Marine Ecological Modelling
Global Climate Change

Improving transferability of Ecological Niche Modelling

Jorge Assis, PhD // jmassis@ualg.pt // jorgemfa.medium.com
2020, Centre of Marine Sciences, University of Algarve

Improving model transferability

The ability of a model to accurately predict to different data is especially relevant for forecasting species distributions under future climate change. This task is dependent on:

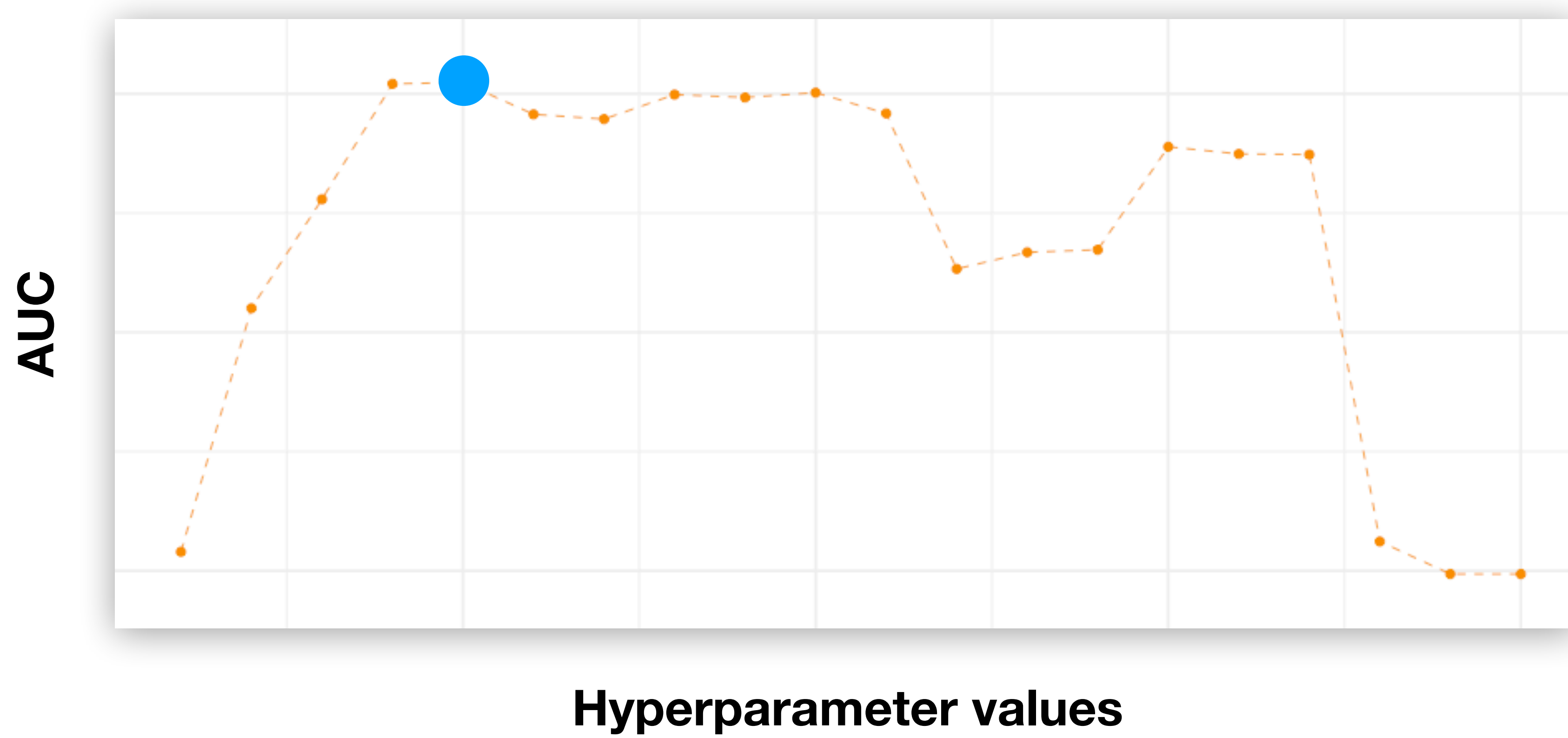
1. Data for model fitting (inc. pseudo-absences; previous sessions);
2. **Proper choice of model parameterisation;**
3. **The complexity of the model.**

Ultimately, transferability is directly related to under / overfitting.

