Developing an Ecological Niche-Based Model for Assessing the Invasion Risk of Anopheles stephensi

Koissi Savi (Ph.D.)

Harvard School of Medicine

January 16, 2024

1. Species Distribution Models (SDMs)
   - Concepts
   - Best practice standard for SDM from Araujo et al. 2019
2. Classification Tree Analysis (CTA) for SDM
3. Ensemble model
4. Envelope Models
5. MaxEnt

Concepts

- **Species Distribution models also known as Ecological Niche Modeling** is a tool to study mosquito geographic distributions. This type of model relates presence-absence or presence-only occurrence data to explanatory landscape factors, producing estimates of suitable habitat

Concepts

- **Species Distribution models also known as Ecological Niche Modeling** is a tool to study mosquito geographic distributions. This type of model relates presence-absence or presence-only occurrence data to explanatory landscape factors, producing estimates of suitable habitat
- SDMs typically include geolocated data on the presence of the species of interest as the response variable, often in the form of occurrence records derived from literature reviews, databases, or aggregated abundance sampling

Concepts

- **Species Distribution models also known as Ecological Niche Modeling** is a tool to study mosquito geographic distributions. This type of model relates presence-absence or presence-only occurrence data to explanatory landscape factors, producing estimates of suitable habitat
- SDMs typically include geolocated data on the presence of the species of interest as the response variable, often in the form of occurrence records derived from literature reviews, databases, or aggregated abundance sampling
- SDMs are applied to determine which predictors limit a species distribution, often referred to as variable importance (Bardie and Leung, 2017)

Concepts

- **Species Distribution models also known as Ecological Niche Modeling** is a tool to study mosquito geographic distributions. This type of model relates presence-absence or presence-only occurrence data to explanatory landscape factors, producing estimates of suitable habitat
- SDMs typically include geolocated data on the presence of the species of interest as the response variable, often in the form of occurrence records derived from literature reviews, databases, or aggregated abundance sampling
- SDMs are applied to determine which predictors limit a species distribution, often referred to as variable importance (Bardie and Leung, 2017)

Type of SDM from Lippi et al. 2023

From: Trends in mosquito species distribution modeling: insights for vector surveillance and disease control

Modeling method	Definition (example)
CTA	Classification tree analysis methods, including classification and regression trees, boosted regression trees, and random forest
Ensemble	A weighted or unweighted average, or combination, of models built with different methods
Envelope	Models that identify the boundaries of species' ecological tolerance directly from data, without the use of machine learning (e.g., BIOCLIM, CLIMEX, CliMond, DOMAIN)
GARP	Genetic algorithm for rule-set production, generates mathematical rules for estimating species presence
MaxEnt	Maximum entropy, for expressing probability distributions
Mechanistic	Process-based models, often using parameters of physiological limits to estimate distributions (e.g., species thermal limits)
Mixed	Uses two or more methods to estimate species distributions, but does not average or combine output into a model ensemble
Regression	Non-machine learning regression models (e.g., logistic regression, generalized linear models, generalized additive models, etc.)
Other	Less commonly used methods that did not fit into another category (e.g., ecological niche factor analysis, environmental suitability thresholds, logic thresholds)

Potential purposes of SDM from from Zurell et al. 2020

- Araújo et al. (2019) suggested three distinct SDM purposes
    1. **Explanation**: (also termed inference) regards detailed analyses of speciesenvironment relationships and aims to provide or test specific hypotheses about the main factors driving the species distributions.

Potential purposes of SDM from from Zurell et al. 2020

- Araújo et al. (2019) suggested three distinct SDM purposes
  1. **Explanation**: (also termed inference) regards detailed analyses of speciesenvironment relationships and aims to provide or test specific hypotheses about the main factors driving the species distributions.
  2. **Mapping**: (also termed interpolation) means that the estimated speciesenvironment relationships are used to map (or interpolate) the species distributions in the same geographic area and time period in which the model was calibrated.

