

A new workflow to flag outliers in species distribution models

Anthony Basooma

PhD Student

anthony.basooma@boku.ac.at

University of Natural Resources and Life Sciences Vienna (BOKU)

A. Schmidt-Kloiber, S. Domisch, Y. Torres-Cambas, M. Smederevac-Lalić, V. Bremerich, P. Meulenbroek, M. Tschikof, A. Funk, T. Hein, F. Borgwardt

April 15-17, 2024 | Oosterpoort, Groningen, The Netherlands

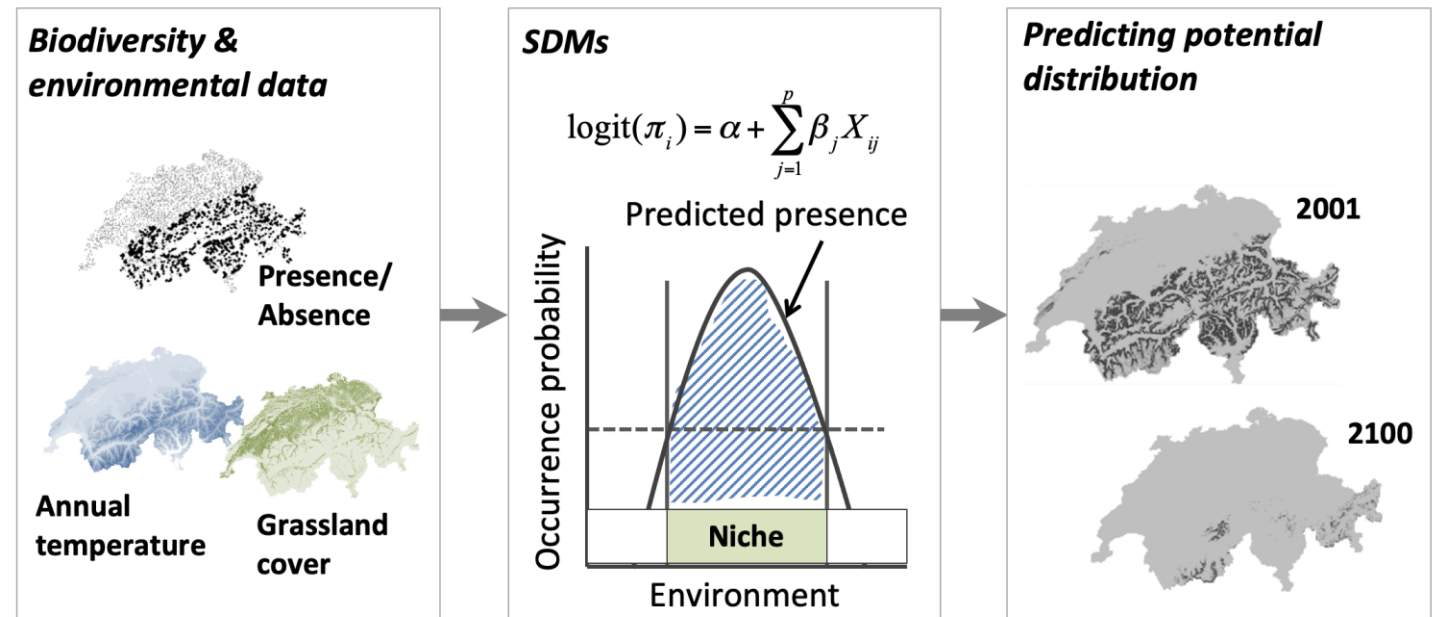
SDMs correlate **species records** and **environmental data** such as **stream power index, elevation, land use, annual temperature, or grassland cover.**

SDMs describe species distribution patterns and make predictions (Franklin & Leahwick 2010).

SDM- a tool that aids conservation planning.

Heterogenous data sources affect data quality, e.g., local or online repositories.

