ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =8.152, July-December 2025, Submitted in July 2025

Computational Multivariate Methods for Ecological Pattern Recognition and Biodiversity Prediction

Jag Pratap Singh Yadav, Research Scholar, Mathematics Department, Sikkim Alpine University, Namchi (Sikkim)
Dr. Umashanker, Assistant Professor, Physical Mathematics Department, Sikkim Alpine University, Namchi (Sikkim)

Abstract

Ecological communities produce data with numerous dimensions, including the abundance of many species reported over multiple locations, timeframes, and environmental gradients all at once. Mathematically sound multivariate and computational techniques are necessary for pattern extraction from this type of data and pattern application to the prediction of biodiversity in the face of novel or changing environmental factors. This work offers a unified approach to ecological pattern recognition and biodiversity prediction using equation-based computational learning algorithms and principal multivariate statistical approaches. Together with the diversity indices and model-evaluation statistics, we formalize ordination methods, classification and clustering methods, and modern computational approaches. These methods include k-means, hierarchical clustering, maximum entropy species distribution models, support vector machines, artificial neural networks, and correspondence analysis. Additionally, we formalize methods for ordination, redundancy analysis, non-metric multidimensional scaling, and linear discriminant analysis. The governing equations, optimization criterion, and ecological interpretation of each approach are presented. Our consideration of the methodologies' mathematical and practical limits, as well as potential future areas for advancement, comes to a close.

Keywords: multivariate statistics, ordination, machine learning, species distribution modelling, biodiversity, pattern recognition, eigenanalysis.

1. Introduction

There is an inherent multivariate nature to biodiversity data. Ecological surveys often document a wide range of environmental factors, including temperature, precipitation, soil pH, elevation, and land-use intensity, in addition to the presence or abundance of hundreds to thousands of species (the response variables) across a number of sites. A matrix Y with dimensions n × S can represent the raw community data, whereas a n × p matrix X can represent the environmental data, where S is the number of species and n is the number of locations. Calculating a function f that predicts the presence or absence of a single focal species, the full community matrix Y, or a diversity index computed from Y is the central mathematical problem in computational ecology. It involves finding low-dimensional, interpretable structure within Y and X.

Following is the structure of this document. You will learn the fundamental data structures and notation used throughout in Section 2. In Section 3, the development of classical ordination approaches for unsupervised pattern recognition is detailed. In Section 4, we outline the approaches for ecological community delineation through clustering and categorization. In Section 5, we go into the development of computational and machine-learning methods that anticipate biodiversity. In Section 6, we see diversity indices, which are common variables for responses. Statistical methods for evaluating prediction models are detailed in Section 7. An operational numerical example demonstrates the procedures in Section 8. Section 10 wraps up, while Section 9 delves into the subject of restrictions.

Regardless of the specific computational details, a common thread running through this paper is that all of the methods presented here are, in essence, the answers to specific optimization problems, such as those involving eigenvalues, likelihoods, losses, or constrained entropy. By presenting the methods in this cohesive, equation-first fashion, we hope to persuade readers to view computational ecology as an integrated field of applied mathematics, where the data structure and ecological question dictate the method of choice.

ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =8.152, July-December 2025, Submitted in July 2025

2. Mathematical Preliminaries and Notation

Let Y = [y_{ij}] be the n × S community data matrix, where y_ij is the abundance (or presence/absence) of species j at site i, for i = 1, …, n and j = 1, …, S. Let X = [x_{ik}] be the n × p environmental data matrix, where x_ik is the value of environmental variable k at site i, for k = 1, …, p. We denote the mean of column j of Y by $\dot{y}_j$ and the mean of column k of X by $\bar{x}_k$. The centred matrices are written $\tilde{Y} = Y - 1\dot{y}^T$ and $\tilde{X} = X - 1\bar{x}^T$, where 1 is an n × 1 vector of ones.

Throughout, boldface capital letters denote matrices, boldface lowercase letters denote vectors, and italic lowercase letters denote scalars. The superscript T denotes matrix transpose, and the superscript −1 denotes matrix inverse. The symbol ‖·‖ denotes the Euclidean (L2) norm of a vector, and tr(·) denotes the trace operator (the sum of the diagonal elements of a square matrix).

3. Ordination Methods for Ecological Pattern Recognition

Ordination approaches extract the most important patterns of variation among sites and species by projecting the high-dimensional community matrix Y onto a small set of synthetic axes. Ecologists rely on these artificial axes to depict and understand community structure.

