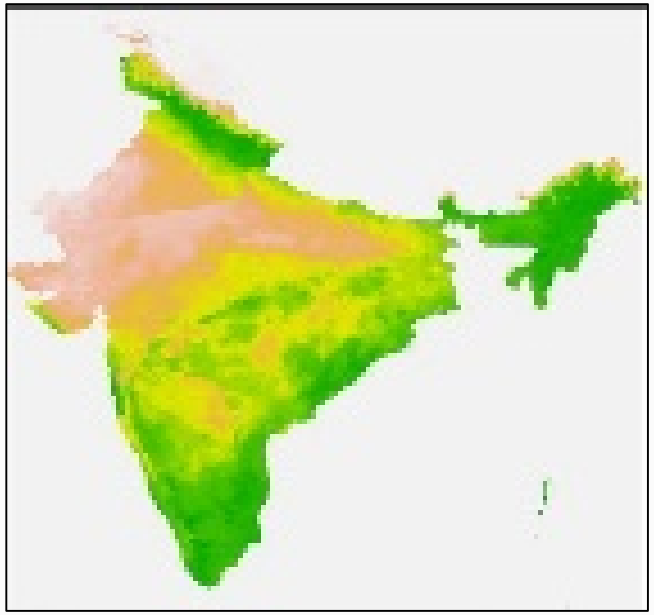
Bio5 : The maximum temperature in the warmest month
Bio6 : the minimum temperature in the coldest month
Bio13 : the maximum precipitation in the wettest month
Bio14 : the minimum precipitation in the driest month

Species


Growth Form
Herb
Shrub
Tree
Vine
Graminoid
Duration
Annual
Perennial
Quadrants of Climatic niche traits
First
Second
Third
Fourth

