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 independent data** is particularly **relevant when forecasting species distributions** under **climate change**.

This is mainly dependent on:

1. Environmental and biodiversity data to train the models;
2. **Proper choice / tuning of model parameterisation;**
3. **Reduced complexity of the models.**

Choice of model parameterisation

Machine learning algorithms have specific parameters.

There is a set of **optimal hyperparameters that best control the learning process** (optimal hyperparameters will solve the learning problem; these will return the best fit).

There is **no rule of thumb for the hyperparameters**; the **same algorithm may require different parameter values to generalize different datasets**.

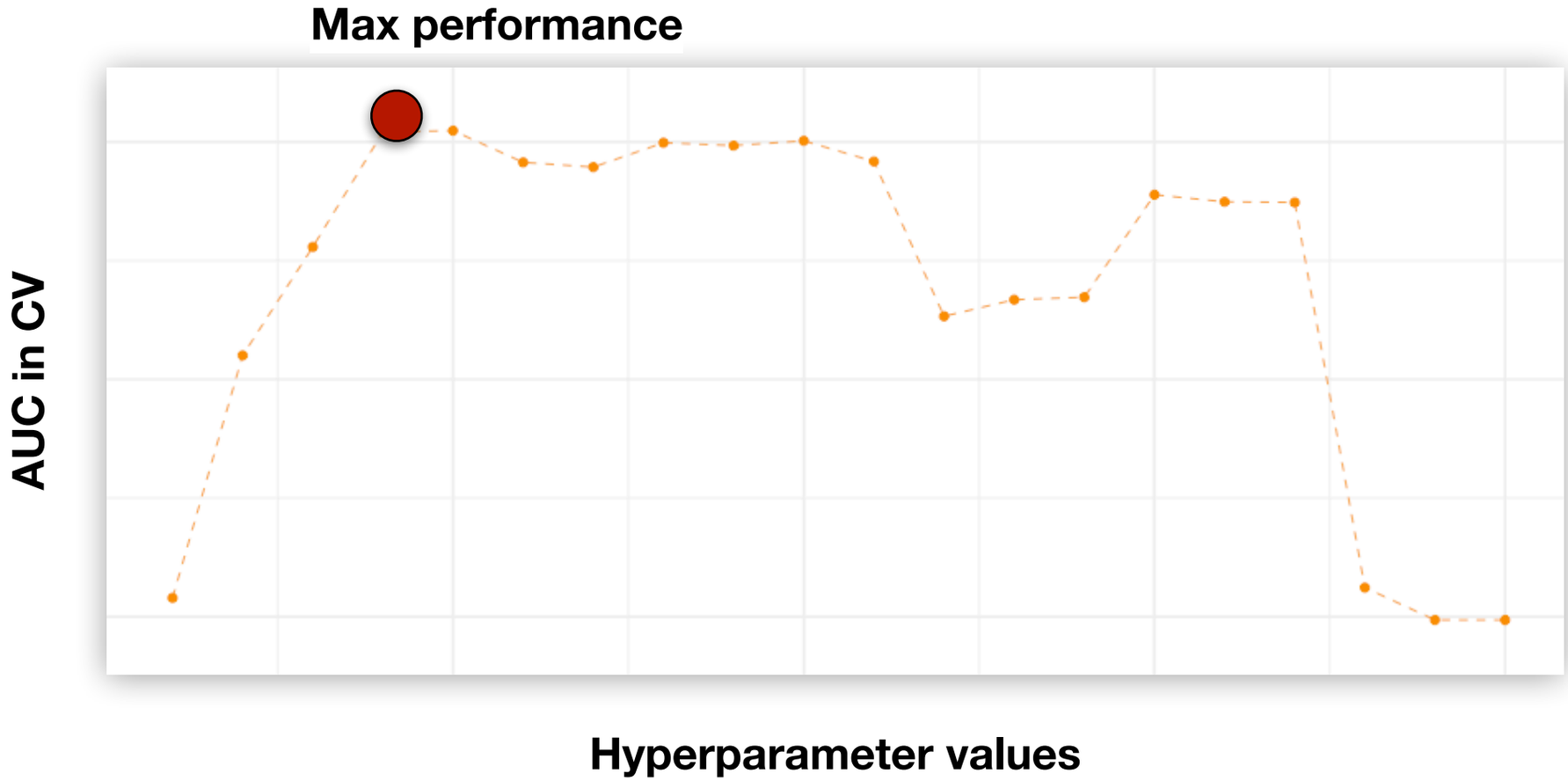
What is the optimal parameter values improving models?
How can we tune these parameters?