Hyperparameter optimization

In machine learning, hyperparameter optimization **finds the best parameters used to control the learning process** (proper hyperparameters optimally solve the learning problem). The same machine learning algorithm can require different parameter values to generalize different data patterns.

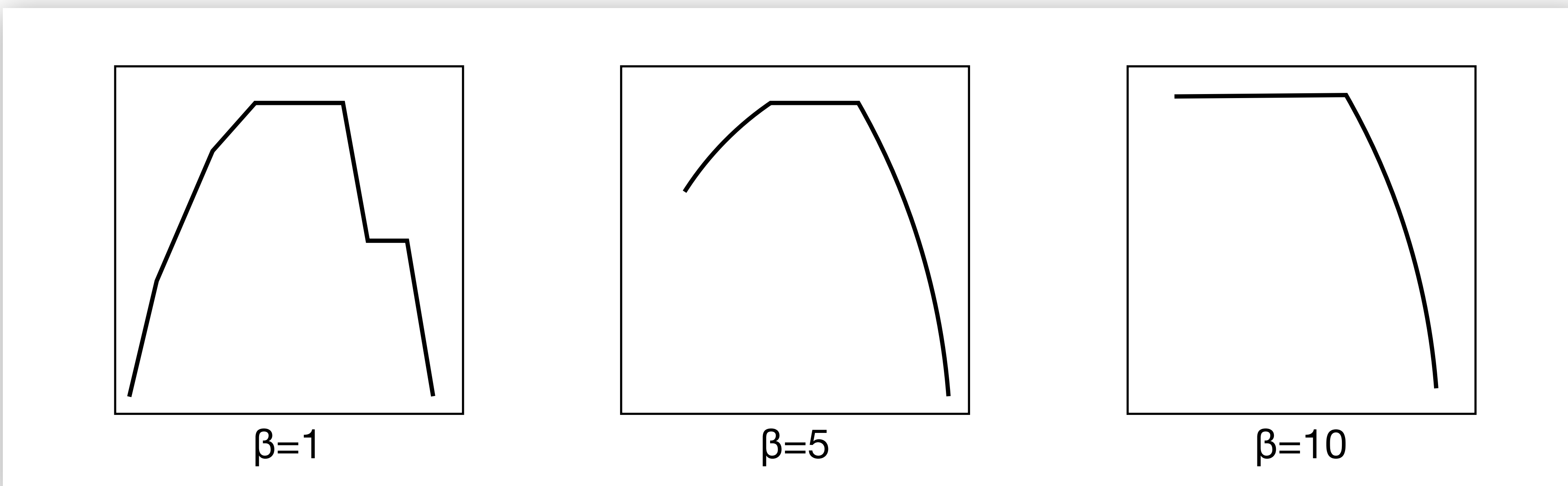
The approach relies on testing **multiple parameter values that minimize a loss function on given independent data**. Cross-validation is often used to estimate generalization performance.



Maxent hyperparameters

An important parameter of Maxent is **regularization**, which **reduces overfitting of the model** in two ways:

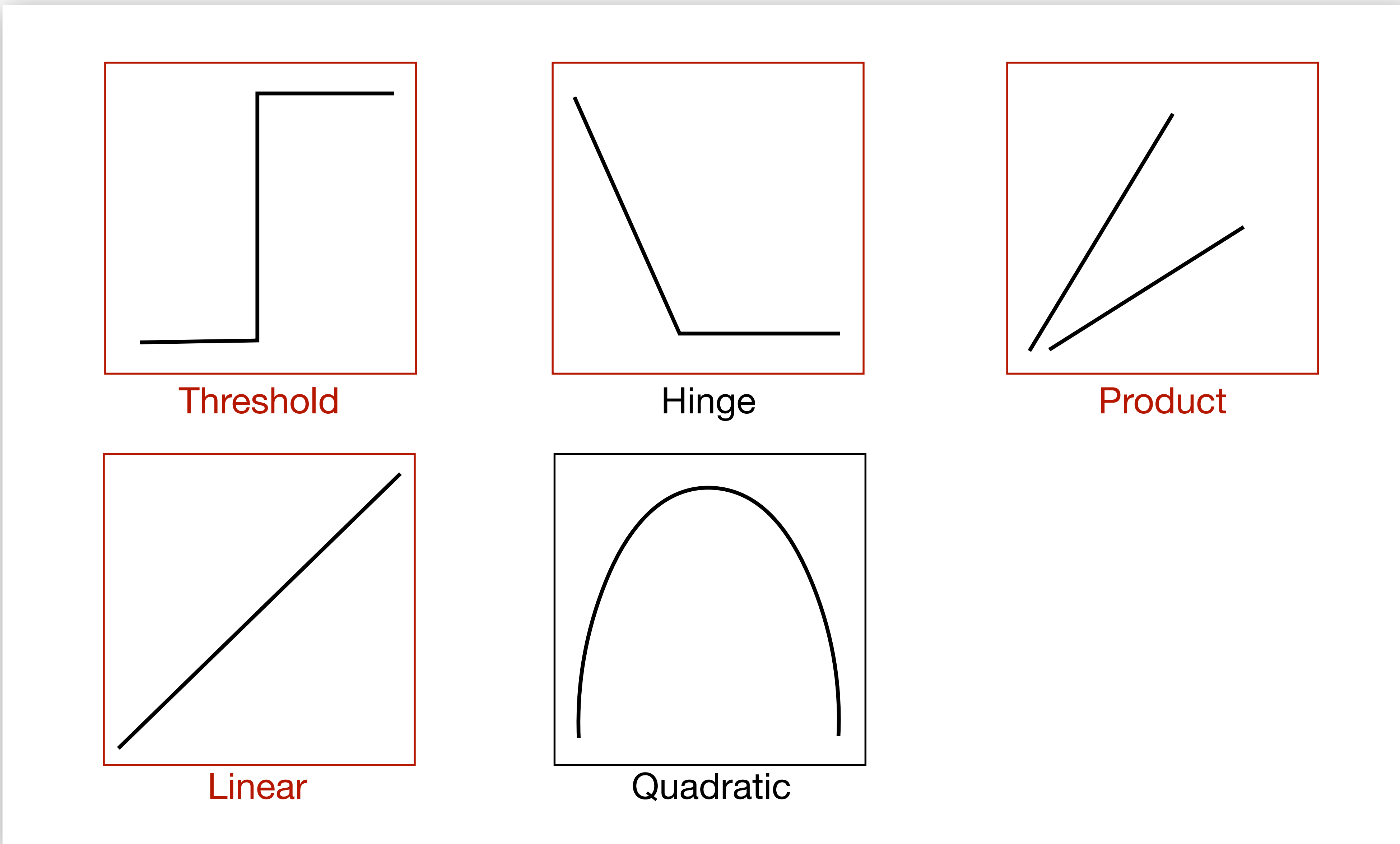
1. **Relaxing the constraints**: instead of fitting the model using the exact constraints of the variables, it takes into account confidence intervals around the constraints, **preventing the model from being fitted to closely around the input data**.



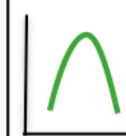
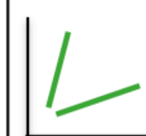
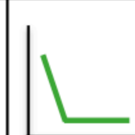
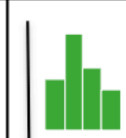
Higher **regularization (β multiplier)** lead to smoother response curves

Elith et al. 2010 Methods Ecology & Evolution

2. **Penalizing complexity**: the model excludes feature types that do not add a significant improvement to the model.