Species Distribution in Sub Categories

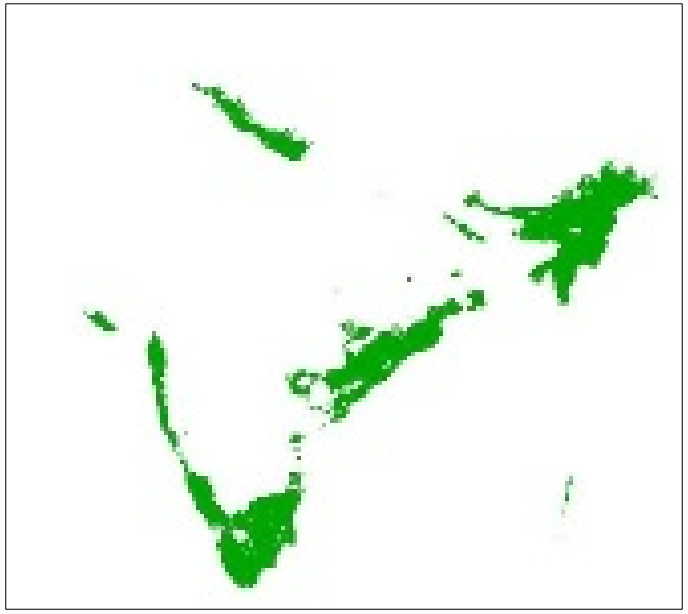


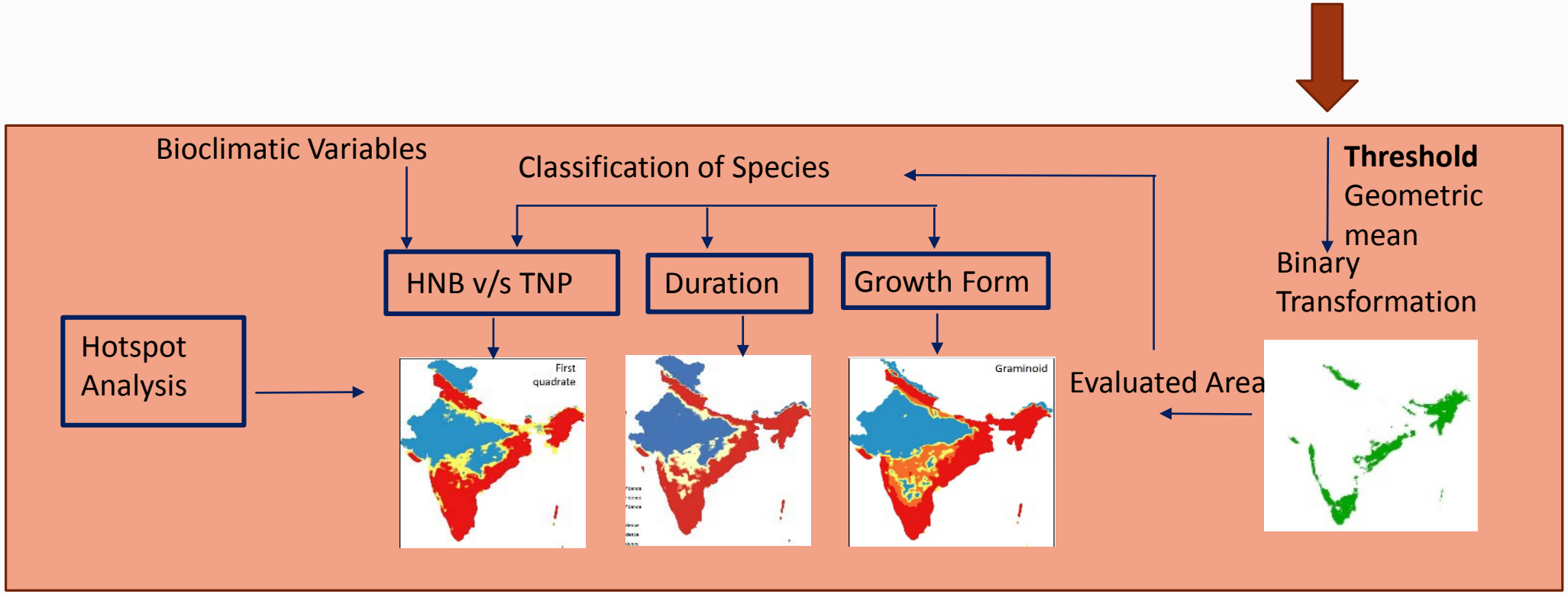


Threshold

Geometric Mean = $\sqrt{specificity * senstivity}$

(Diego et al., 2022; Tharwat, 2020)

