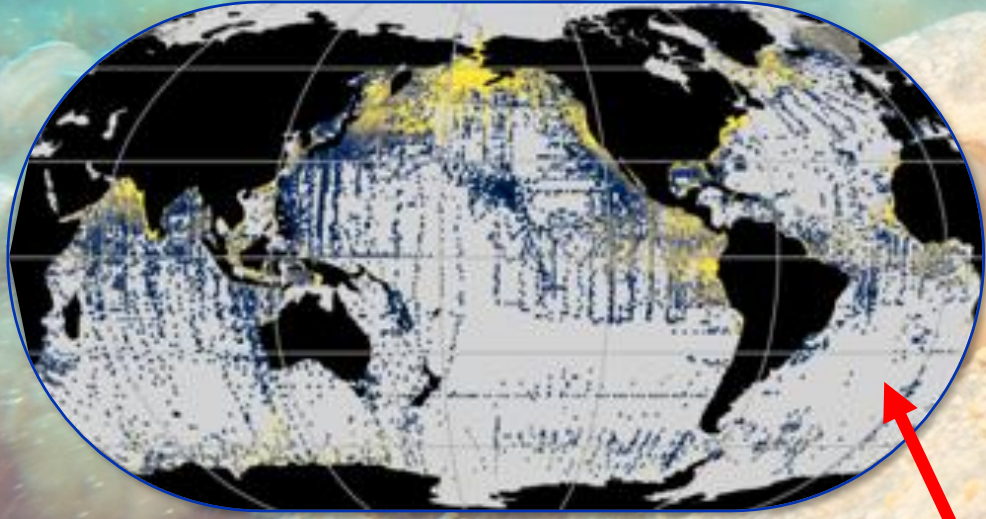


Ecosystem Workbench for Ecosystem-level EOVs

Alexandre Schickele¹, Corentin Clerc¹, Urs Hoffmann
Elizondo¹, Dominic Eriksson¹, Siyi Zhang¹, **Matthias Münnich**¹,
Jean-Olivier Irisson², Virginie Sonnet², Stéphane Pesant³, Meike
Vogt¹

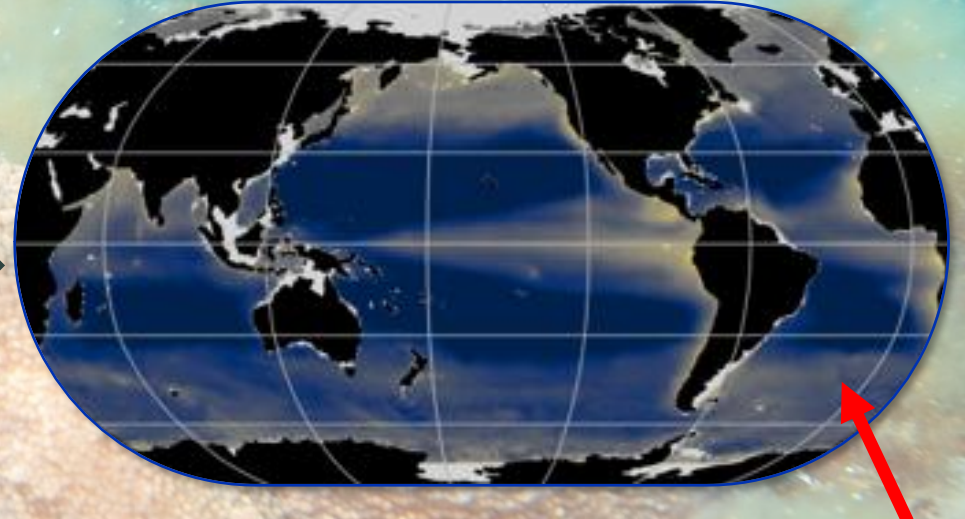
Marine Biological Observations
(here Zooplankton)

Global Map
(Zooplankton)



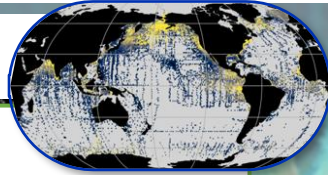
**Ecosystem
Workbench**

Habitat Modelling



?