3.1 Principal Component Analysis (PCA)

Using a series of orthogonal linear combinations of the initial variables, principal component analysis attempts to capture the maximum variance in a sequential fashion. The sample's covariance matrix is defined as follows, given the centered data matrix $\tilde{Y}$:

$$C = (1 / (n - 1)) \cdot \tilde{Y}^T \tilde{Y} \quad (1)$$

Principal component analysis solves the eigenvalue problem

$$C\, v_k = \lambda_k v_k, \quad k = 1, \ldots, S \quad (2)$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda\_S \geq 0$ are the eigenvalues of C and v_k is the corresponding eigenvector, normalized so that $\|v_k\| = 1$. The k^{th} principal component score for site i is then

$$z_{ik} = \tilde{y}_i^T v_k \quad (3)$$

where $\tilde{y}_i$ is the i^{th} row of $\tilde{Y}$ written as a column vector. The proportion of total variance explained by the first m components is\

$$R^2_m = \sum_{k=1}^{m} \lambda k \,/\, \sum_{k=1}^{s} \lambda k \quad (4)$$

In ecological practice, sites that are close together in the space spanned by the first two or three components (z_i^1, z_i^2, z_i^3) are interpreted as having similar species composition, and the loadings v_k identify which species drive the separation along each axis.

3.2 Correspondence Analysis (CA) and Canonical Correspondence Analysis (CCA)

Because raw abundance data are often non-negative counts with many zeros, ecologists frequently prefer correspondence analysis, which is based on the chi-square distance rather than the Euclidean distance. Let $y_{++} = \sum_{i=1}^{n} \sum_{J=1}^{S} yij$ be the grand total of the community matrix, and define the row and column masses

$$r_i = \sum_{J=1}^{S} yij \,/\, y_{++} \quad , \quad c_j = \sum_{i=1}^{n} yij \,/\, y_{++} \quad (5)$$

The matrix of standardized residuals is

$$q_{ij} = y_{ij}/y_{++} \; - r_i c_j \;/ \sqrt{ricj} \quad (6)$$

Using a singular value decomposition of Q, where Q = U D V^T, correspondence analysis finds the site scores along axis k by rescaling the kth column of U by the row masses. Canonical correspondence analysis (CCA) is a method where the ordination axes are further limited to being linear combinations of recorded environmental variables X. In a formally speaking, CCA-determined linear combination

$$u = \tilde{X} b \quad (7)$$

maximization of the weighted variance of the matching row scores derived from Q, with ‖u‖ = 1. Simplifying the eigenvalue problem to its generic form yields the vector b.

$$\tilde{X}^T Q\, Q^T \tilde{X} b = \mu\, \tilde{X}^T \tilde{X} b \quad (8)$$

where the eigenvalue linked to the restricted axis is denoted by μ. The main reason CCA is so popular is that it gives a numerical value to the extent to which observed environmental gradients explain the observed community variation.

3.3 Redundancy Analysis (RDA)

The counterpart of CCA in terms of Euclidean distance is redundancy analysis. This model represents the community matrix by means of an environmental matrix that is subject to multivariate linear regression.

$$\tilde{Y} = \tilde{X} B + E \quad (9)$$

this is where the matrix B, which is p × S, contains the predicted regression coefficients using ordinary least squares.

$$\hat{B} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y} \quad (10)$$

and E is the n × S matrix of residuals. RDA then performs PCA on the matrix of fitted values $\hat{Y} = \tilde{X}\,\hat{B}$, so that the resulting axes represent the variation in community composition that is explainable by the environmental variables, ranked by the eigenvalues of the fitted covariance matrix $(1/(n-1))\,\hat{Y}^T \hat{Y}$.

3.4 Non-metric Multidimensional Scaling (NMDS)

Non-metric multidimensional scaling attempts to find a set of points in a two- or three-dimensional space whose distance rank corresponds to the rank of the original dissimilarities d_{ij} (usually the Bray-Curtis dissimilarity), rather than assuming a linear relationship between the ordination distances and the original dissimilarities. The goal of NMDS is to minimize the stress function, where δ_{ij} is the distance in degrees between the ordination points of sites i and j.

$$\text{Stress} = \sqrt{\frac{\sum_{i<j} (\delta_{ij} - f(d_{ij}))^2}{\sum_{i<j} \delta_{ij}^2}} \quad (11)$$

where f is a rank-preserving regression function that is monotonic and fitted between the original dissimilarities and the ordination distances. A standard method for optimizing the point configuration is an iterative gradient-descent approach. Since it does not make as many assumptions about distributions as PCA or RDA, NMDS is more suited for community ecology data that are extremely non-linear or contain a large number of zero values.