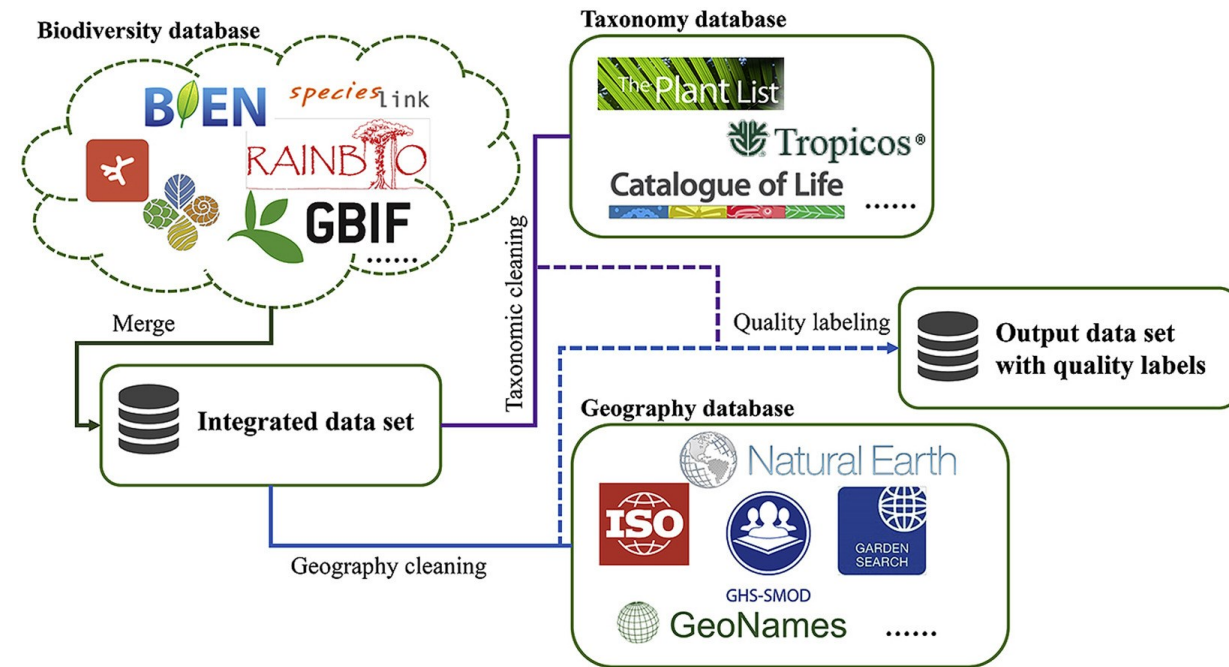
Data cleaning is indispensable.



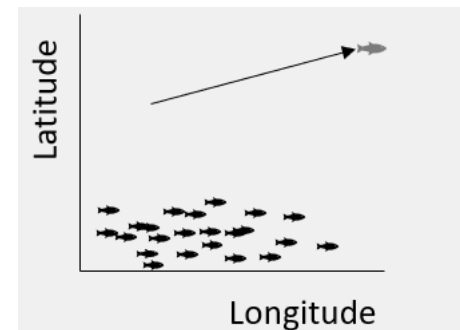
Species records (temporal, taxonomic, or coordinates) (Zizka et al. 2019; Jin & Yang 2020).

Environmental outliers are ignored??

Environmental outliers are records that are distant in environmental space (Newbold, 2010)



Jin & Yang 2020





IGB

Leibniz-Institut für Gewässerökologie
und Binnenfischerei



BOKU
UNIVERSITY

Solution to environmental outliers

Ecological ranges for species such as pH, temperature, depth, or distribution
(**ecological thresholds**).

Home ranges for species (Maarten et al., 2022).

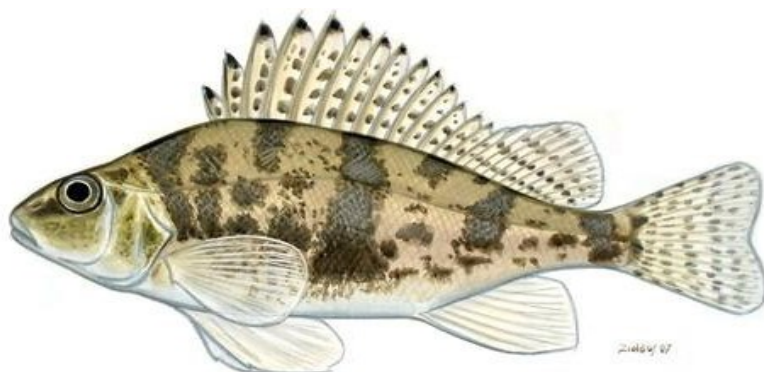
Detect records that are distant in environmental space (**ensemble outlier
detection**).



