HABITAT SUITABILITY MODELLING OF BIOTIC COMMUNITIES

[MODELIZACIÓN DE LA IDONEIDAD DE HÁBITAT PARA COMUNIDADES BIOLÓGICAS]

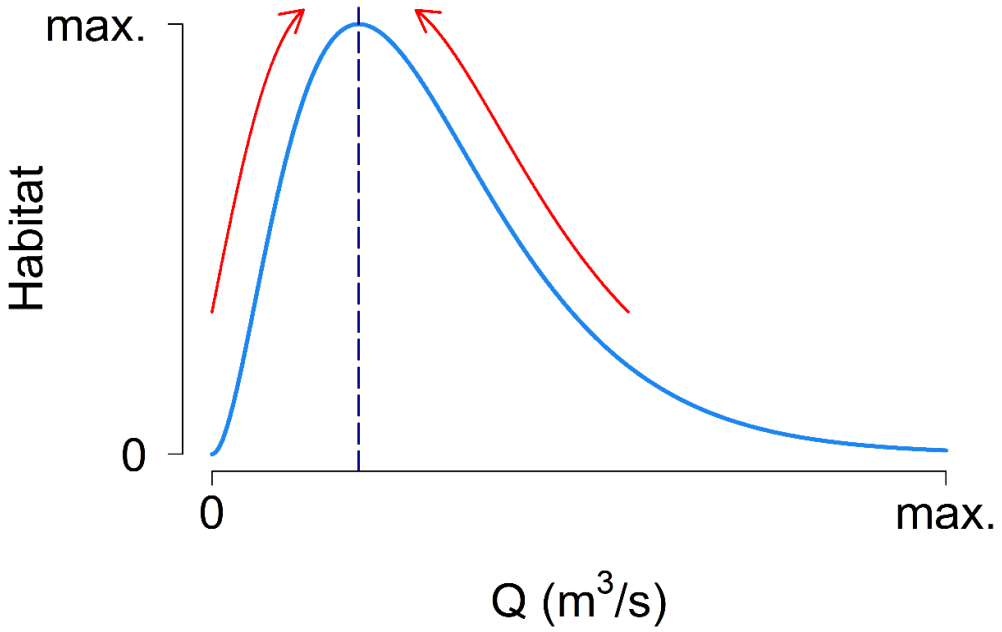
RAFEL MUÑOZ MAS

Universitat de Girona
Grup de Recerca
en Ecologia Aquàtica Continental
GRECO

www.udg.edu/GRECO

ENVIRONMENTAL FLOW ASSESSMENT (EFA)

- *"[t]he quantity, timing, and quality of water flows required to **sustain freshwater and estuarine ecosystems** and the human livelihoods and wellbeing that depend on these ecosystems"*



- **HABITAT SUITABILITY MODELS**
  - Habitat suitability curves
  - Statistical models
  - **Machine learning**