Potential purposes of SDM from from Zurell et al. 2020

- Araújo et al. (2019) suggested three distinct SDM purposes
  1. **Explanation**: (also termed inference) regards detailed analyses of speciesenvironment relationships and aims to provide or test specific hypotheses about the main factors driving the species distributions.
  2. **Mapping**: (also termed interpolation) means that the estimated speciesenvironment relationships are used to map (or interpolate) the species distributions in the same geographic area and time period in which the model was calibrated.
  3. **Transfer**: (also termed forecast or projection; but these terms are less precise) means that the estimated speciesenvironment relationships are transferred to a different geographic region or time period future or past (Yates et al. 2018).

Potential purposes of SDM from from Zurell et al. 2020

- Araújo et al. (2019) suggested three distinct SDM purposes
  1. **Explanation**: (also termed inference) regards detailed analyses of speciesenvironment relationships and aims to provide or test specific hypotheses about the main factors driving the species distributions.
  2. **Mapping**: (also termed interpolation) means that the estimated speciesenvironment relationships are used to map (or interpolate) the species distributions in the same geographic area and time period in which the model was calibrated.
  3. **Transfer**: (also termed forecast or projection; but these terms are less precise) means that the estimated speciesenvironment relationships are transferred to a different geographic region or time period future or past (Yates et al. 2018).

1. Species Distribution Models (SDMs)
   - Concepts
   - Best practice standard for SDM from Araujo et al. 2019
2. Classification Tree Analysis (CTA) for SDM
3. Ensemble model
4. Envelope Models
5. MaxEnt

Standard of SDM models

- **Gold standard**: Usually requires ideal data (seldom available) and next-generation modeling approaches that remain under development, as well as results obtained through multiple sources of evidence including manipulative experiments. Therefore, this level typically provides targets for excellence and directions for future research.

Standard of SDM models

- **Gold standard**: Usually requires ideal data (seldom available) and next-generation modeling approaches that remain under development, as well as results obtained through multiple sources of evidence including manipulative experiments. Therefore, this level typically provides targets for excellence and directions for future research.
- **(Silver standard**: Typically involving imperfect (but best available) data combined with analyses that allow uncertainty and bias to be reduced, accounted for, or at least estimated.

Standard of SDM models

- **Gold standard**: Usually requires ideal data (seldom available) and next-generation modeling approaches that remain under development, as well as results obtained through multiple sources of evidence including manipulative experiments. Therefore, this level typically provides targets for excellence and directions for future research.
- **(Silver standard**: Typically involving imperfect (but best available) data combined with analyses that allow uncertainty and bias to be reduced, accounted for, or at least estimated.
- **Bronze standard**: Encompasses data and procedures that represent the minimum currently acceptable practices for models to be included in SDM. It includes approaches to characterize and address limitations of data and models, and to interpret their implications on the results.

Standard of SDM models

- **Gold standard**: Usually requires ideal data (seldom available) and next-generation modeling approaches that remain under development, as well as results obtained through multiple sources of evidence including manipulative experiments. Therefore, this level typically provides targets for excellence and directions for future research.
- **(Silver standard**: Typically involving imperfect (but best available) data combined with analyses that allow uncertainty and bias to be reduced, accounted for, or at least estimated.
- **Bronze standard**: Encompasses data and procedures that represent the minimum currently acceptable practices for models to be included in SDM. It includes approaches to characterize and address limitations of data and models, and to interpret their implications on the results.
- **Deficient**: Involves the use of data and/or modeling practices that are considered unacceptable for models used in driving policy and practice.

Standard of SDM models

- **Gold standard**: Usually requires ideal data (seldom available) and next-generation modeling approaches that remain under development, as well as results obtained through multiple sources of evidence including manipulative experiments. Therefore, this level typically provides targets for excellence and directions for future research.
- **(Silver standard**: Typically involving imperfect (but best available) data combined with analyses that allow uncertainty and bias to be reduced, accounted for, or at least estimated.
- **Bronze standard**: Encompasses data and procedures that represent the minimum currently acceptable practices for models to be included in SDM. It includes approaches to characterize and address limitations of data and models, and to interpret their implications on the results.
- **Deficient**: Involves the use of data and/or modeling practices that are considered unacceptable for models used in driving policy and practice.