4. Classification and Clustering Methods

Sites (or species) are partitioned into discrete groups that correspond to diverse ecological communities or habitat types using clustering and discriminant approaches, while ordination decreases dimensionality.

4.1 K-means Clustering

Given a desired number of clusters K, k-means partitions the n sites into K disjoint sets $C_1, \ldots, C_K$ by minimizing the within-cluster sum of squares

$$J = \sum_{k=1}^{K} \sum_{i \in Ck} \| yi - \mu k \|^2 \quad (12)$$

where $\mu_k = 1/|C_k| \sum_{i \in Ck} yi$ where cluster k's centroid exists. The usual approach iteratively decreases J monotonically until convergence to a local optimum by either (i) assigning each site to its nearest centroid or (ii) recalculating centroids, since the exact reduction of J is NP-hard.

4.2 Hierarchical Clustering

Hierarchical clustering builds a nested series of partitions by successively merging the two closest clusters according to a chosen linkage rule. Under average linkage, the distance between clusters A and B is

$$D(A,B) = \frac{1}{|A||B|} \sum_{i \in A} \sum_{j \in B} d(\mathbf{y}_i, \mathbf{y}_j) \quad (13)$$

where d(·,·) is a chosen dissimilarity measure, most commonly the Bray-Curtis dissimilarity,

$$d_{BC}(\mathbf{y}_i, \mathbf{y}_j) = \frac{\sum_{s=1}^{S} |y_{is} - y_{js}|}{\sum_{s=1}^{S} (y_{is} + y_{js})} \quad (14)$$

The output is a dendrogram, and ecologically meaningful clusters are typically obtained by cutting the dendrogram at a chosen dissimilarity threshold.

4.3 Linear Discriminant Analysis (LDA)

When groups of sites are already known (for example, distinct habitat types), linear discriminant analysis finds the linear combination of variables that best separates the groups. Let μ_g be the mean vector of group g, and define the within-group and between-group scatter matrices

$$S_W = \sum_{g=1}^{G} \sum_{i \in g} (y_i - \mu_g)(y_i - \mu_g)^T \quad (15)$$

$$S_B = \sum_{g=1}^{G} n_g (\mu_g - \bar{y})(\mu_g - \bar{y})^T \quad (16)$$

where n_g is the number of sites in group g and $\bar{\mu}$ is the overall mean. LDA finds the projection vector w that maximizes Fisher's criterion

$$J(w) = (w^T S_B w) / (w^T S_W w) \quad (17)$$

The solution is the eigenvector corresponding to the largest eigenvalue of the generalized eigenvalue problem $S_B w = \lambda S_W w$. LDA is used both to classify new sites into predefined community types and to identify which environmental or species variables contribute most to the separation between them.

5. Computational and Machine-Learning Methods for Biodiversity Prediction

Whereas classical multivariate statistics assume linear relationships and known distributional forms, computational learning algorithms make weaker assumptions and can capture non-linear, interactive relationships between environmental predictors and biodiversity outcomes. These methods are now the dominant tools for species distribution modelling and large-scale biodiversity prediction.

5.1 Logistic Regression for Species Distribution Modelling

The simplest predictive model for a single species' presence ($y_i = 1$) or absence ($y_i = 0$) at site i is logistic regression, which models the probability of presence as

$$P(y_i = 1 \mid x_i) = 1 / (1 + exp(-(\beta_0 + \beta^T x_i))) \quad (18)$$

The coefficients β are estimated by maximizing the log-likelihood

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \ln P(y_i = 1 \mid \mathbf{x}_i) + (1 - y_i) \ln \left(1 - P(y_i = 1 \mid \mathbf{x}_i)\right) \right] \quad (19)$$

using iteratively reweighted least squares or gradient-based optimization.