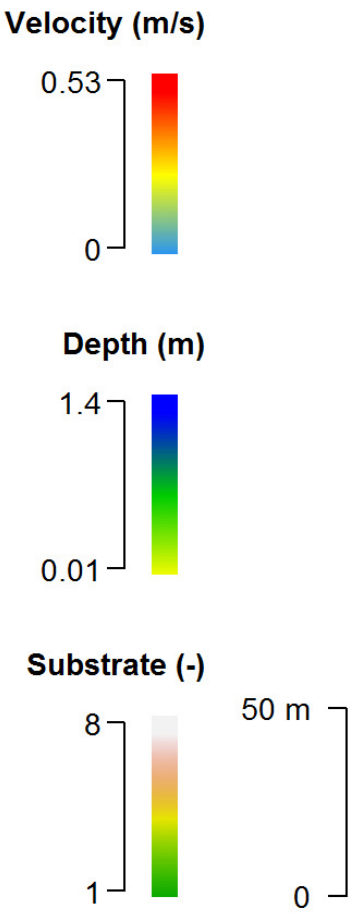
HABITAT SUITABILITY MODELLING – DATA COLLECTION

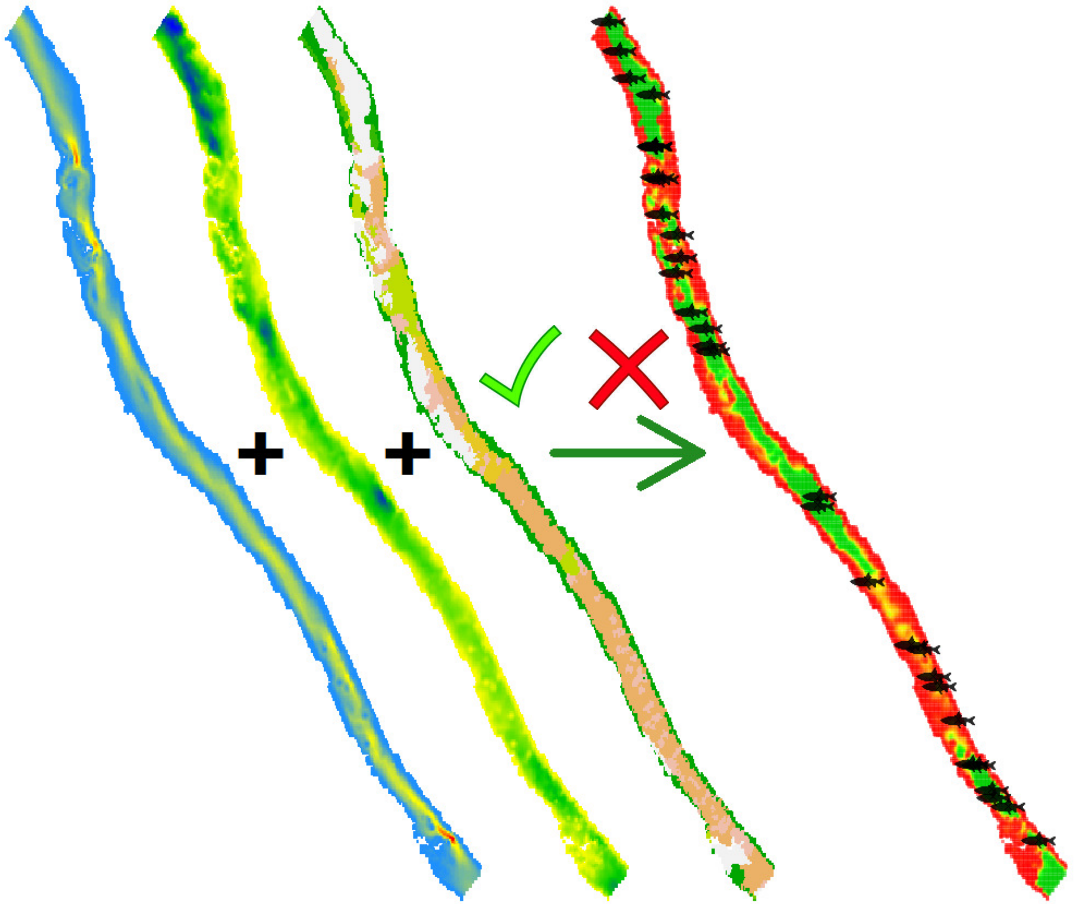
Field survey

- Species presence/absence or abundance
- Habitat characterisation
  - Depth
  - Velocity
  - Substrate
  - Refuge
  - …

III

HABITAT SUITABILITY MODELLING – MAIN OBJECTIVE





1. Variable interactions
2. Non linear
3. Heterogeneous data
4. …

IV

Machine learning – Strengths and Weaknesses

- Non-linear
- Variables interactions
- Multiple input/output variable types: continuous, categorical, ordinal …
- Require programming
- Low interpretability
- Data/knowledge demanding
- No Free Lunch: multiple available techniques + interaction with baseline data + hyper-parameters' selection + variable selection …
- Overfitting
- …

SUPERVISED MACHINE LEARNING CLASSIFICATION TECHNIQUES

Artificial Neural Networks

- **Multi-Layer Perceptrons**
- Probabilistic Neural Networks
- Self-Organising Maps
- ...