Key Criteria affecting the quality of the SDM

- Quality of the response variable (usually species occurrence data)

Key Criteria affecting the quality of the SDM

- Quality of the response variable (usually species occurrence data)
- Quality of the predictor variables (usually environmental data)

Key Criteria affecting the quality of the SDM

- Quality of the response variable (usually species occurrence data)
- Quality of the predictor variables (usually environmental data)
- "Model building i.e., fitting a statistical relationship be-tween species occurrence data and environmental data.

Key Criteria affecting the quality of the SDM

- Quality of the response variable (usually species occurrence data)
- Quality of the predictor variables (usually environmental data)
- "Model building i.e., fitting a statistical relationship be-tween species occurrence data and environmental data.
- "Model evaluation aligns a balance blend of realism, accuracy, and generality which can be assessed through these questions:

Key Criteria affecting the quality of the SDM

- Quality of the response variable (usually species occurrence data)
- Quality of the predictor variables (usually environmental data)
- "Model building i.e., fitting a statistical relationship be-tween species occurrence data and environmental data.
- "Model evaluation aligns a balance blend of realism, accuracy, and generality which can be assessed through these questions:
  1. How robust is the model to departures from the assumptions?

Key Criteria affecting the quality of the SDM

- Quality of the response variable (usually species occurrence data)
- Quality of the predictor variables (usually environmental data)
- "Model building i.e., fitting a statistical relationship be-tween species occurrence data and environmental data.
- "Model evaluation aligns a balance blend of realism, accuracy, and generality which can be assessed through these questions:
  1. How robust is the model to departures from the assumptions?
  2. How meaningful are the evaluation metrics used?

Key Criteria affecting the quality of the SDM

- Quality of the response variable (usually species occurrence data)
- Quality of the predictor variables (usually environmental data)
- "Model building i.e., fitting a statistical relationship be-tween species occurrence data and environmental data.
- "Model evaluation aligns a balance blend of realism, accuracy, and generality which can be assessed

  through these questions:

  1. How robust is the model to departures from the assumptions?
  2. How meaningful are the evaluation metrics used?
  3. How predictive is the model when tested against independent data?

Key Criteria affecting the quality of the SDM

- Quality of the response variable (usually species occurrence data)
- Quality of the predictor variables (usually environmental data)
- "Model building i.e., fitting a statistical relationship be-tween species occurrence data and environmental data.
- "Model evaluation aligns a balance blend of realism, accuracy, and generality which can be assessed

  through these questions:

  1. How robust is the model to departures from the assumptions?
  2. How meaningful are the evaluation metrics used?
  3. How predictive is the model when tested against independent data?

1. Species Distribution Models (SDMs)
2. Classification Tree Analysis (CTA) for SDM
   - Model building and evaluation
3. Ensemble model
4. Envelope Models
5. MaxEnt

Description

- Classification Tree Analysis (CTA), also known as Decision Tree Analysis, is a machine learning algorithm used for classification and regression tasks. In the context of SDM, CTA is applied to predict the presence or absence of a species based on environmental predictor variables. The algorithm recursively splits the dataset into subsets based on the predictor variables, creating a tree structure where each node represents a decision based on a specific variable. Here are some of the advantages of CTA:

Description

- Classification Tree Analysis (CTA), also known as Decision Tree Analysis, is a machine learning algorithm used for classification and regression tasks. In the context of SDM, CTA is applied to predict the presence or absence of a species based on environmental predictor variables. The algorithm recursively splits the dataset into subsets based on the predictor variables, creating a tree structure where each node represents a decision based on a specific variable. Here are some of the advantages of CTA:
- **Interpretability**: Decision trees are easy to interpret and visualize, making it simpler for researchers and stakeholders to understand the model's decision-making process.

Description

- Classification Tree Analysis (CTA), also known as Decision Tree Analysis, is a machine learning algorithm used for classification and regression tasks. In the context of SDM, CTA is applied to predict the presence or absence of a species based on environmental predictor variables. The algorithm recursively splits the dataset into subsets based on the predictor variables, creating a tree structure where each node represents a decision based on a specific variable. Here are some of the advantages of CTA:
- **Interpretability**: Decision trees are easy to interpret and visualize, making it simpler for researchers and stakeholders to understand the model's decision-making process.
- **Non-linearity Handling**: CTA can capture non-linear relationships between environmental variables and species occurrence, which might be missed by linear models.