Hyperparameter optimization

The approach relies on testing a span of parameter values and finding which value maximizes the performance of models.

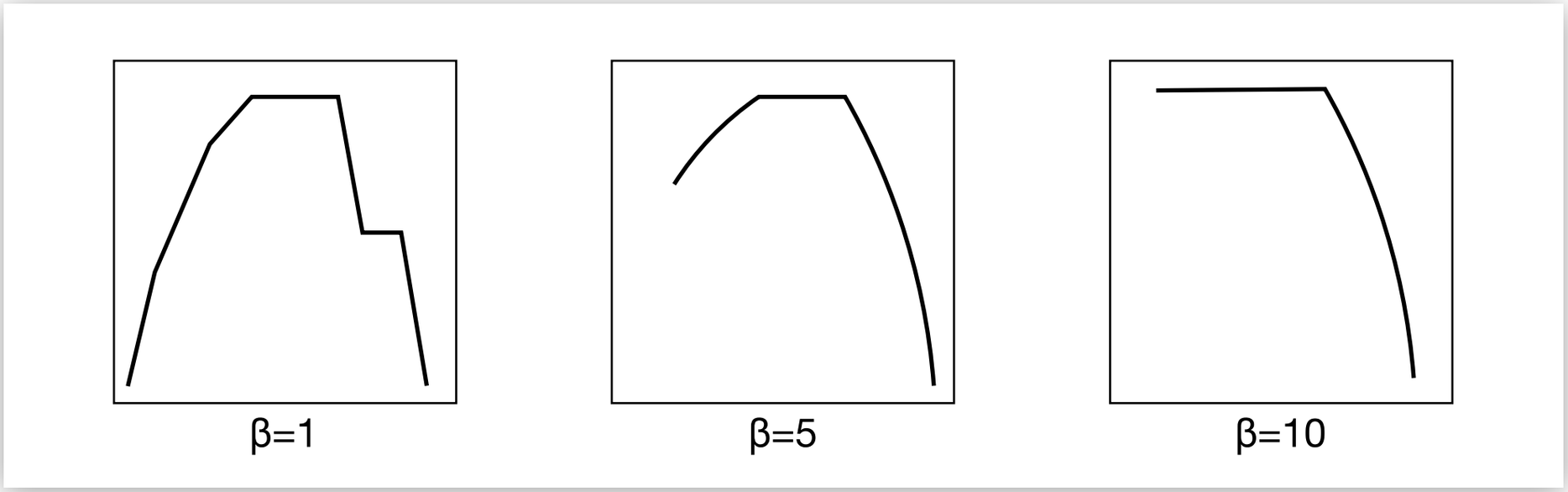
Cross-validation (testing in independent data) is used in the process of hyperparameter optimization.

Which parameter value has higher performance in independent data?



Maxent hyperparameters

A key parameter of **Maxent** is **regularization**, which controls **fit constraints and therefore strongly impacts overfitting**.