Decision tree-based or recursive partitioning

- **Classification & Regression Trees (CART)**
- C4.5/C5.0
- ...
- **Random forests**
- Gradient boosting machines
  or
  Boosted regression trees
- ...

Support vector-based

- **Support vector machines**

Fuzzy logic-based

- **Mamdani Fuzzy Rule-Based Systems**
- 0-order Takagi-Sugeno-Khan FRBSs
- ...

Bayesian-based

- Bayesian belief networks
- ...

Piecewise linear

- Multivariate adaptive regression splines

Statistical

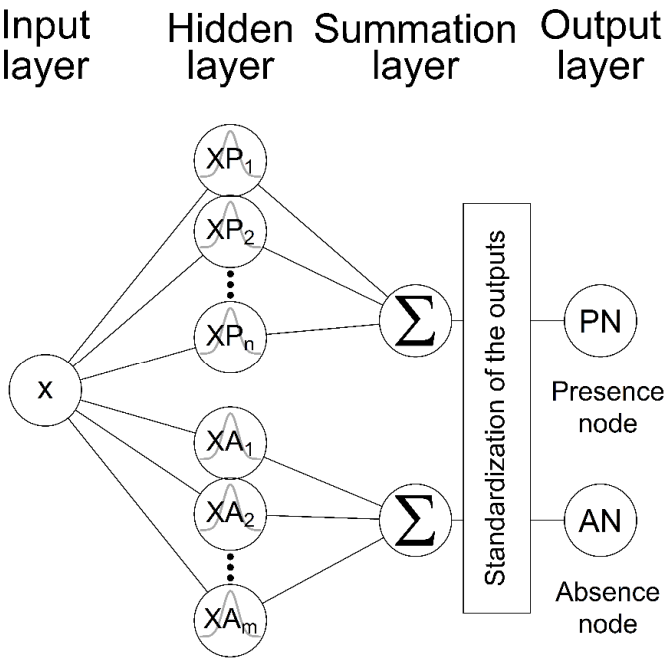
- Generalised Linear Models (GLMs)
- Generalised Additive Models (GAMs)
- ...

Ensemble learning

SUPERVISED MACHINE LEARNING CLASSIFICATION TECHNIQUES

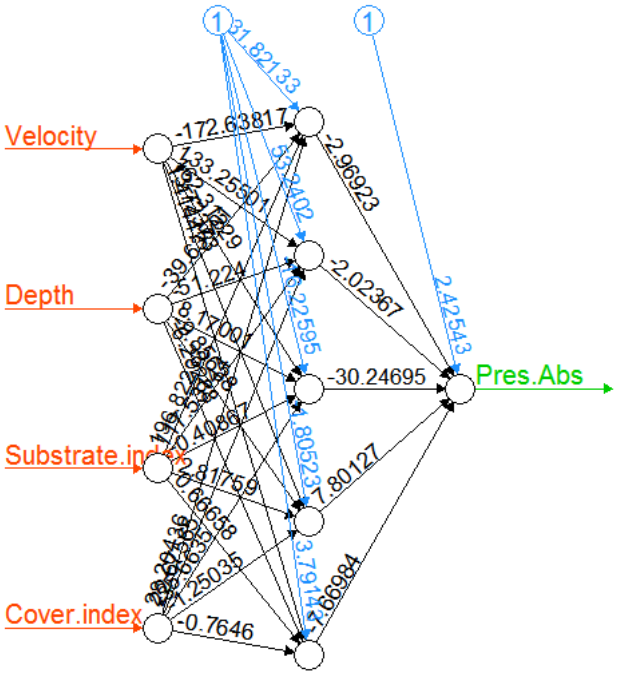
1. Multi – Layer Perceptrons (MLPs)
2. Decision trees (CART)
3. Random Forests (RFs)
4. Support Vector machines (SVMs)
5. Fuzzy Rule-Based Systems (FRBSs)

- Ensemble learning



Multi-Layer Perceptrons (MLPs)