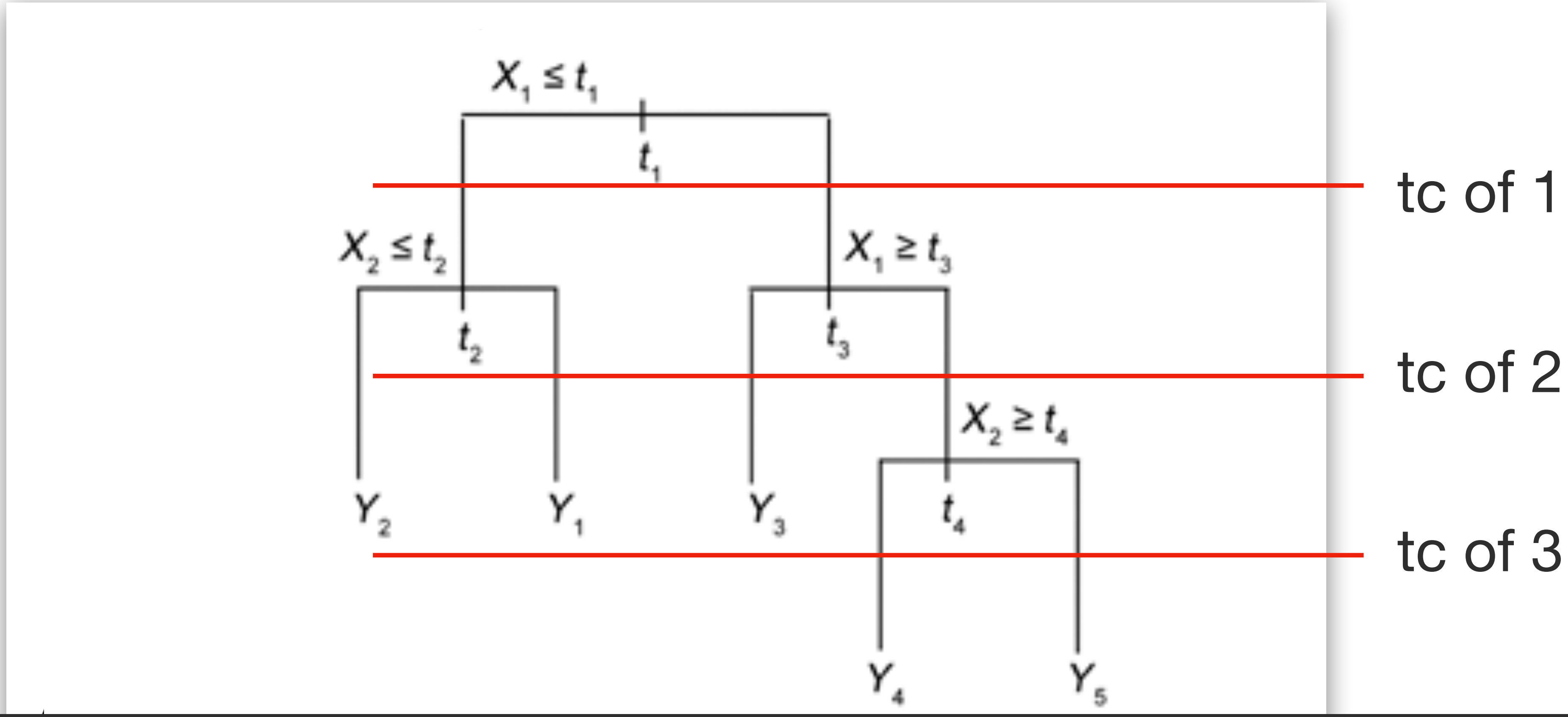
Elith et al. 2010 Methods Ecology & Evolution

Feature type	Interpretation	Constraint	Shape
Linear	Continuous variable	The *mean* of each environmental variable at an unknown location should be close to the mean of that variable in known occurrence locations.	
Quadratic	Square of the variable	The *variance* of each environmental variable at an unknown location should be close to the variance of that variable in known occurrence locations.	
Product	Pairs of continuous variables – allows for interactions	The *co-variance* of two environmental variables at an unknown location should be close to the co-variance of those variables in known occurrence locations.	
Threshold	Conversion into binary response based on a threshold	The proportion of predicted occurrences with values above the threshold (binary response = 1) should be close to the proportion of known occurrences.	
Hinge	As threshold type, but response after the threshold (knot) is linear	The mean above the knot of each environmental variable at an unknown location should be close to the mean above the knot of that variable in known occurrence locations.	
Categorical	Categorical variable	The proportion of predicted occurrences in each category should be close to the proportion of observed occurrences in each category.	

Boosted Regression Trees hyperparameters are:

1. Learning Rate (a.k.a, Shrinkage) determines the contribution of each tree to the growing model. Higher / Lower values mean more / less time to learn, which can translate into under / overfitting.