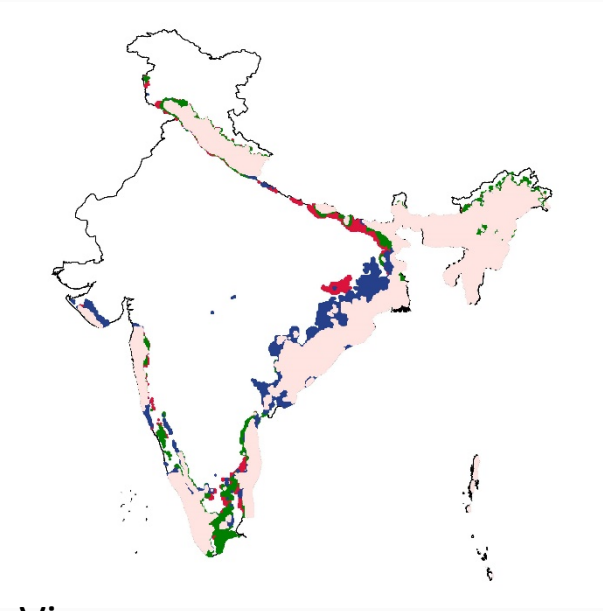




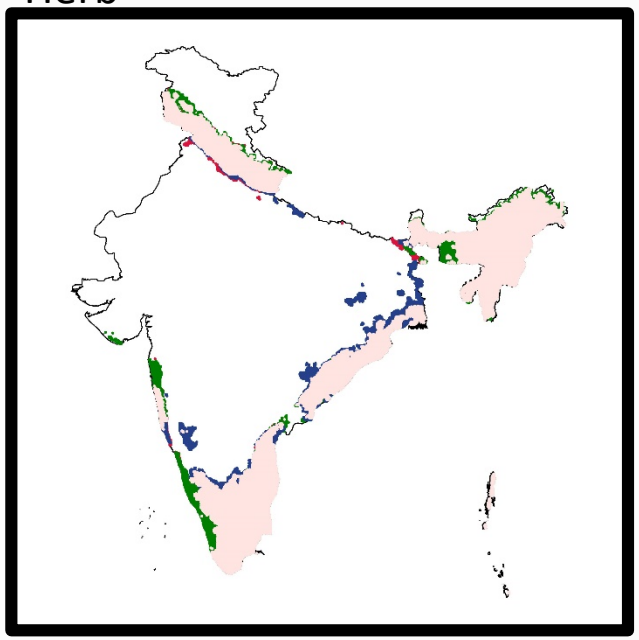
Post-Modelling Analysis

Growth Form : The 99% hotspot for current, 2050 and 2070 with overlapped region of current, 2050 and 2070 outlined as conserved region.