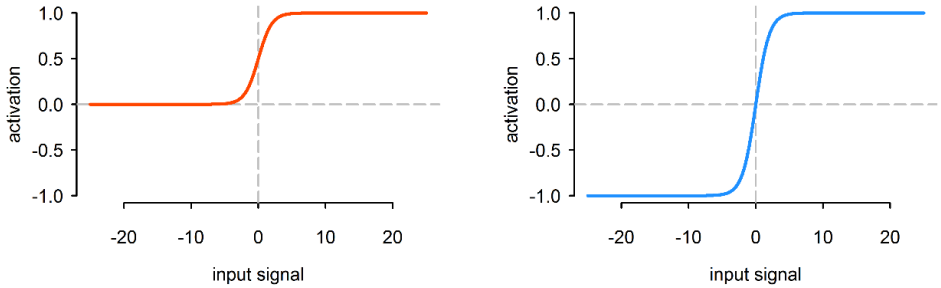
VII



Error: 36.567656 Steps: 8096

$$y_k = g\left(\sum_{j=0}^{M} w_{kj}^{(2)} h\left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i\right)\right)$$

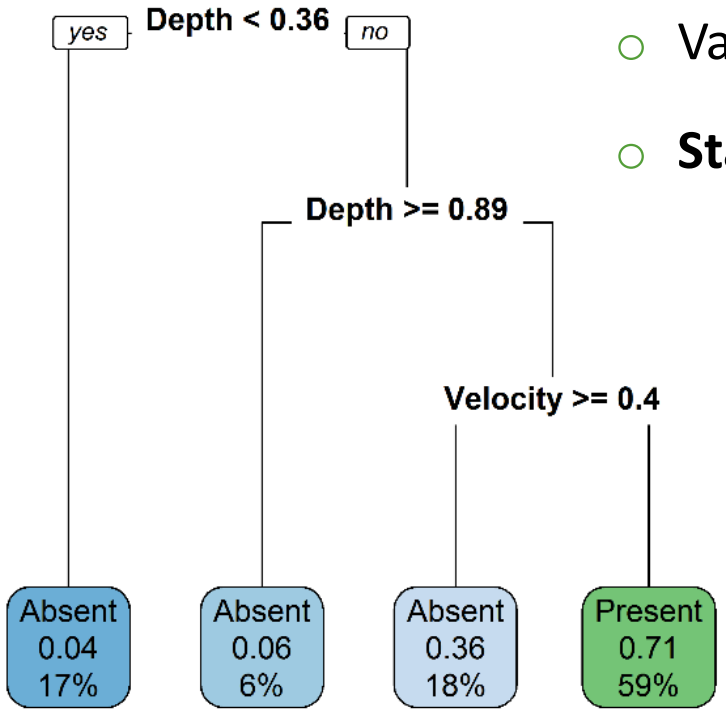
- 1st machine learning technique (McCulloch & Pitts, **1943**).
- Backpropagation algorithm (Rumelhart et al., **1986**).
- Inspired by human brain.
- Only **numerical data.**
- They are non-interpretable → **Black box.**