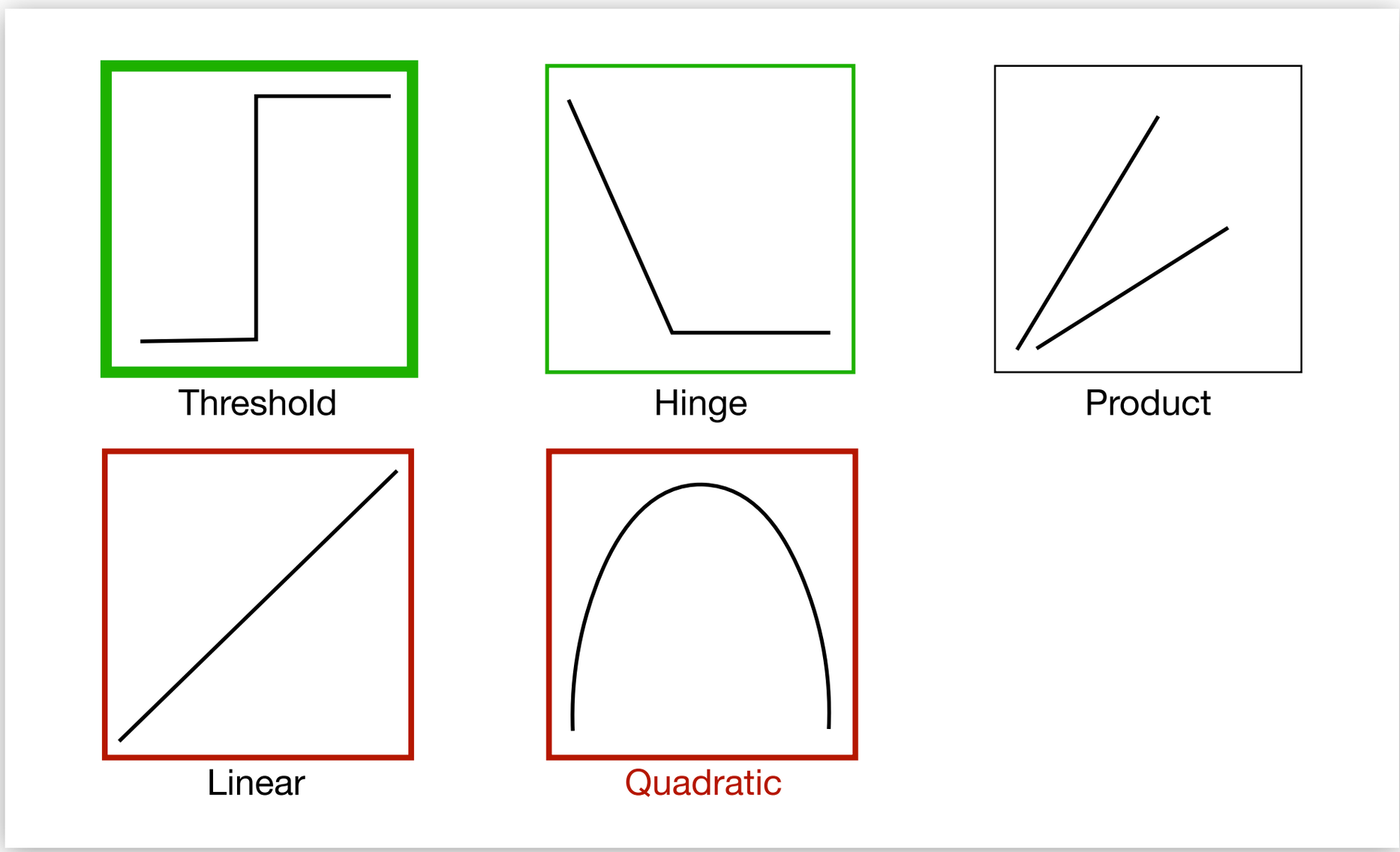


Higher regularization (β multiplier) values reduced the smoothing / relaxation of curves, which in term, can reduce overfitting.

Elith et al. 2010 Methods Ecology & Evolution

An additional parameter of Maxent is feature type.