eOSC | AquaINFRA



1. Species ecological ranges



***Gymnocephalus baloni* (Danube ruffe)**

<https://fishbase.mnhn.fr/summary/9400>

- pH range: 6.8 - 7.8
- Temperature: 10°C - 20°C

Solution to environmental outliers

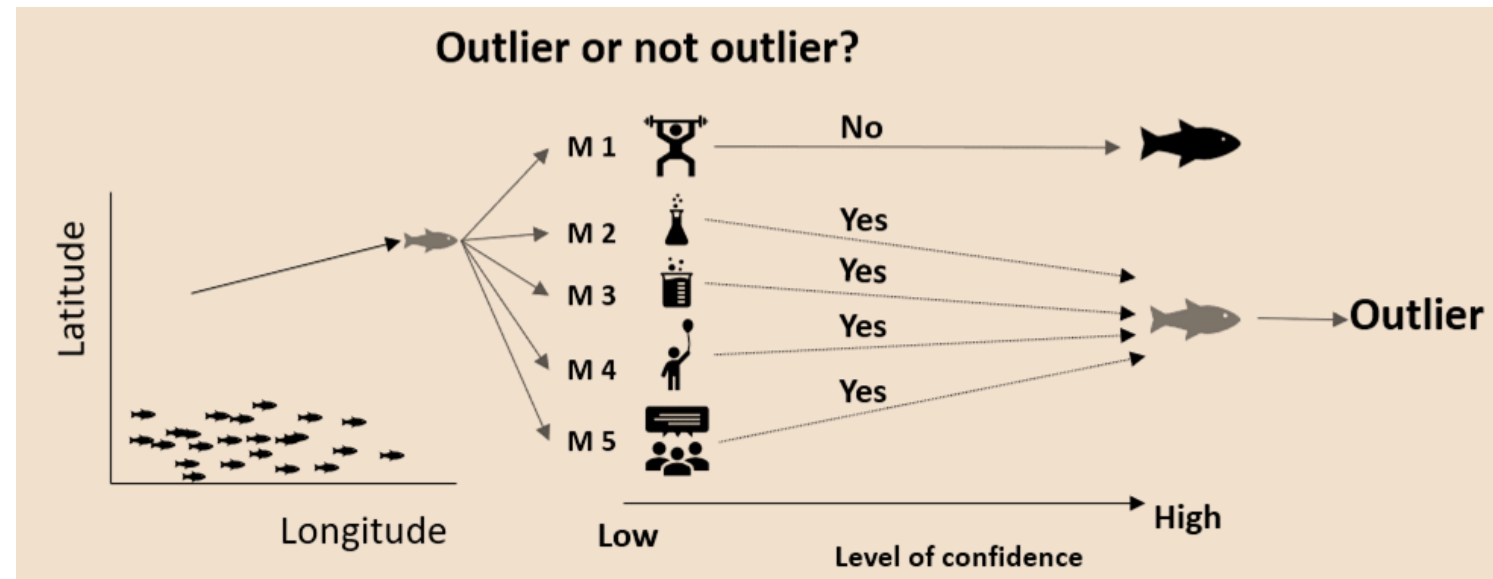
2. Ensemble outlier detection

Multiple outlier detection methods to confirm a particular record as an **absolute outlier**.

Advantage of EOD

Individual method weaknesses considered.

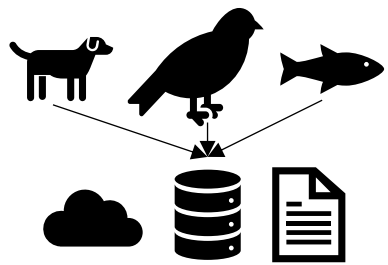
The Z-score assumes that data is normally distributed.