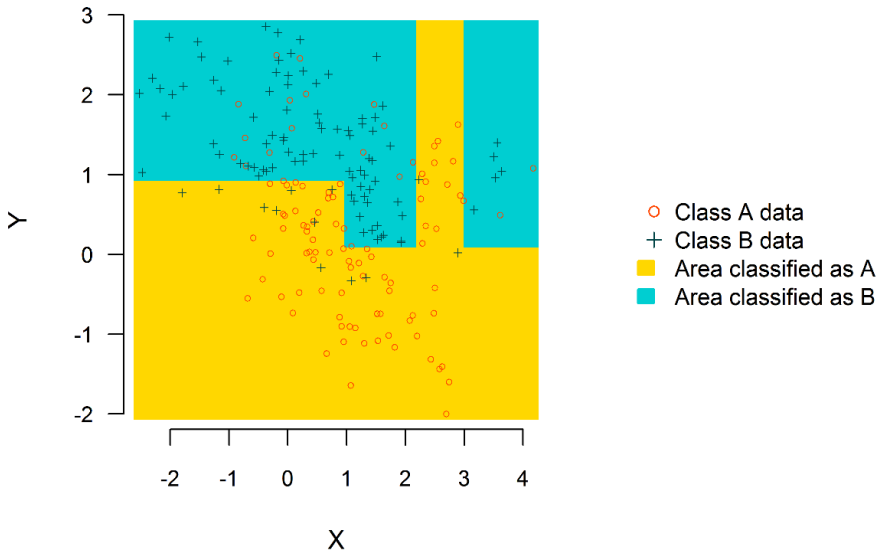


- They were gradually abandoned until the popularization of deep learning.

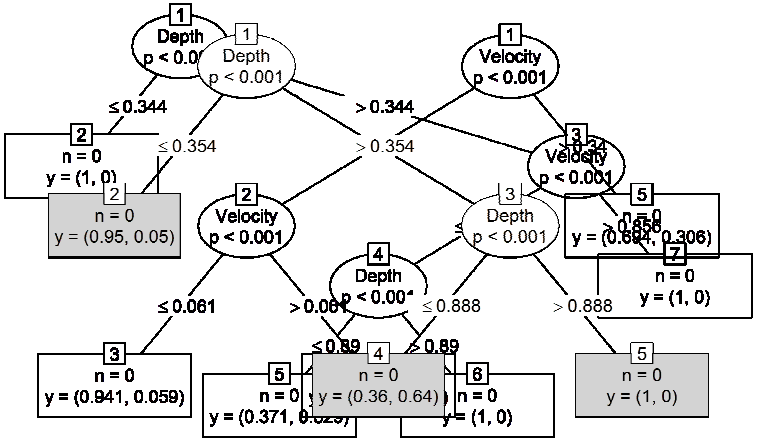
Decision trees - CART

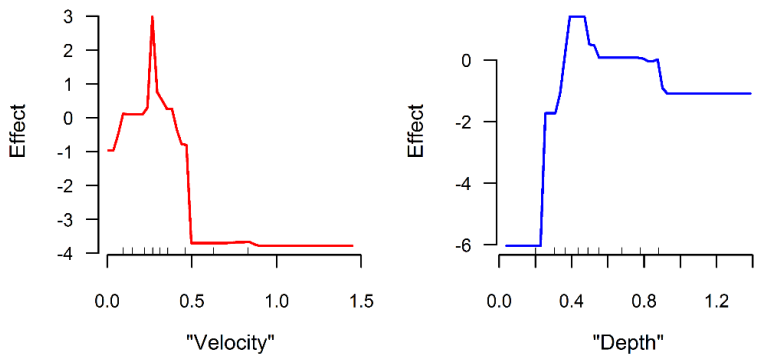
- Classification And Regression Trees (CART) (Breiman 1984)
- They are considered **interpretable** ML techniques
- Variables effects modeled as a **hierarchy** → Evolutionary trees
- **Stair-like decision surfaces** → Multivariate decision trees