We tend to **exclude features that are more prone to overfitting.**

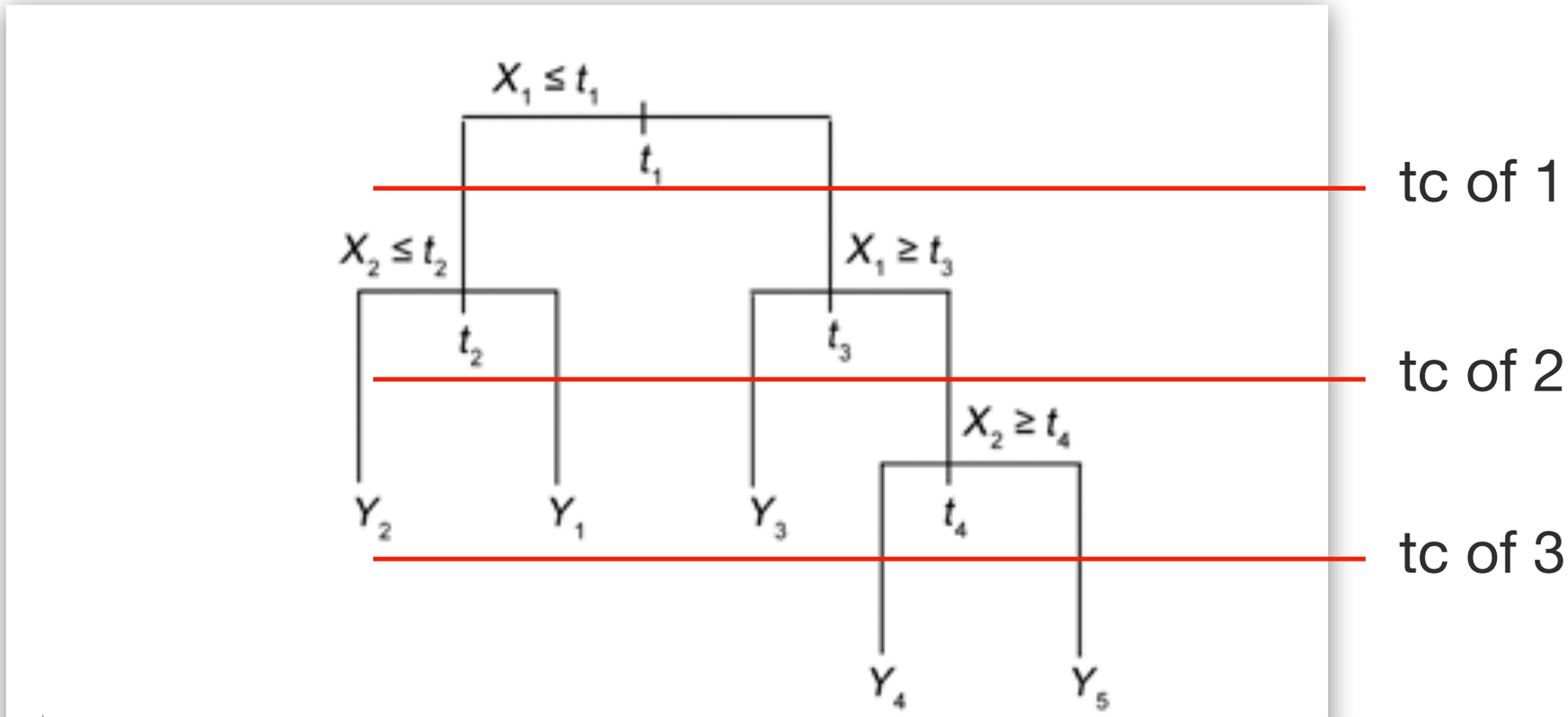


Elith et al. 2010 Methods Ecology & Evolution

Boosted Regression Trees hyperparameters

1. **Tree Complexity controls the degree of fitted interactions**.