Implementing ensemble outlier detection

specleanr package

Data acquisition & harmonization



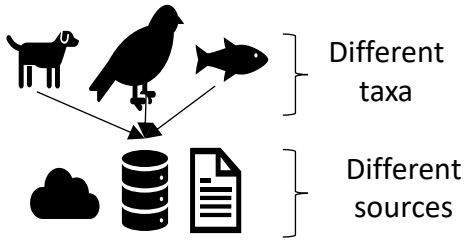
} Different taxa

} Different sources



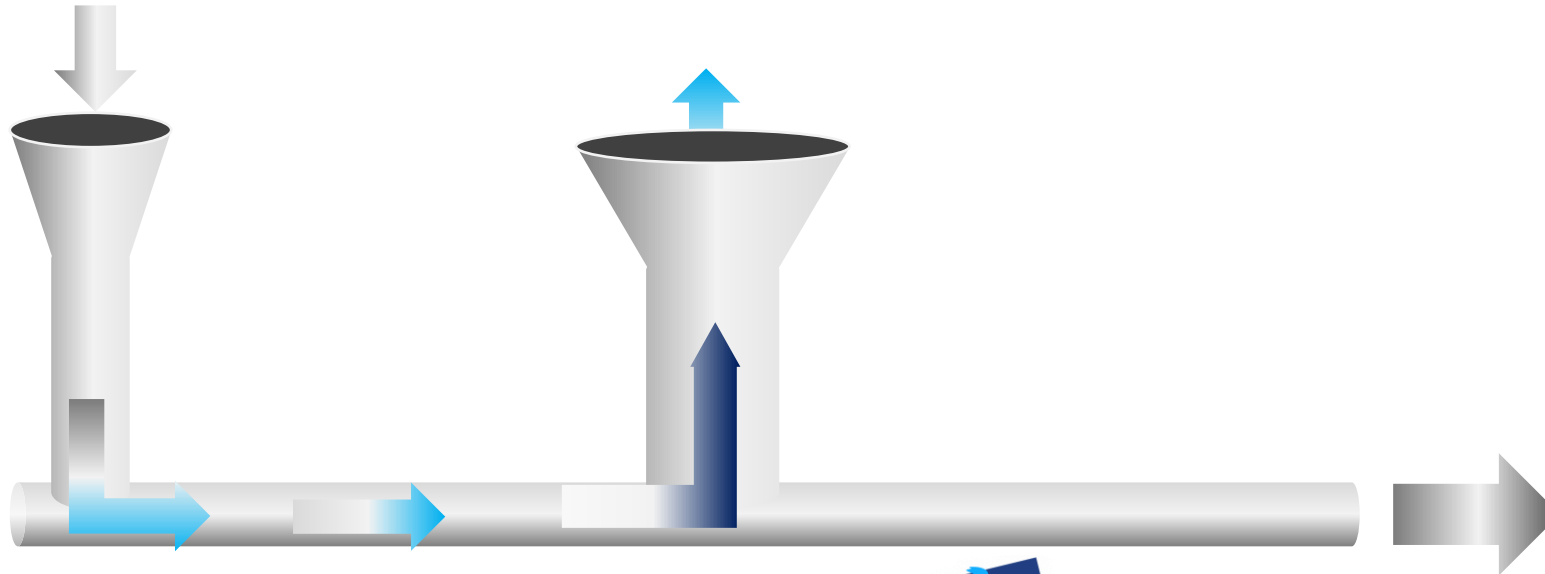
specleanr package

Data acquisition & harmonization



Geospatial checks

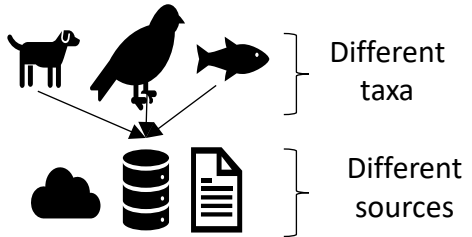
- Basin ranges
- Natura2000
- Distribution ranges
- Latitudinal ranges