Random Forests - RFs

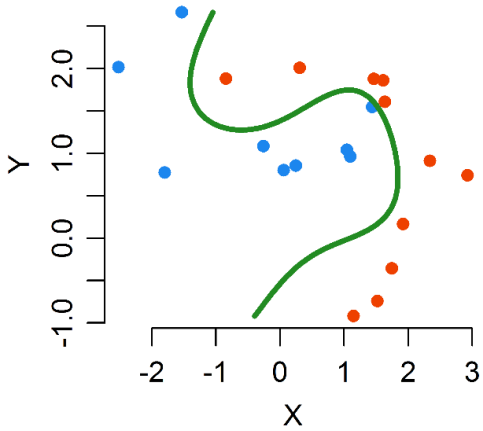


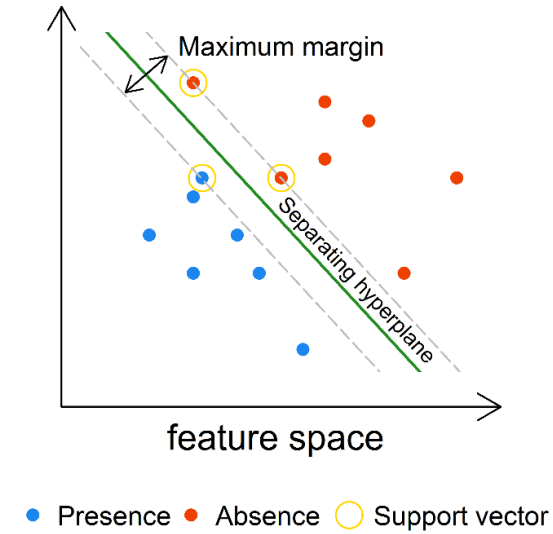


- **Ensemble ML technique** based on the **aggregation** of **CARTs** (Breiman 2001).
- Keep decision tree advantages & render paramount **accuracy** (**1st**).
- Solve multiple CART drawbacks (e.g. hierarchy) but **THEY DO OVERFIT.**
- Others drawbacks emerged → **Conditional random forests.**
- They are non-interpretable → **Black box.**
- Triggered the development of multiple **variable importance** approaches.

Support Vector Machines - SVMs

- Support Vector machines (Cortes & Vapnik 1995) are non-linear classifiers that use the **kernel trick** to create maximum-margin discriminant/separating hyperplanes.
- Paramount **accuracy** (**2nd**).
- Only **numerical data**
- They are **non-probabilistic**, although approaches exit (e.g. Platt 2000)
- They are non-interpretable → **Black box**.

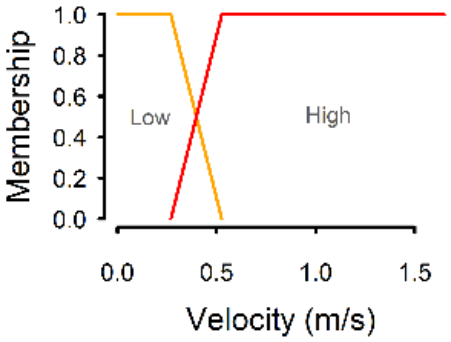


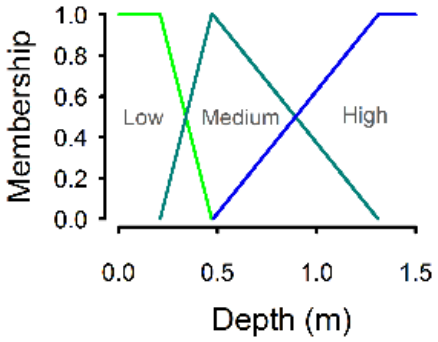


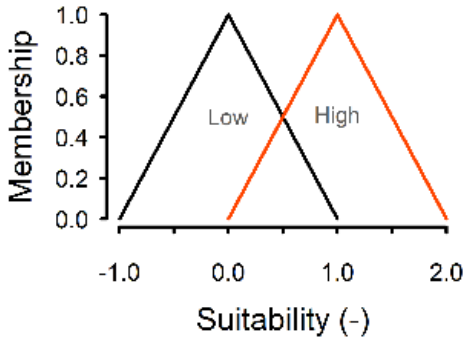
Fuzzy Rule-Based Systems - FRBSs

- Mimic human reasoning → **Interpretable**
  - ***IF** velocity is Low, depth is High, substrate is Medium **THEN** the habitat suitability is High.*
- Based on Zadeh's fuzzy set theory (Zadeh 1965)
- Can be **data-driven**, **expert-knowledge/literature-based** or both (**hybrid**)

