

Species



Goals – 1. the compilation of a quality-controlled dataset of marine species occurrence data which can be used for benchmarking Species Distribution Model algorithms (SDM), 2. linked to environmental data sets and 3. trait databases.

Hurdles – selection of a set of *identifiable species* evenly distributed over *taxonomic groups*, *range sizes* and *ecoregions* with sufficient unique distribution records.

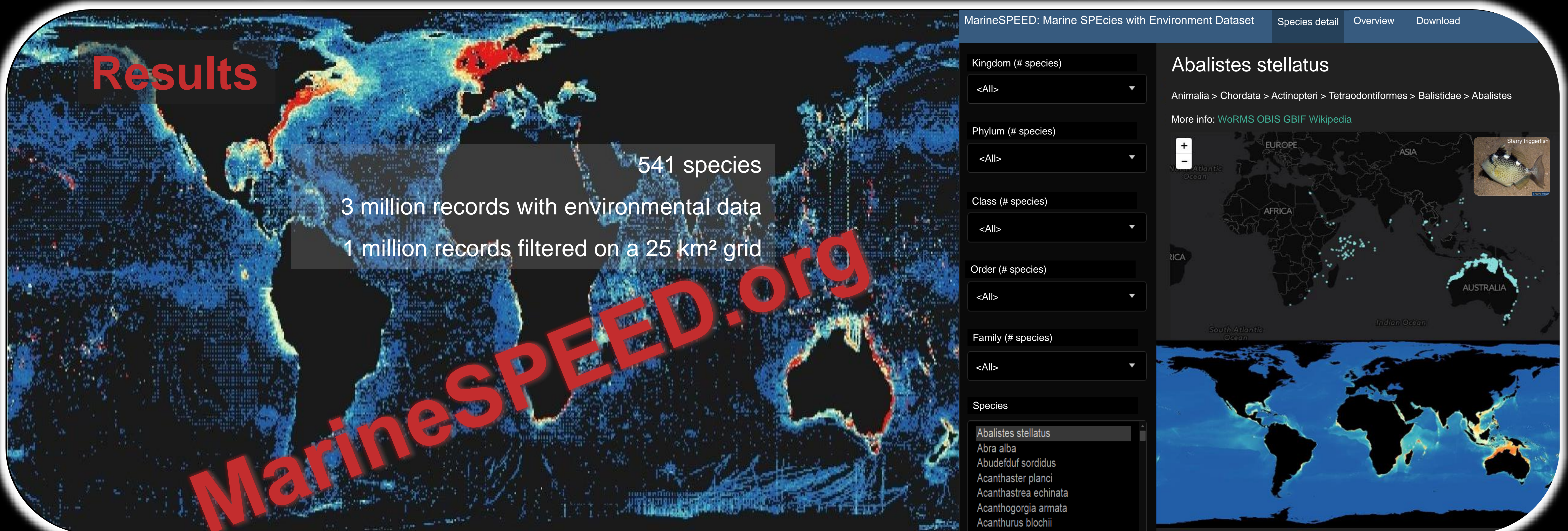
Sources and criteria – distribution records from public repositories such as OBIS, GBIF, EMODNET and Reef Life Survey, filtered for emblematic species with sufficient records, known distributions, allowing identification of gaps and errors.

Links – taxa are linked to WoRMS and species trait databases.

Environment

Environmental data from Bio-ORACLE and MARSPEC were added to all species distribution records with the *sdmpredictors* R package.

Results



Download

At marinespeed.org you can ...

- Download links for 3 versions of the dataset :
 - Raw version: all distribution records with environmental data and random background data
 - Pre-processed data at 25km² Grid filtered 5 fold random cross validation
 - Pre-processed data at 25km² grid filtered 5 fold with spatial cross validation

All versions include taxonomic and traits information for all species from WoRMS.

Visualize all species and there distributions records, overlay environmental layers and ecoregions.

Applications

The benchmark dataset is primarily designed to test the efficiency of species distribution algorithms, even though its use should not be limited to SDMs. Current applications include:

- Predictor selection for different taxa
- Comparison of different methods correcting for sampling bias
- Comparison of different SDM algorithms
- Defining optimal parameter settings
- Comparing results from different papers
- Easy to access data for SDM courses
- Set up a Kaggle competition