Implementing ensemble outlier detection

specleanr package

Data acquisition & harmonization

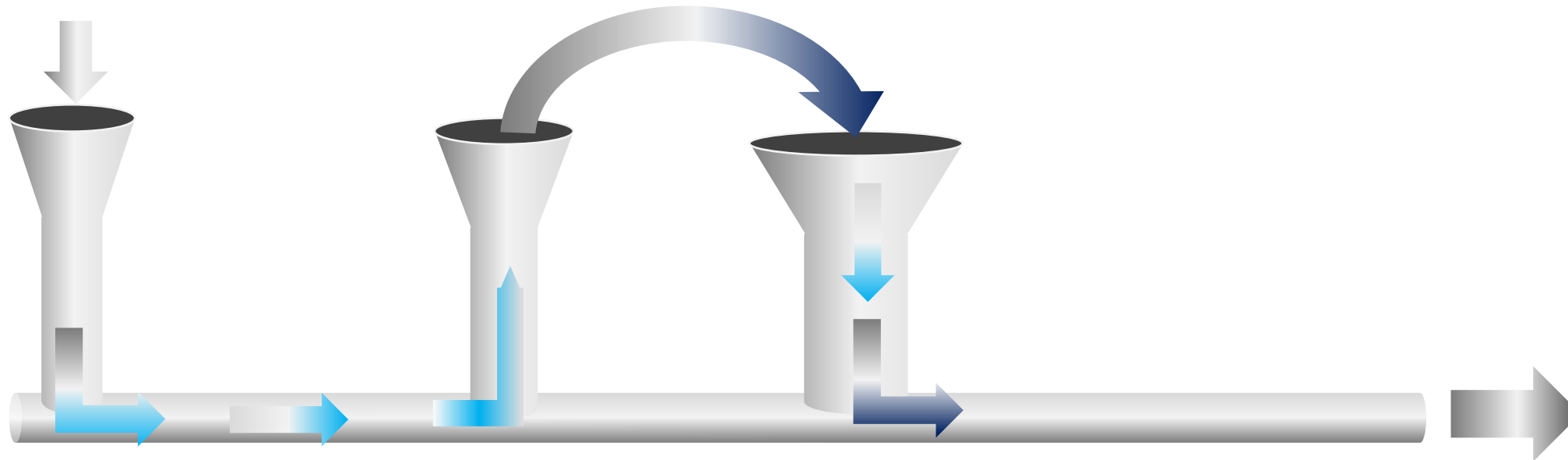


Geospatial checks

Basin ranges
Natura2000
Distribution ranges
Latitudinal ranges

Outlier detection

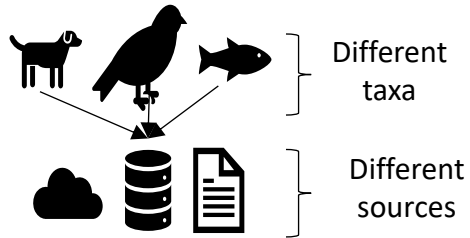
Univariate methods
Multivariate methods
Ecological thresholds