Description

- Classification Tree Analysis (CTA), also known as Decision Tree Analysis, is a machine learning algorithm used for classification and regression tasks. In the context of SDM, CTA is applied to predict the presence or absence of a species based on environmental predictor variables. The algorithm recursively splits the dataset into subsets based on the predictor variables, creating a tree structure where each node represents a decision based on a specific variable. Here are some of the advantages of CTA:
- **Interpretability**: Decision trees are easy to interpret and visualize, making it simpler for researchers and stakeholders to understand the model's decision-making process.
- **Non-linearity Handling**: CTA can capture non-linear relationships between environmental variables and species occurrence, which might be missed by linear models.
- **No Assumption of Linearity**: Unlike some parametric models, CTA does not assume a linear relationship between predictors and the response, making it more flexible.

Description

- Classification Tree Analysis (CTA), also known as Decision Tree Analysis, is a machine learning algorithm used for classification and regression tasks. In the context of SDM, CTA is applied to predict the presence or absence of a species based on environmental predictor variables. The algorithm recursively splits the dataset into subsets based on the predictor variables, creating a tree structure where each node represents a decision based on a specific variable. Here are some of the advantages of CTA:
- **Interpretability**: Decision trees are easy to interpret and visualize, making it simpler for researchers and stakeholders to understand the model's decision-making process.
- **Non-linearity Handling**: CTA can capture non-linear relationships between environmental variables and species occurrence, which might be missed by linear models.
- **No Assumption of Linearity**: Unlike some parametric models, CTA does not assume a linear relationship between predictors and the response, making it more flexible.
- **Variable Importance**: CTA provides a measure of variable importance, helping identify which environmental factors are most influential in predicting species distribution.

Use cases and weaknesses of CTA

- Habitat Suitability Modeling; Predicting Range Shifts; Understanding the ecological requirements and preferences of a species.

Use cases and weaknesses of CTA

- Habitat Suitability Modeling; Predicting Range Shifts; Understanding the ecological requirements and preferences of a species.
- Decision trees can be prone to overfitting, capturing noise in the training data and leading to poor generalization on new data.

Use cases and weaknesses of CTA

- Habitat Suitability Modeling; Predicting Range Shifts; Understanding the ecological requirements and preferences of a species.
- Decision trees can be prone to overfitting, capturing noise in the training data and leading to poor generalization on new data.
- Small changes in the data can result in different tree structures, impacting the stability of the model.

Use cases and weaknesses of CTA

- Habitat Suitability Modeling; Predicting Range Shifts; Understanding the ecological requirements and preferences of a species.
- Decision trees can be prone to overfitting, capturing noise in the training data and leading to poor generalization on new data.
- Small changes in the data can result in different tree structures, impacting the stability of the model.
- CTA can be sensitive to outliers, potentially leading to biased predictions.

Use cases and weaknesses of CTA

- Habitat Suitability Modeling; Predicting Range Shifts; Understanding the ecological requirements and preferences of a species.
- Decision trees can be prone to overfitting, capturing noise in the training data and leading to poor generalization on new data.
- Small changes in the data can result in different tree structures, impacting the stability of the model.
- CTA can be sensitive to outliers, potentially leading to biased predictions.

CTA model building and evaluation

- Application starting from line 144 here
- **Accuracy**: Accuracy is the ratio of correctly predicted instances to the total instances. It measures the overall correctness of the model.

CTA model building and evaluation

- Application starting from line 144 here
- **Accuracy**: Accuracy is the ratio of correctly predicted instances to the total instances. It measures the overall correctness of the model.

$$Accuracy = \frac{TruePositives + TrueNegative}{Total} \tag{1}$$

- **Sensitivity**: Sensitivity is the ratio of correctly predicted positive observations to the total actual positives. It measures the ability of the model to capture all positive instances.