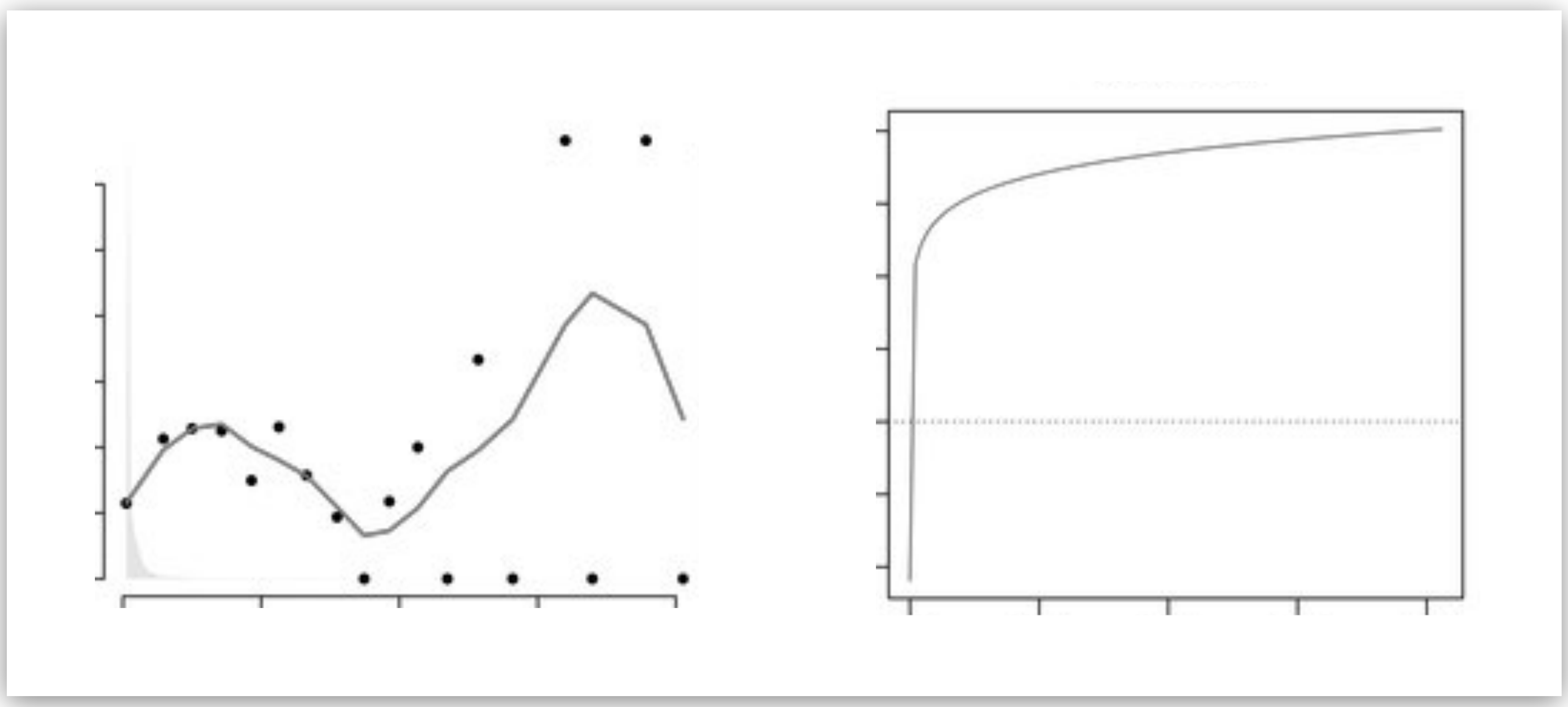
2. Tree Complexity (a.k.a, Interaction Depth) controls whether interactions are fitted. A tc of 1 (single decision stump; two terminal nodes), a tc of 2 fits a model with up to two-way interactions, and so on. Higher / Lower values mean more / less complexity, which can translate into over / underfitting.



Boosted Regression Trees additional hyperparameter is:

3. Monotonicity relationships between the output and predictor variables, which strongly reduces overfitting.

Overfitted response vs. positive monotonicity



Reducing the complexity of the model

Excessively complexity risk overfitting data and can erroneously attribute patterns to random noise, **leading to biased predictions that are too specific to the baseline system** to be transferable.

Greater transferability is expected in parsimonious models with few predictors. **As complexity grows**, so do potential predictor combinations and the likelihood of mismatch between baseline and target conditions, which can **result in incorrect interpolation and extrapolation**.

One approach is to remove all the variables that have a permutation importance lower than 5%. This removes the lowest ranked variable, trains a new model and computes a new rank. The process is repeated until all the remaining variables have an importance greater than 5%.

Yates et al., 2018

Reducing the complexity of the model

A more reliable approaches are the forward and backward propagation methods. For both approaches, a full model built (i.e., with all predictor variables included) and performance assessed (e.g., with AUC).

Forward method :: models are built by adding variables one by one, sorted from the higher to the lowest contributive. This is performed until adding lower contributive predictors does not add predictive power (i.e., equal to the performance of the full model).

Backward method :: the variables are removed from the full model, one by one, sorted from the lowest to the higher contributive until the model starts to loose performance.

The backward method is generally the preferred method, because the forward method produces so-called suppressor effects. These suppressor effects occur when predictors are only significant when another predictor is held constant.

Yates et al., 2018