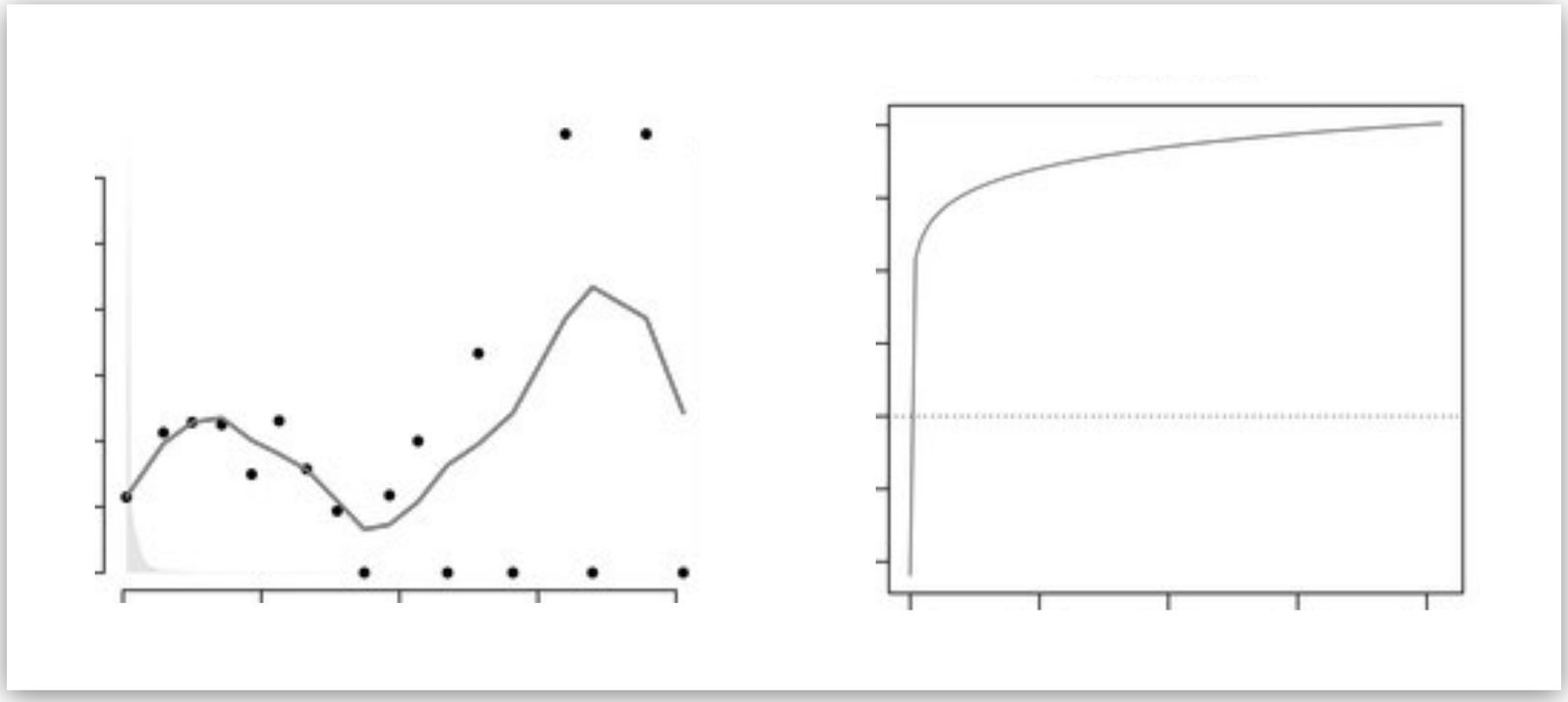
A tc of 1 (single decision; two terminal nodes), a tc of 2 fits a model with up to two-way interactions, and so on. Higher / Lower tc values mean more / less complexity, which can translate into over / underfitting.



2. **Monotonicity relationships between the output and predictor variables**, strongly **reduces overfitting**.

The choice of monotonicity is made a priori **based on ecological theory** (no monotonicity, negative or positive monotonicity).

Overfitted response vs. positive monotonicity



Reducing the complexity of models

Excessive model complexity (too many predictors) risks overfitting data, which leads to biased predictions outside the domain where the models where fitted (baseline).

Greater transferability is expected with simpler models, with few predictors. As complexity grows, so does the predictor combinations and therefore likelihood of mismatch between baseline and target (e.g., future) conditions.

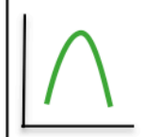
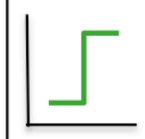
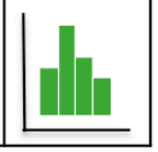
Yates et al., 2018

Reducing the complexity of models

The approach is to remove variables that have a contribution lower than 5%.