Plankton data



Input

Ecosystem Workbench

Output



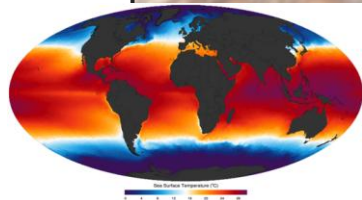
Species distribution modelling pipeline

- Comprehensive suite of methods: (GLM, GAM, SVM, MLP, RF, BRT, MBTR)
- Quality checks, uncertainty estimation

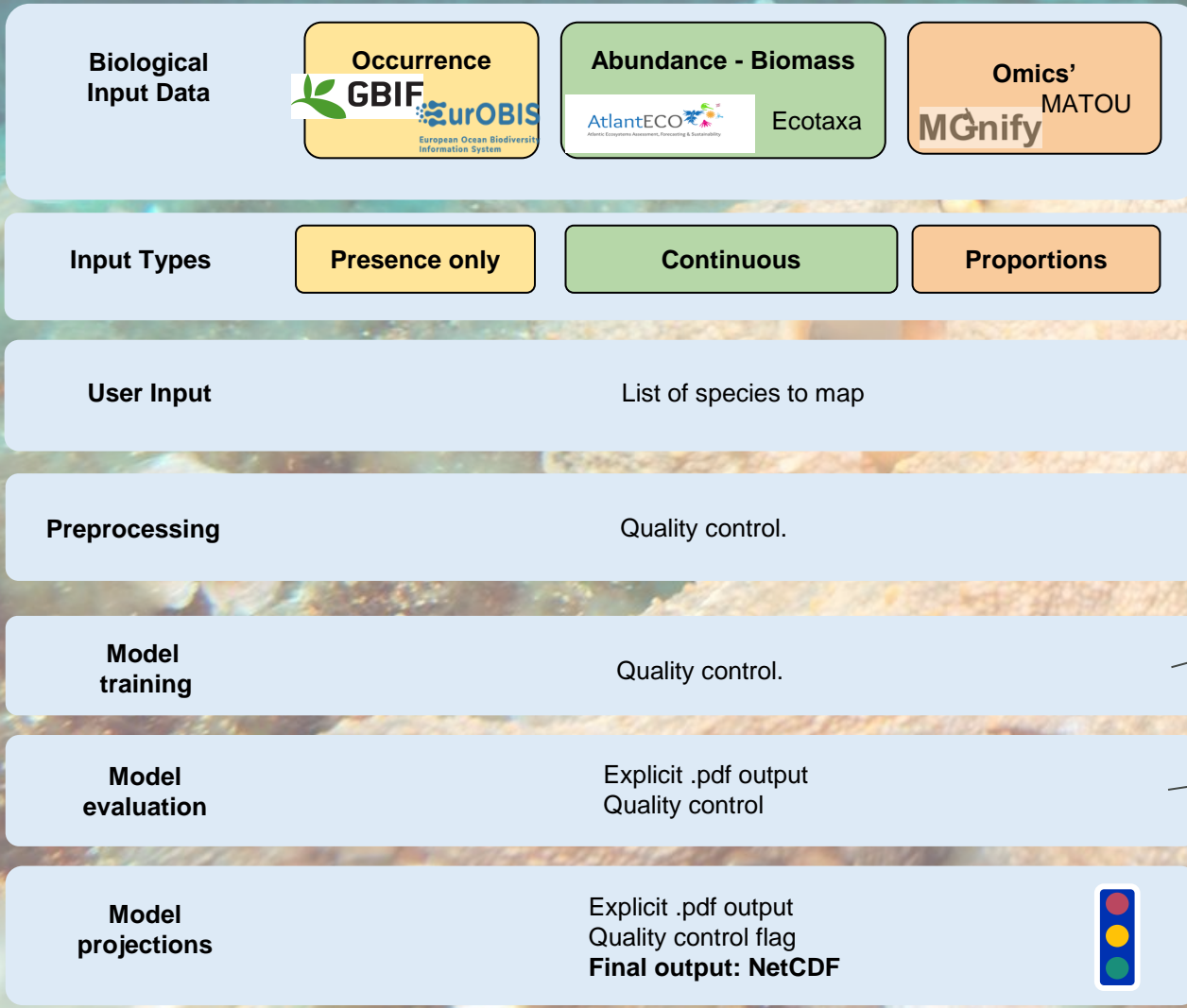
Input

Global Environmental data

Temperature, O₂, Chlorophyl, Alkalinity, EKE, HCO₃, PAR, DIC, ...