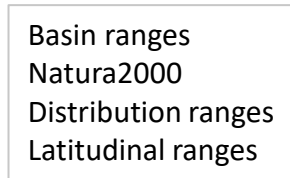
Implementing ensemble outlier detection

specleanr package

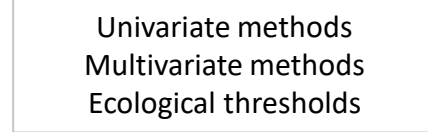
Data acquisition & harmonization



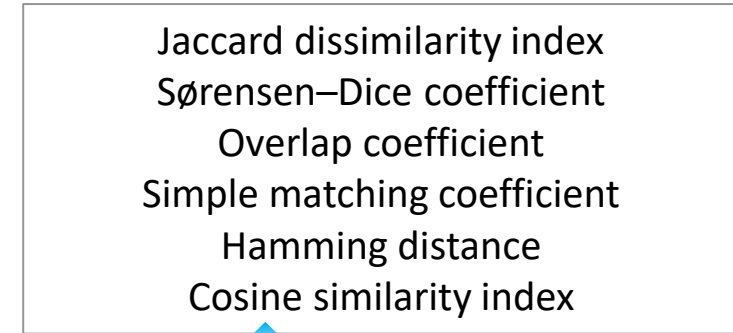
Geospatial checks



Outlier detection



Best method selection



Absolute outliers

Majority votes

Clean data



SDM

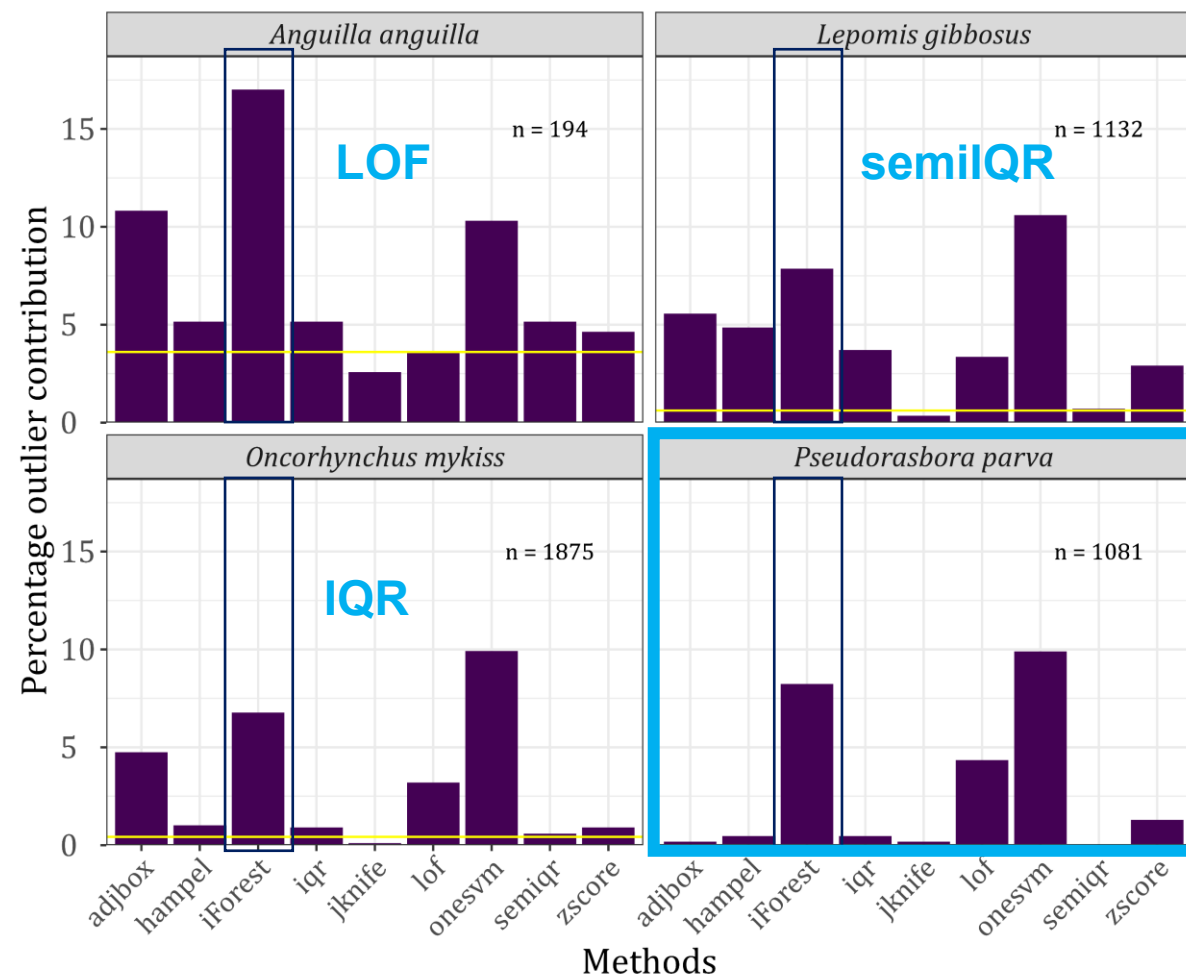
Case study-Invasive species in DRB

WORDCLIM data-environmental data
(BIO6).

Outliers flagged by each method vary.

The absolute outliers vary across
species.

At the threshold of 0.6, stone moroko
(*Pseudorasbora parva*) had no absolute
outliers.



Conclusions

Subjective or ad-hoc method selection is problematic.

EOD is vital in handling weaknesses in method outlier detection

The workflow is applicable across taxa and biogeographical models.

Warmly welcome to test the package ([specleanr*](#))

Thank you for listening