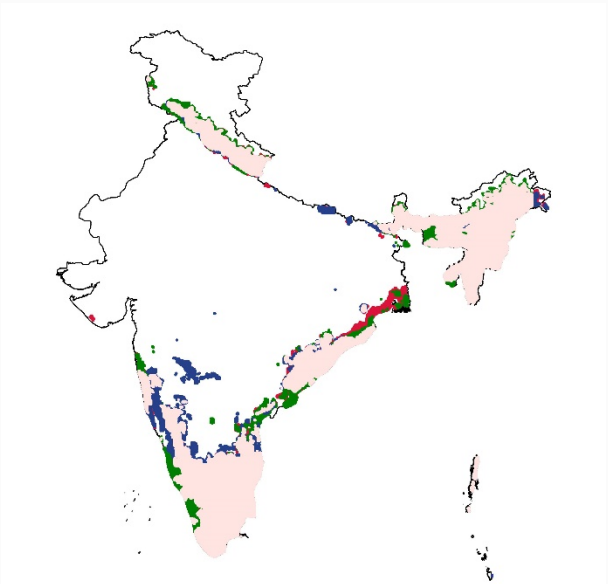
Graminoid



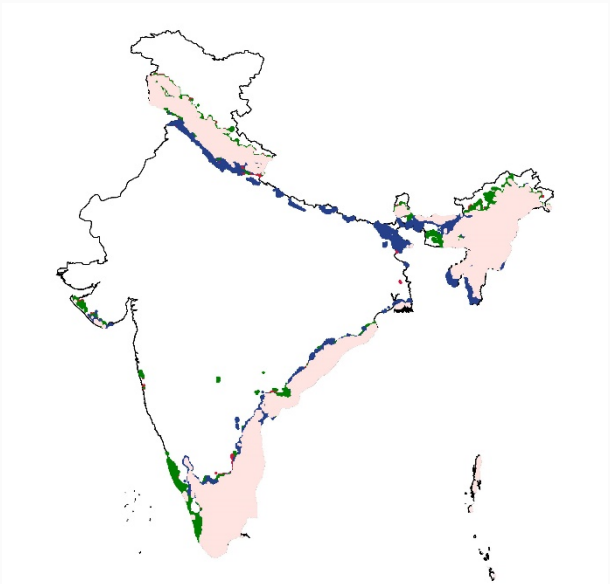
Herb



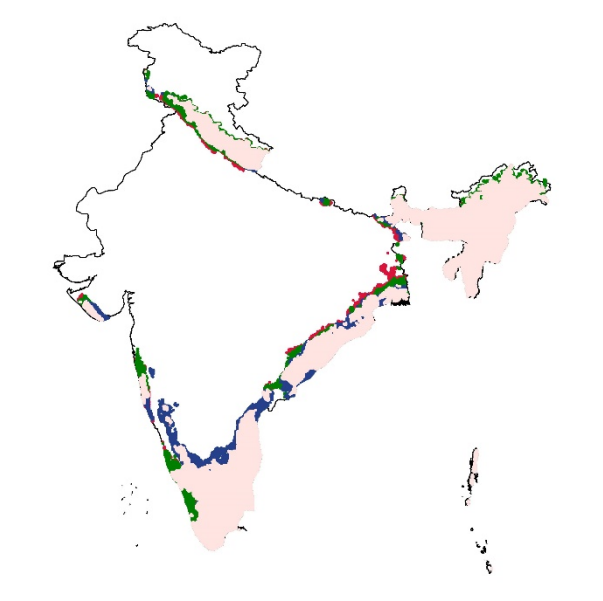
Shrub

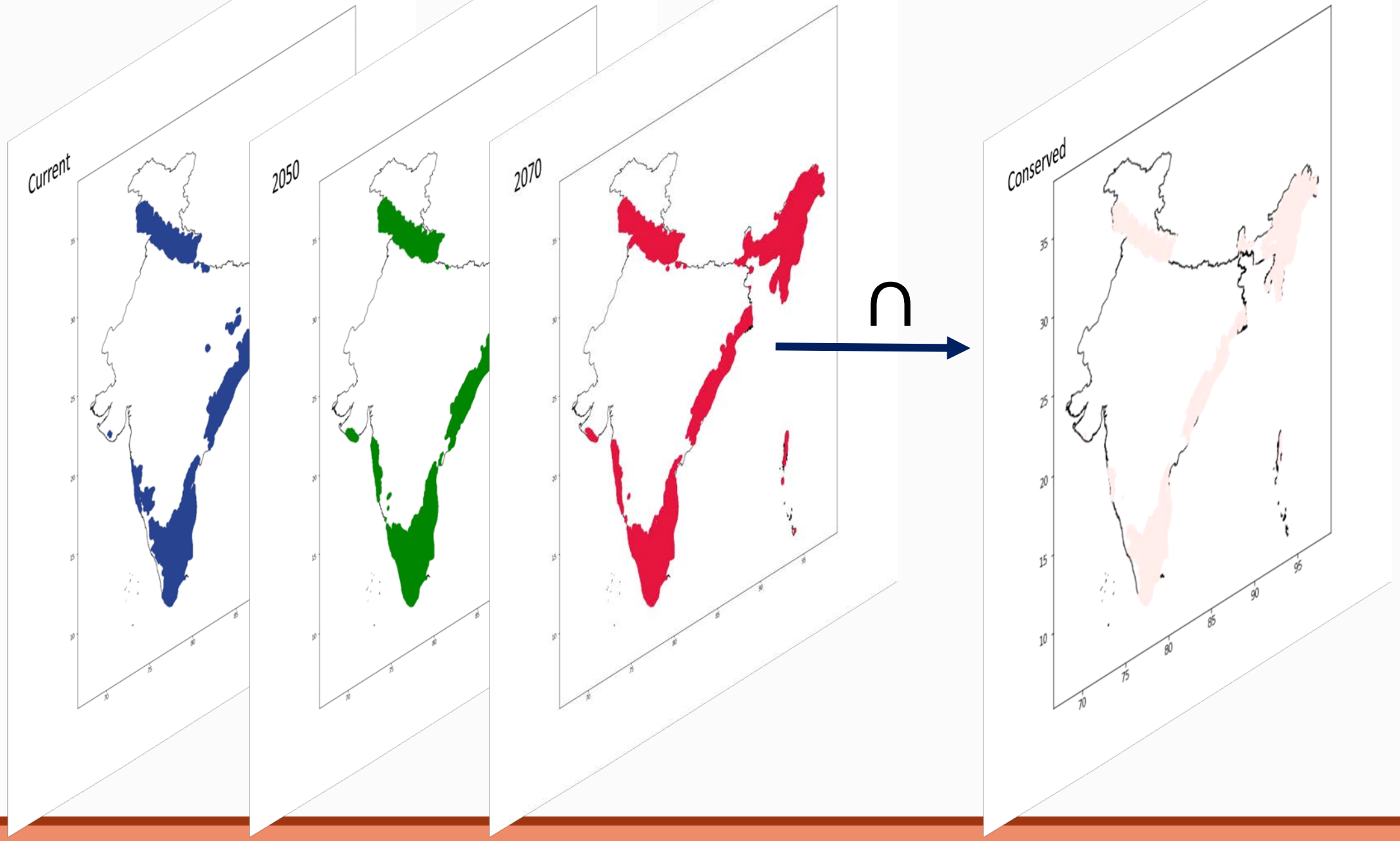


Tree



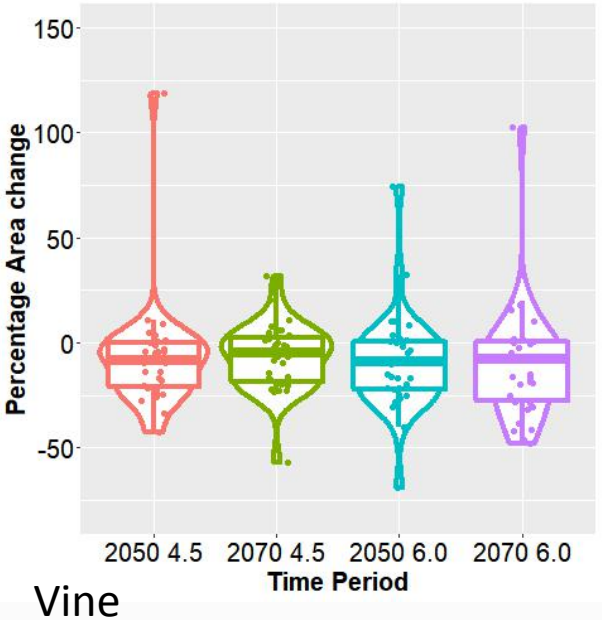