CTA model building and evaluation

- Application starting from line 144 here
- **Accuracy**: Accuracy is the ratio of correctly predicted instances to the total instances. It measures the overall correctness of the model.

$$Accuracy = \frac{TruePositives + TrueNegative}{Total} \tag{1}$$

- **Sensitivity**: Sensitivity is the ratio of correctly predicted positive observations to the total actual positives. It measures the ability of the model to capture all positive instances.

$$Sensitivity = \frac{TruePositives}{ActualPositive} \tag{2}$$

- **Specificity**: Specificity is the ratio of correctly predicted negative observations to the total actual negatives. It measures the ability of the model to avoid false positives.

CTA model building and evaluation

- Application starting from line 144 here
- **Accuracy**: Accuracy is the ratio of correctly predicted instances to the total instances. It measures the overall correctness of the model.

$$Accuracy = \frac{TruePositives + TrueNegative}{Total} \tag{1}$$

- **Sensitivity**: Sensitivity is the ratio of correctly predicted positive observations to the total actual positives. It measures the ability of the model to capture all positive instances.

$$Sensitivity = \frac{TruePositives}{ActualPositive} \tag{2}$$

- **Specificity**: Specificity is the ratio of correctly predicted negative observations to the total actual negatives. It measures the ability of the model to avoid false positives.

$$Specificity = \frac{TrueNegatives}{ActualNegative} \tag{3}$$

- **Precision**: Precision is the ratio of correctly predicted positive observations to the total predicted positives. It measures the accuracy of positive predictions made by the model.

CTA model building and evaluation

- Application starting from line 144 here
- **Accuracy**: Accuracy is the ratio of correctly predicted instances to the total instances. It measures the overall correctness of the model.

$$Accuracy = \frac{TruePositives + TrueNegative}{Total} \tag{1}$$

- **Sensitivity**: Sensitivity is the ratio of correctly predicted positive observations to the total actual positives. It measures the ability of the model to capture all positive instances.

$$Sensitivity = \frac{TruePositives}{ActualPositive} \tag{2}$$

- **Specificity**: Specificity is the ratio of correctly predicted negative observations to the total actual negatives. It measures the ability of the model to avoid false positives.

$$Specificity = \frac{TrueNegatives}{ActualNegative} \tag{3}$$

- **Precision**: Precision is the ratio of correctly predicted positive observations to the total predicted positives. It measures the accuracy of positive predictions made by the model.

$$Precision = \frac{TruePositive}{PredictedPositive} \tag{4}$$

1. Species Distribution Models (SDMs)
2. Classification Tree Analysis (CTA) for SDM
3. Ensemble model
   - Concept
   - Application
4. Envelope Models
5. MaxEnt

Concept

- In the context of Species Distribution Modeling (SDM), an ensemble model refers to a modeling approach where multiple individual models are combined to produce a more robust and accurate prediction.

Concept

- In the context of Species Distribution Modeling (SDM), an ensemble model refers to a modeling approach where multiple individual models are combined to produce a more robust and accurate prediction.
- The idea is that by aggregating predictions from multiple models, the weaknesses of individual models can be mitigated, and the overall performance can be improved. Common techniques for creating ensemble models include averaging, stacking, and boosting.

Concept

- In the context of Species Distribution Modeling (SDM), an ensemble model refers to a modeling approach where multiple individual models are combined to produce a more robust and accurate prediction.
- The idea is that by aggregating predictions from multiple models, the weaknesses of individual models can be mitigated, and the overall performance can be improved. Common techniques for creating ensemble models include averaging, stacking, and boosting.
- Ensemble models often outperform individual models by reducing overfitting and capturing a broader range of patterns in the data.

Concept

- In the context of Species Distribution Modeling (SDM), an ensemble model refers to a modeling approach where multiple individual models are combined to produce a more robust and accurate prediction.
- The idea is that by aggregating predictions from multiple models, the weaknesses of individual models can be mitigated, and the overall performance can be improved. Common techniques for creating ensemble models include averaging, stacking, and boosting.
- Ensemble models often outperform individual models by reducing overfitting and capturing a broader range of patterns in the data.
- Ensembles are more robust to noise and variations in the data, as they aggregate information from multiple sources.