1. **Compute the contribution of variables** and **removing the lowest ranked variable.**
2. **A new model is trained and variable contribution is computed again.** The process is **repeated until all the remaining variables have an importance greater than 5%.**

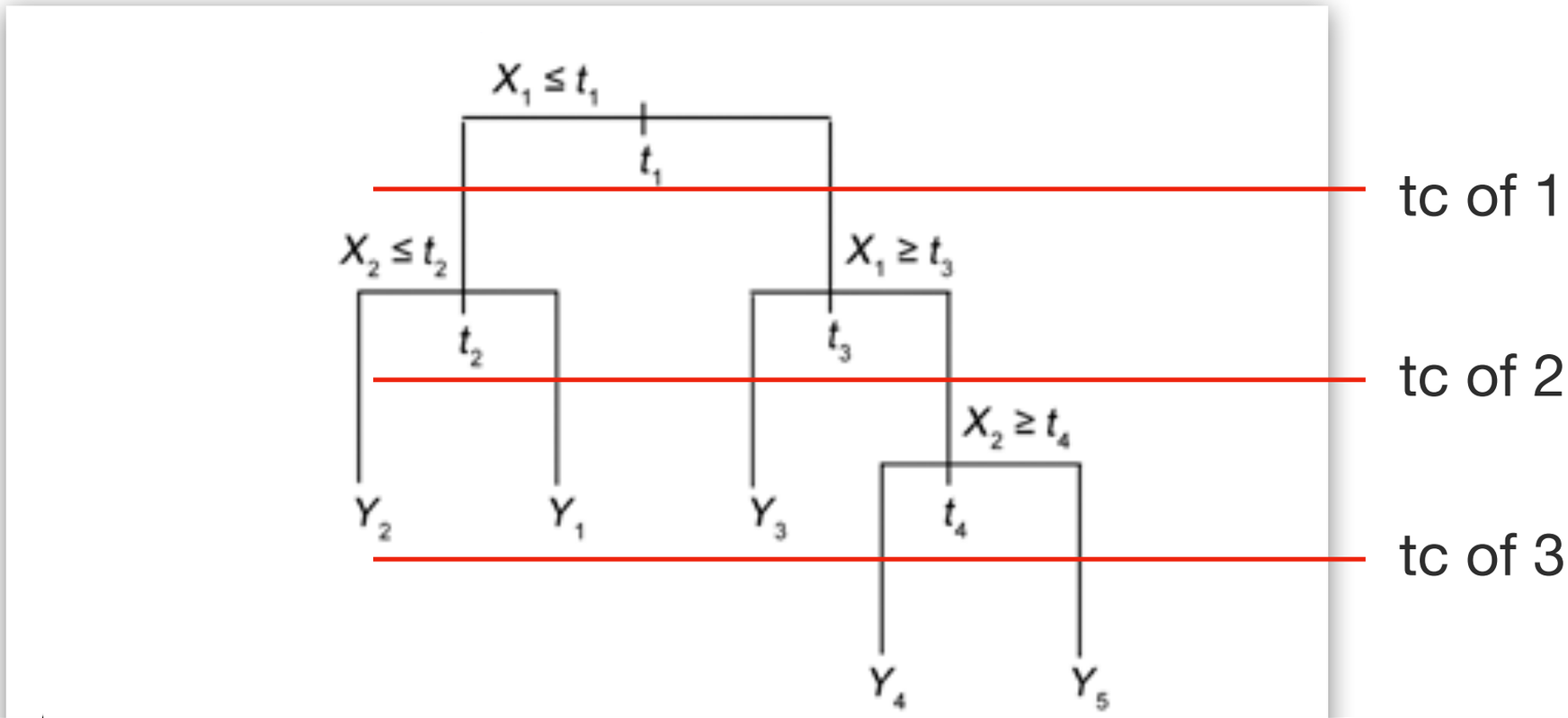
Yates et al., 2018

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

1. **Learning Rate** determines the **contribution of each tree to the final model.** Higher / Lower values mean less / more time to learn, which can translate into under / overfitting.

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



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

Forward method :: deprecated.

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