5.2 Maximum Entropy Modelling (MaxEnt)

MaxEnt is one of the most widely used species distribution models in ecology, particularly for presence-only data. It estimates the probability distribution π(x) over environmental space that has maximum entropy

$$H(\pi) = -\sum_x \pi(x) \log \pi(x) \quad (20)$$

subject to the constraint that the expected value of each environmental feature f_k under π matches its empirical average at the observed presence locations,

$$\mathbb{E}_\pi[f_k(\mathbf{x})] = \frac{1}{n} \sum_{i=1}^{n} f_k(\mathbf{x}_i), k = 1, \ldots, p \quad (21)$$

The solution to this constrained optimization has the exponential (Gibbs) form

$$\pi(x) = 1 / Z(\lambda) \, exp(\sum_{k=1}^{p} \lambda k f k(x)) \quad (22)$$

where $Z(\lambda)$ is a normalizing constant and the λ_k are Lagrange multipliers estimated numerically. The resulting $\pi(x)$ is interpreted as a habitat-suitability surface across geographic or environmental space.

5.3 Random Forests

A random forest is an ensemble of B decision trees, each trained on a bootstrap sample of the data using a random subset of predictor variables at each split. Each tree T_b partitions the predictor space into regions and predicts a constant value within each region; the split at each node is chosen to minimize impurity, commonly measured by the Gini index for classification,

$$G(t) - 1 - \sum_{c=1}^{C} pc(t)^2) \quad (23)$$

where $p_c(t)$ is the proportion of class c observations in node t, or by the variance reduction criterion for regression of a continuous response such as species richness. The forest's final prediction is the average (for regression) or majority vote (for classification) across all B trees,

$$\hat{f}(x) = 1/B \sum_{b=1}^{B} Tb(x) \quad (24)$$

Random forests are particularly well suited to biodiversity prediction because they capture non-linear responses and interactions among environmental variables without requiring the analyst to specify a functional form in advance, and they provide a natural measure of variable importance based on the average decrease in impurity attributable to each predictor.

5.4 Gradient Boosted Trees

A related ensemble method, gradient boosting, builds an additive model of M trees sequentially, where each new tree is fitted to the negative gradient (the residual error) of the loss function evaluated at the current model. The model after m stages is

$$f_m(x) = f_{m-1}(x) + \eta \cdot T_m(x) \quad (24.1)$$

where T_m is a shallow regression tree fitted to the pseudo-residuals $r_i = -\partial L(y_i, f_{m-1}(x_i)) / \partial f_{m-1} x_i$, and η is a shrinkage (learning-rate) parameter that controls the contribution of each successive tree. Gradient boosting frequently achieves higher predictive accuracy than a single random forest on structured biodiversity datasets, at the cost of greater sensitivity to the choice of η, tree depth, and the number of boosting stages M, all of which must be tuned by cross-validation (Section 7) to avoid overfitting.

5.5 Artificial Neural Networks

A feed-forward neural network with one hidden layer of H units computes its prediction as

$$h_m = \sigma\left(\sum_{k=1}^{p} w_{mk}^{(1)} x_k + b_m^{(1)}\right), m = 1, \ldots, H \quad (25)$$

$$\hat{y} = \sigma\left(\sum_{m=1}^{H} w_m^{(2)} h_m + b^{(2)}\right) \quad (26)$$

where σ is a non-linear activation function (commonly the logistic sigmoid or the rectified linear unit, $\mathrm{ReLU}(u) = \max(0, u)$), and the weight matrices $W^{(1)}$, $W^{(2)}$ and bias vectors $b^{(1)}$, $b^{(2)}$ are estimated by minimizing a loss function such as the mean squared error

$$L(W) = 1/n \sum_{i=1}^{n} (yi - \hat{y}i)^2 \quad (27)$$

via stochastic gradient descent, in which the weights are updated iteratively as

$$W \leftarrow W - \eta \nabla_W L(W) \quad (28)$$

where η is the learning rate. Deep networks, with multiple hidden layers, are increasingly used to predict biodiversity indices directly from remote-sensing imagery or acoustic recordings, in which case the input x_i is itself a high-dimensional array (an image or spectrogram) processed by convolutional layers before being passed to the fully connected layers described above.