Global plankton EOVs and EBVs (NetCDF files & Summary PDF)



New data icomint in through AtlantECO



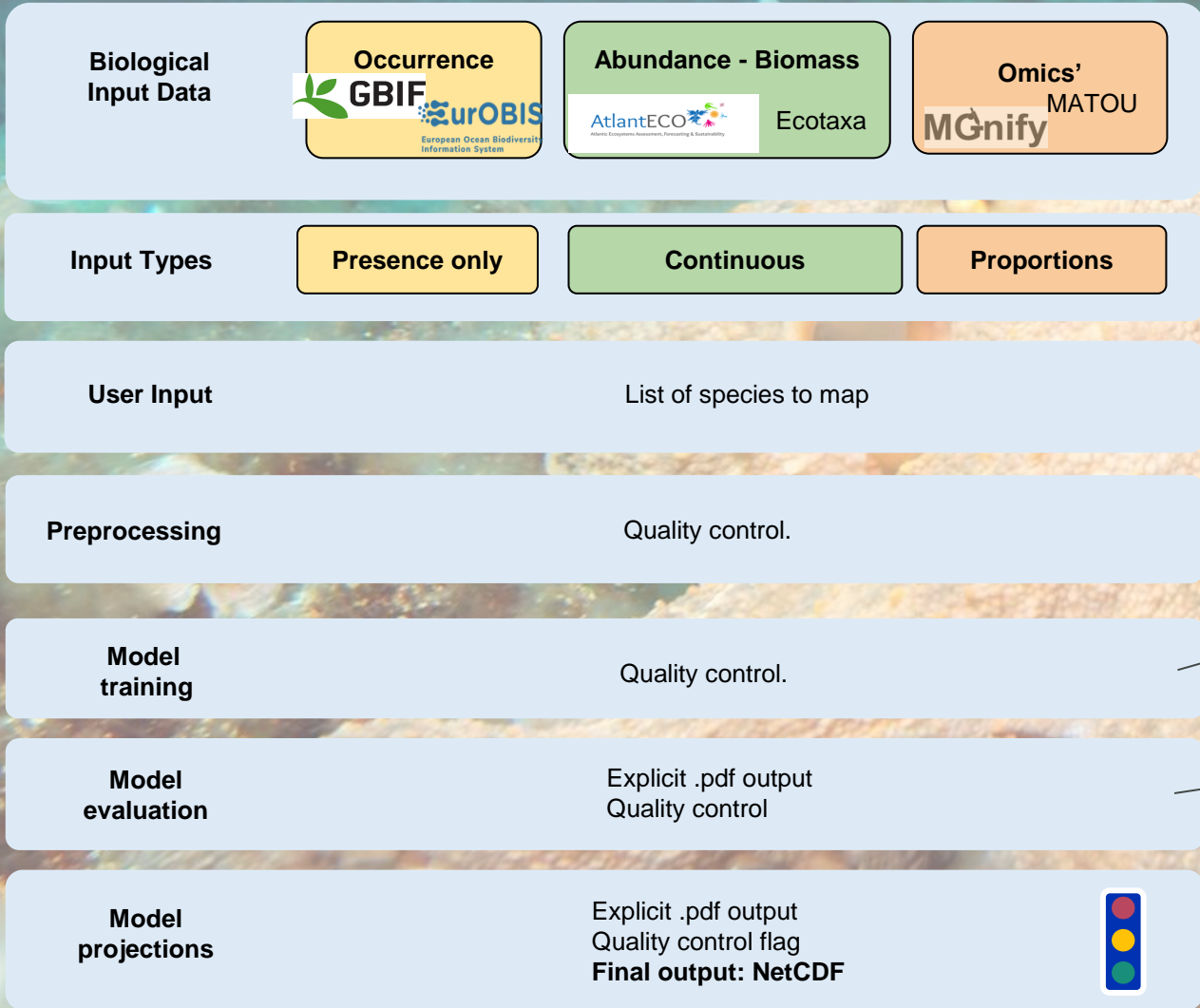
Atlantic Ecosystems Assessment, Forecasting & Sustainability