Concept

- In the context of Species Distribution Modeling (SDM), an ensemble model refers to a modeling approach where multiple individual models are combined to produce a more robust and accurate prediction.
- The idea is that by aggregating predictions from multiple models, the weaknesses of individual models can be mitigated, and the overall performance can be improved. Common techniques for creating ensemble models include averaging, stacking, and boosting.
- Ensemble models often outperform individual models by reducing overfitting and capturing a broader range of patterns in the data.
- Ensembles are more robust to noise and variations in the data, as they aggregate information from multiple sources.
- Ensemble models can be more complex and computationally intensive than individual models, especially when using a large number of base models.

Concept

- In the context of Species Distribution Modeling (SDM), an ensemble model refers to a modeling approach where multiple individual models are combined to produce a more robust and accurate prediction.
- The idea is that by aggregating predictions from multiple models, the weaknesses of individual models can be mitigated, and the overall performance can be improved. Common techniques for creating ensemble models include averaging, stacking, and boosting.
- Ensemble models often outperform individual models by reducing overfitting and capturing a broader range of patterns in the data.
- Ensembles are more robust to noise and variations in the data, as they aggregate information from multiple sources.
- Ensemble models can be more complex and computationally intensive than individual models, especially when using a large number of base models.
- The interpretability of ensemble models may be lower compared to individual models, as it involves combining predictions from multiple sources.

Concept

- In the context of Species Distribution Modeling (SDM), an ensemble model refers to a modeling approach where multiple individual models are combined to produce a more robust and accurate prediction.
- The idea is that by aggregating predictions from multiple models, the weaknesses of individual models can be mitigated, and the overall performance can be improved. Common techniques for creating ensemble models include averaging, stacking, and boosting.
- Ensemble models often outperform individual models by reducing overfitting and capturing a broader range of patterns in the data.
- Ensembles are more robust to noise and variations in the data, as they aggregate information from multiple sources.
- Ensemble models can be more complex and computationally intensive than individual models, especially when using a large number of base models.
- The interpretability of ensemble models may be lower compared to individual models, as it involves combining predictions from multiple sources.

Application

- Ensemble models are particularly effective when dealing with diverse and heterogeneous datasets.

Application

- Ensemble models are particularly effective when dealing with diverse and heterogeneous datasets.
- If there is uncertainty about which individual model is the best, ensemble models can help capture different perspectives.

Application

- Ensemble models are particularly effective when dealing with diverse and heterogeneous datasets.
- If there is uncertainty about which individual model is the best, ensemble models can help capture different perspectives.
- Ensembles can be useful in situations where overfitting is a concern with individual models.

Application

- Ensemble models are particularly effective when dealing with diverse and heterogeneous datasets.
- If there is uncertainty about which individual model is the best, ensemble models can help capture different perspectives.
- Ensembles can be useful in situations where overfitting is a concern with individual models.

Concept

- An envelope model is a type of ecological niche model that identifies the environmental conditions within which a species is likely to occur. It essentially delineates the environmental boundaries or envelope that encapsulates the known occurrences of a species. This model could be applied in the context of Anopheles stephensi to decipher ecological conditions that favor the species' growth.

Concept

- An envelope model is a type of ecological niche model that identifies the environmental conditions within which a species is likely to occur. It essentially delineates the environmental boundaries or envelope that encapsulates the known occurrences of a species. This model could be applied in the context of Anopheles stephensi to decipher ecological conditions that favor the species' growth.
- Envelope models are often simpler and computationally less demanding than some machine learning-based models.

Concept

- An envelope model is a type of ecological niche model that identifies the environmental conditions within which a species is likely to occur. It essentially delineates the environmental boundaries or envelope that encapsulates the known occurrences of a species. This model could be applied in the context of Anopheles stephensi to decipher ecological conditions that favor the species' growth.
- Envelope models are often simpler and computationally less demanding than some machine learning-based models.
- The results are often easy to interpret, making it accessible for practitioners and researchers with varying levels of statistical expertise.