5.6 Support Vector Machines (SVM)

For binary classification of species presence/absence, a support vector machine seeks the separating hyperplane $w^T x + b = 0$ that maximizes the margin between classes, which is obtained by solving the constrained optimization problem

$$\min_{w,b} \tfrac{1}{2} \|w\|^2 \quad \text{subject to} \quad y_i (w^T x_i + b) \geq 1 - \xi_i, \ \xi_i \geq 0 \quad (29)$$

where the slack variables ξ_i allow for imperfectly separable data and are penalized in the objective by an additional term $C \Sigma_i \xi_i$, with C a user-chosen regularization constant. When the relationship between predictors and species occurrence is non-linear, the inner product $x_i^T x_j$ is replaced by a kernel function $K(x_i, x_j)$, most commonly the Gaussian radial basis function

$$K(x_i, x_j) = \exp(- \|x_i - x_j\|^2 / (2\sigma^2)) \quad (30)$$

allowing the SVM to fit complex, non-linear decision boundaries in the original predictor space.

6. Diversity Indices as Response Variables

Many prediction problems in computational ecology take a scalar diversity index, rather than the full community matrix, as the response variable to be predicted from environmental data. The most widely used indices are defined below in terms of the proportional abundance $p_j = y_{ij} / \sum_{j=1}^{S} yij$ of species j at a given site.

Species richness is simply the count of species present,

$$R = \sum_{j=1}^{S} 1((yij > 0) \quad (31)$$

where $\mathbb{1}(\cdot)$ is the indicator function. The Shannon diversity index is

$$H' = 1 - \sum_{j=1}^{S} pj \log(pj) \quad (32)$$

and the Simpson diversity index, which places greater weight on abundant species, is

$$D = 1 - \sum_{j=1}^{S} pj^2 \quad (33)$$

Pielou's evenness index rescales Shannon diversity relative to its maximum possible value,

$$J = H' / \log(R) \quad (34)$$

These indices are frequently regressed against, or predicted from, environmental covariates using any of the models described in Sections 4 and 5, treating the index as a univariate continuous response.

7. Model Evaluation and Validation Statistics

Regardless of which method is used, predictive models must be validated on data not used in model fitting, typically via k-fold cross-validation, in which the data are partitioned into k subsets, the model is trained on $k - 1$ subsets and evaluated on the remaining subset, and the process is repeated k times.

For continuous predictions (e.g., species richness or a diversity index), the coefficient of determination is used,

$$R^2 = 1 - \sum_{i=1}^{n}(yi - \hat{y}i)^2 / \sum_{i=1}^{n}(yi - \bar{y})^2 \quad (35)$$

and the root-mean-square error,

$$RMSE = \sqrt{[(1/n) \sum_{i=1}^{n}(yi - \hat{y}i)^2} \quad (36)$$

For binary classification of species presence/absence, the model's discrimination ability is summarized by the area under the receiver operating characteristic curve (AUC), which can be interpreted as the probability that a randomly chosen presence site is ranked higher by the model than a randomly chosen absence site,

$$AUC = P(\hat{y}(x_+) > \hat{y}(x_-)) \quad (37)$$

and by Cohen's kappa statistic, which corrects the observed classification accuracy p_o for the accuracy p_e expected by chance,

$$\kappa = (p_o - p_e) / (1 - p_e) \quad (38)$$

8. Numerical Example

Here we see the methods described in Sections 3–5 in action by taking a look at a survey of forty forest plots. In these plots, we record the abundance of twenty-five different tree species along with six environmental variables: elevation, mean yearly rainfall, mean annual temperature, soil pH, canopy openness, and distance to the nearest river. The following is the standard flow of an analysis pipeline. To begin, the 40 × 25 community matrix Y is subjected to principal component analysis (PCA) or non-dimensional feature

selection (NMDS) in order to get a two-dimensional ordination. This ordination enables the visual diagnosis of potential community kinds. Secondly, the ordination scores are used in conjunction with hierarchical clustering or k-means to formally divide the 40 plots into, example, K = 4 community kinds. Alternatively, the Bray-Curtis dissimilarity matrix can be used directly. Additionally, the third step is to determine which of the six environmental variables has the greatest impact on community composition by utilizing CCA or RDA. This will help quantify the amount of variance among the 25 species' abundances (equation 9) that can be explained jointly by these variables. The fourth step is to use the six environmental factors to forecast the presence or abundance of one or more particular species of conservation concern at new, unsurveyed locations across the landscape. This is done by fitting a random forest or MaxEnt model (equations 22 to 24). Fifth, a predicted diversity surface for the whole study region is produced by regressing the environmental factors (equations 25–27) against the Shannon diversity index (equation 32) calculated at each of the forty plots. Lastly, k-fold cross-validation and the R^2, RMSE, AUC, and κ statistics from Section 7 are used to evaluate the predictive accuracy of every fitted model. Only models that demonstrate satisfactory performance on held-out data are then utilized to create maps or forecasts of landscape-scale biodiversity.