Vine



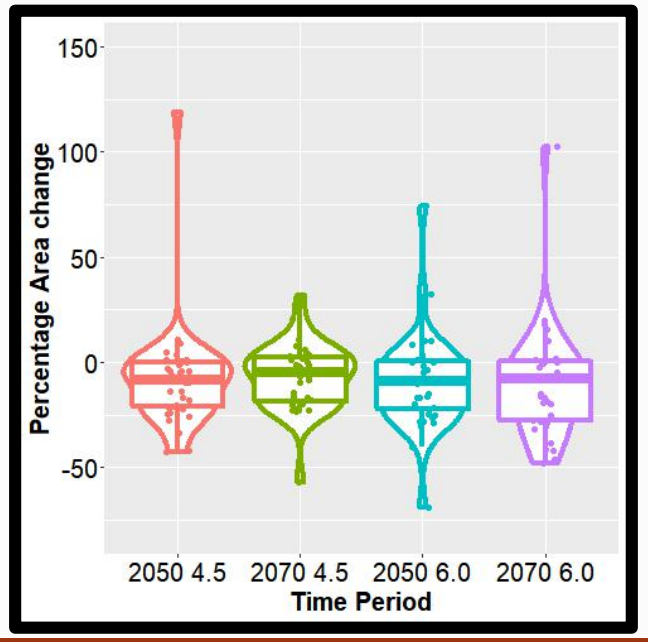

Current
2050
2070
Conserved
35
30
25
20
15
10
70
75
80
85
90
95

Growth Form : The comparison of overall area changes emerged in Time period 2050 and 2070 at RCP 4.5 and RCP 6.0 respectively.