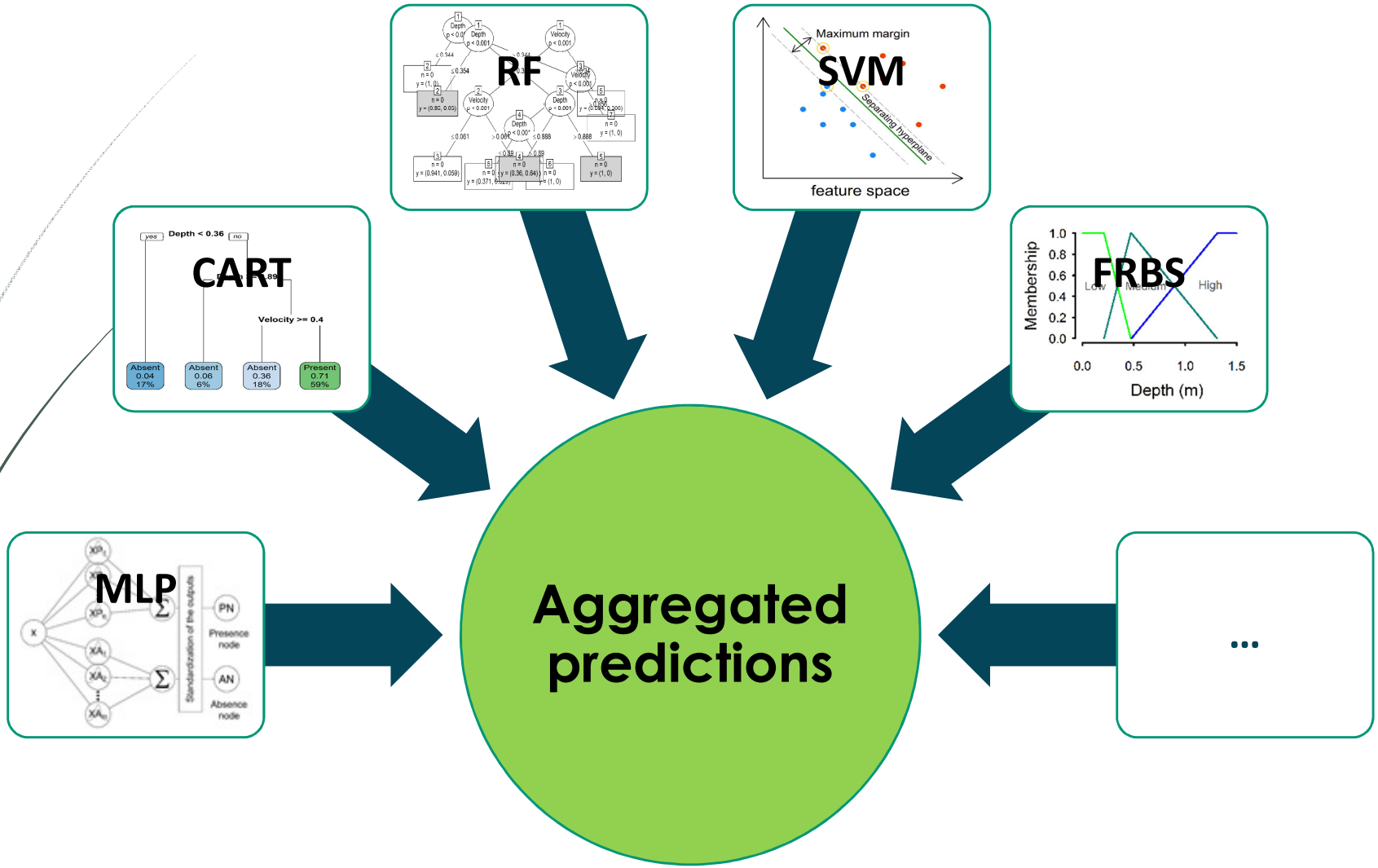
Velocity	Depth	Suitability
Low	Low	Low
High	Low	Low
Low	Medium	High
High	Medium	Low
Low	High	Low
High	High	Low