Concept

- An envelope model is a type of ecological niche model that identifies the environmental conditions within which a species is likely to occur. It essentially delineates the environmental boundaries or envelope that encapsulates the known occurrences of a species. This model could be applied in the context of Anopheles stephensi to decipher ecological conditions that favor the species' growth.
- Envelope models are often simpler and computationally less demanding than some machine learning-based models.
- The results are often easy to interpret, making it accessible for practitioners and researchers with varying levels of statistical expertise.
- The performance of envelope models can be sensitive to the quality and representativeness of input data.

Concept

- An envelope model is a type of ecological niche model that identifies the environmental conditions within which a species is likely to occur. It essentially delineates the environmental boundaries or envelope that encapsulates the known occurrences of a species. This model could be applied in the context of Anopheles stephensi to decipher ecological conditions that favor the species' growth.
- Envelope models are often simpler and computationally less demanding than some machine learning-based models.
- The results are often easy to interpret, making it accessible for practitioners and researchers with varying levels of statistical expertise.
- The performance of envelope models can be sensitive to the quality and representativeness of input data.

Application

- Envelope models assume that the species' niche is in equilibrium with the environmental conditions. This may not hold true in cases of rapid environmental change.

Application

- Envelope models assume that the species' niche is in equilibrium with the environmental conditions. This may not hold true in cases of rapid environmental change.
- Non-ML envelope models may assume linear relationships between species occurrences and environmental variables, which may not capture complex ecological responses.

Application

- Envelope models assume that the species' niche is in equilibrium with the environmental conditions. This may not hold true in cases of rapid environmental change.
- Non-ML envelope models may assume linear relationships between species occurrences and environmental variables, which may not capture complex ecological responses.
- Application starting from line 255 here

Concept

- MaxEnt (Maximum Entropy) is a popular machine learning-based algorithm used for Species Distribution Modeling (SDM). It's widely used due to its ability to handle presence-only data and its capacity to model complex relationships between species occurrences and environmental variables.

Concept

- MaxEnt (Maximum Entropy) is a popular machine learning-based algorithm used for Species Distribution Modeling (SDM). It's widely used due to its ability to handle presence-only data and its capacity to model complex relationships between species occurrences and environmental variables.
- MaxEnt operates based on the principle of maximum entropy, seeking a probability distribution that is maximally uncertain (maximum entropy) while satisfying constraints imposed by the available information (species occurrences and environmental variables). In the context of SDM, MaxEnt aims to find the probability distribution of environmental variables given the observed species occurrences.

Concept

- MaxEnt (Maximum Entropy) is a popular machine learning-based algorithm used for Species Distribution Modeling (SDM). It's widely used due to its ability to handle presence-only data and its capacity to model complex relationships between species occurrences and environmental variables.
- MaxEnt operates based on the principle of maximum entropy, seeking a probability distribution that is maximally uncertain (maximum entropy) while satisfying constraints imposed by the available information (species occurrences and environmental variables). In the context of SDM, MaxEnt aims to find the probability distribution of environmental variables given the observed species occurrences.
- The objective function in MaxEnt is formulated as:

Concept

- MaxEnt (Maximum Entropy) is a popular machine learning-based algorithm used for Species Distribution Modeling (SDM). It's widely used due to its ability to handle presence-only data and its capacity to model complex relationships between species occurrences and environmental variables.
- MaxEnt operates based on the principle of maximum entropy, seeking a probability distribution that is maximally uncertain (maximum entropy) while satisfying constraints imposed by the available information (species occurrences and environmental variables). In the context of SDM, MaxEnt aims to find the probability distribution of environmental variables given the observed species occurrences.
- The objective function in MaxEnt is formulated as:

$$MaximizeH(p) = -\sum_{i} p(x_i) log(p(x_i)) + \lambda \sum_{j} w_j f_j(x) \tag{5}$$

where $H(p)$ is the entropy; $p(x_i)$ is the probability of the environmental variable x_i; λ is a regularization parameter; w_j are weights assigned to environmental features $f_j(x)$

Thank You