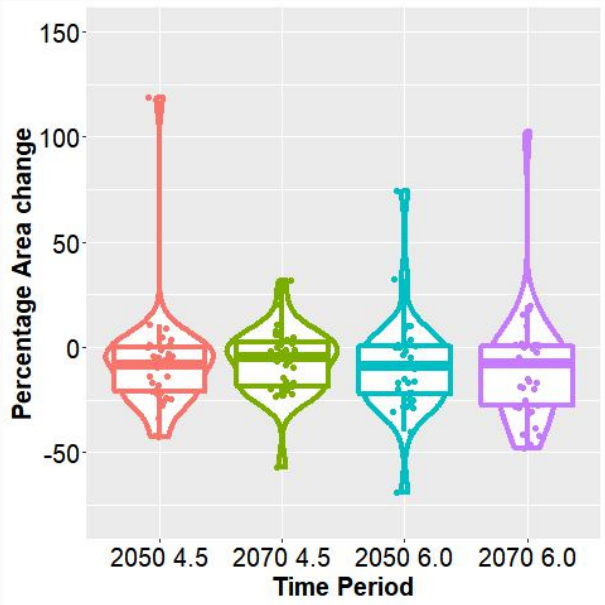
Graminoid



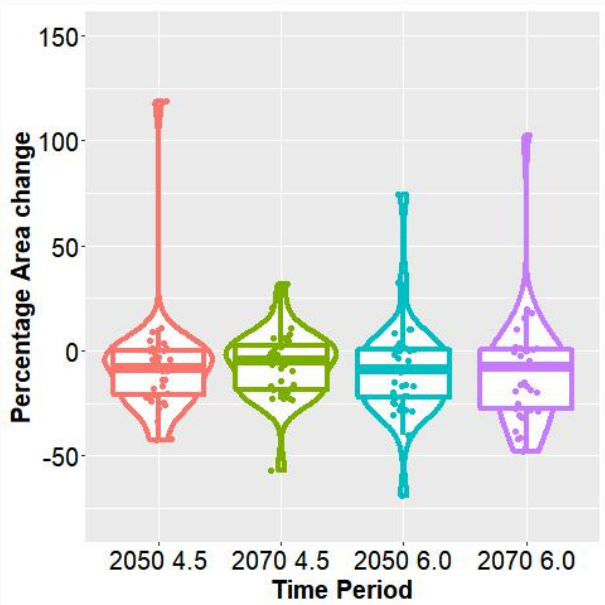
Herb



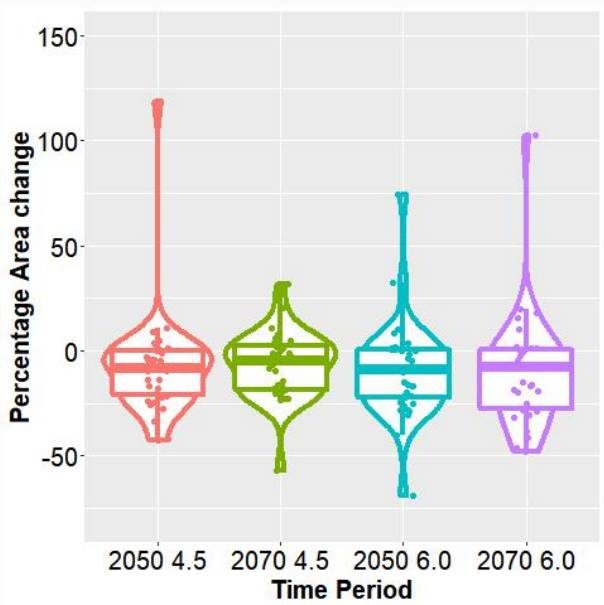
Shrub



Tree



Vine



List of Species showing expansion and contraction

GROUPS	RAPID EXPANSION (RCP4.5)	RAPID CONTRACTION (RCP 4.5)
GRAMINOID	Bromus catharticus Vahl	Pennisetum polystachyon Schult.
HERB	Oxalis pes-caprae L.	Youngia japonica (L.) DC.
SHRUB	Stachytarpheta cayennensis (Rich.) Vahl	Cylindropuntia ramosissima (Engelm.) F.M.Knuth
TREE	Acacia melanoxylon R.Br.	Prosopis juliflora (Sw.) DC.
VINE	Calopogonium mucunoides Desv.	Passiflora edulis Sims
ANNUAL	Acmella radicans (Jacq.) R.K.Jansen	Youngia japonica (L.) DC.
PERENNIAL	Oxalis pes-caprae L.	Cylindropuntia ramosissima (Engelm.) F.M.Knuth
FIRST	Oxalis pes-caprae L.	Cylindropuntia ramosissima (Engelm.) F.M.Knuth
SECOND	Calopogonium mucunoides Desv.	Oxalis latifolia Kunth
THIRD	Acmella radicans (Jacq.) R.K.Jansen	Cirsium arvense (L.) Scop.
FOURTH	Stachytarpheta cayennensis (Rich.) Vahl	Citharexylum spinosum L.

References:

- Hastie, T., Tibshirani, R., Friedman, J., (2009) Support vector machines and flexible discriminants, The elements of statistical learning. Springer, pp. 417-458.
- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J. (2017) dismo: Species distribution modeling. R package version 1, 1-1.
- Mpakairi, K.S., Ndaimani, H., Tagwireyi, P., Gara, T.W., Zvidzai, M., Madhlamoto, D. (2017) Missing in action: Species competition is a neglected predictor variable in species distribution modelling. PLoS One 12, e0181088.
- Mukherjee, A., Williams, D., Wheeler, G., Cuda, J., Pal, S., Overholt, W. (2012) Brazilian peppertree (Schinus terebinthifolius) in Florida and South America: evidence of a possible niche shift driven by hybridization. Biological Invasions 14, 1415-1430.
- Chowdhury, M.Z.I., Turin, T.C. (2020) Variable selection strategies and its importance in clinical prediction modelling. Family medicine and community health 8.
- Clarke, L., Edmonds, J., Jacoby, H., Pitcher, H., Reilly, J., Richels, R., (2007) Scenarios of Greenhouse Gas Emissions and Atmospheric Concentrations.
- Cobos, M., Peterson, A., Barve, N., Osorio, O.L., (2019) kuenm: An R package for detailed development of ecological niche models using Maxent. PeerJ.

Collaborators :

Dr. Amiya R. Bhowmick
ICT, Matunga

Dr. Abhishek Mukherjee,
ISI, Giridih

Dr. Achyut K. Banerjee,
Sun Yat-Sen University,
China

Abhijit Singh,
ISI, Giridih

Acknowledgement

- I would like to express special thanks to my supervisor Amiya Ranjan Bhowmick, co-guide Dr. Abhishek Mukherjee and Dr. Achyut K. Banerjee for giving me this opportunity to work on this topic and for keeping their believe in me.
- I would like to thank ICT and ISI for providing us logistic and financial support.
- I would like express gratitude to my respected family for giving me support in each circumstances.

Thank You !