Regardless of the taxonomic group, ecosystem, or specific software implementation used, this pipeline —ordination, clustering, constrained ordination, predictive modeling, and cross-validated evaluation— is representative of the general structure followed in most published applications of computational multivariate methods to biodiversity data.

9. Mathematical and Practical Limitations

The approaches presented above have a number of drawbacks that should be highlighted. First, when species show unimodal (bell-shaped) responses to environmental gradients, linear ordination methods (PCA, RDA) assume that they respond monotonically to underlying gradients. This assumption is often violated, which is why unimodal-based methods (CCA, correspondence analysis) are used instead.

The second point is that when the ratio of species S to sites n is large, the eigenvalue decompositions of the relevant scatter or covariance matrices become singular or near-singular and regularization is needed. One example of this is the ridge-penalized variant that adds a small constant to the diagonal of S_W before inversion. This is in contrast to the eigenvalue decompositions used in PCA, CA, CCA, and LDA (equations 2, 8, and 17), which assume that the relevant matrices are of full rank and well-conditioned.

Third, as emphasised in Section 7, machine-learning methods like neural networks and random forests (equations 23–28) often necessitate bigger sample sizes than traditional multivariate methods to prevent overfitting. Additionally, their adaptability can result in misleading patterns that do not have any bearing on ecology.

In terms of the fourth point, it is well-known that presence-only methods like MaxEnt (equations 20–22) are susceptible to sampling bias in the areas where species presences were recorded. For example, if the survey effort is focused around roads or research stations, the fitted distribution $\pi(x)$ will reflect this bias instead of the species' actual ecological niche, unless explicit bias-correction terms are utilized in the model.

Lastly, it is important to note that while utilizing these models to inform conservation policy or land-management decisions, they only estimate statistical association between environmental predictors and biodiversity outcomes, not causal mechanism. Having a high predictive accuracy, such as a high R^2 or AUC, does not prove that a specific environmental variable causes species occurrence or diversity.

10. Conclusion

A comprehensive, equation-based overview of the main multivariate and computational methods for ecological pattern recognition and biodiversity prediction has been provided in

this paper. These methods include classical linear ordination (PCA, CA, CCA, RDA), non-linear ordination (NMDS), classification and clustering (k-means, hierarchical clustering, LDA), and modern computational learning methods (logistic regression, MaxEnt, random forests, neural networks, support vector machines). The paper has also defined and assessed predictive performance using diversity indices and validation statistics.

These methods are all mathematically related to one another and all aim to optimize some well-defined objective function, such as variance explained, within-cluster sum of squares, between-to-within group variance ratio, likelihood, random forests, neural networks, or prediction error. The goal is to reduce a high-dimensional community or predictor matrix to a smaller set of interpretable quantities, such as eigenvectors, cluster centroids, discriminant axes, or fitted functions. Predicting biodiversity in a dynamic world using computational multivariate methods requires a solid grasp of these underlying equations, not just treating each method as a computational "black box." Only then can we apply these methods in an ecologically informed and statistically valid manner.

References

1. Anderson, M. J. (2001). *A new method for non-parametric multivariate analysis of variance*. *Austral Ecology, 26*(1), 32–46.
2. Borcard, D., Gillet, F., & Legendre, P. (2018). *Numerical ecology with R* (2nd ed.). Springer.
3. Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs, 27*(4), 325–349.
4. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
5. Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology, 18*(1), 117–143.
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.
7. Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics, 40*, 677–697.
8. Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions, 17*(1), 43–57.
9. Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179–188.
10. Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
11. Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(6), 417–441.
12. Legendre, P., & Legendre, L. (2012). *Numerical ecology* (3rd English ed.). Elsevier.
13. McCune, B., & Grace, J. B. (2002). *Analysis of ecological communities*. MjM Software Design.
14. Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2*(11), 559–572.
15. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.
16. Simpson, E. H. (1949). Measurement of diversity. *Nature, 163*(4148), 688.
17. Ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology, 67*(5), 1167–1179.
18. Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika, 17*(4), 401–419.
19. Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.
20. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.