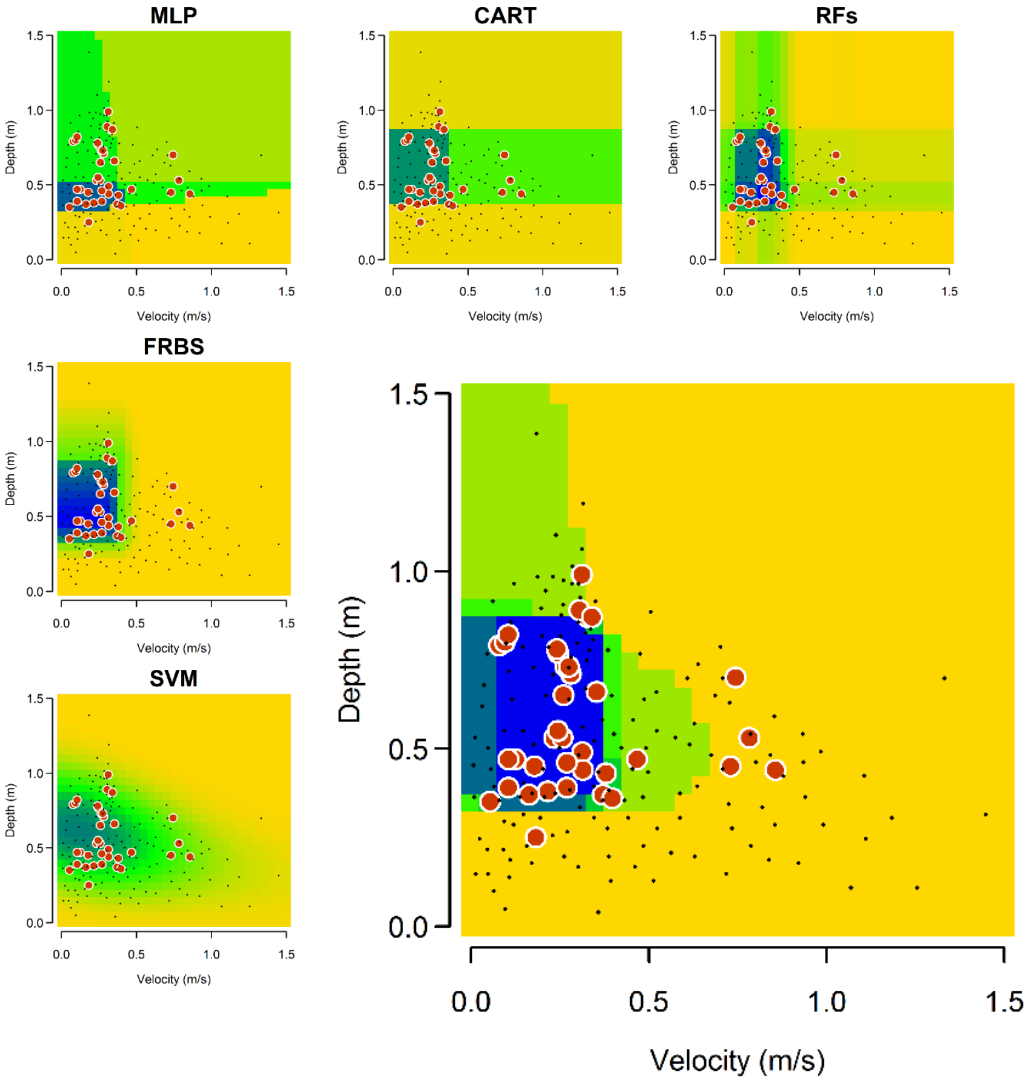




ENSEMBLE LEARNING



ENSEMBLE LEARNING - EXAMPLE



Squalius valentinus

- Small cyprinid
- Vulnerable (VU)
- Serpis & Cabriel rivers

Conclusions

- Suited for habitat suitability/preference modelling → flexible techniques (different data types +non-linear + variables interactions + …)
- Multi-Layer Perceptron (MLP) (1st ML technique) → Deep learning
- Exploration/interpretation → Decision trees (e.g. CART)
- Performance → Random Forests (RFs) (1st option)
- Performance → Support Vector Machines (SVMs) (2nd option)
- Incomplete data → Fuzzy Rule-Based Systems (FRBSs)
- Unclear/unreliable habitat suitability/preferences → Models' ensembles

Thank you for your attention