Phytoplankton
Zooplankton

Bacteria
Viruses






- Taxonomic harmonization
- Biological outliers
- Gridding
- Data biases
- Match-up with predictors
- Multicollinearity in predictors
- Predictor selection

- Hyper-parameter tuning
- Spatial autocorrelation
- Multiple models:
GLM, GAM, SVM, MLP, RF, BRT, MBTR

- Model predictive performance
- Environmental predictor importance

- Projection uncertainty
- Spatial extrapolation
- Ensemble agreement
- **Traffic light quality control** 

The screenshot displays the Blue-Cloud2026 web portal. At the top, there is a navigation bar with a search bar and a 'Go to' dropdown menu. The main content area is divided into several sections:

- Catalogue:** A search bar with the text 'Insert keywords here' and a 'Search' button. Below it, there are icons for 'Deliverable (47)', 'Guidelines (17)', and 'Service (1)'. A 'Go to' dropdown menu is open, showing options: 'Blue-Cloud2026Project', 'Blue-CloudTrainingLab', 'Ecosystem-Workbench', and 'FisheriesAtlas'.
- VLabs for Blue-Cloud...:** A section with two entries for 'Blue-Cloud2026' and 'Blue-Cloud', both marked as 'restricted'. The first entry describes support for project activities and discussions. The second entry describes support for project activities and discussions, mentioning it is equipped with the following.
- Ecosystem-Workbench:** A central section with a 'Statistics' panel showing 'Your Stats in Ecosystem-Workbench' and a 'Shared Folder' panel with a message: 'Ops, cannot reach the server. Please try to reload the page or check your internet connection'. Below the message is a table with columns 'Name', 'Owner', and 'Last modified'. A 'News feed' section shows a post by 'Meike Vogt' dated 'November 08 2023, 3:22 PM' with the text 'Welcome everyone to our ecosystem workb...'. A 'Methods' section lists 'CEPHALOPOD 1.0.3' by 'Urs Hofmann Elizondo', with sub-methods 'Species distro', 'Example', and 'Multiinstance'.
- CEPHALOPOD:** A section titled 'CEPHALOPOD pipeline prototype' with an 'Inputs' form. The form includes fields for 'Runtime' (ursho/cephalopod_image_local_test:latest), 'predictors' (https://api.d4science.org/workspace/items/4fb9be9a-130b-4), 'run_name' (cephalopod_run), and 'output_path' (https://api.d4science.org/workspace/items/371d01d8-d959-).
- Execution Monitor:** A section titled 'Execution Monitor' with a search bar and tabs for 'Live executions' and 'Archived executions'. A live execution is shown for 'CEPHALOPOD 1.0.3' with a 'successful' status, accepted on '31/10/2024 @ 13:24:27', and last updated on '31/10/2024 @ 14:34:21'.

<https://blue-cloud.d4science.org/>

Detailed summary for user

For the following run: cephalopod_run

Document description:

The present .pdf file provides detailed information related to the analysis in- and outputs. The input description displayed hereafter correspond to the user-defined parameters related to the biological sample selection criteria, the type of data used in the analysis and the choices for algorithm training. They are common to all considered species. The outputs are displayed at the species level, in form of graphical outputs for each pipeline steps, and only including those passing the embedded quality checks. The graphical outputs provide information concerning:

- The biological and environmental data location, selection, and importance in the observations.
- The environmental variable importance in the model training.
- The spatial projections and uncertainties
- The biological response to each environmental variable in form of partial dependency plots

Depending on the user selection, a selection of diversity projections across species can be displayed at the end of this document. It only considers species passing the quality checks.

Please note that the Bluecloud2026 Ecosystem Workbench development team did its best to provide a habitat modelling pipeline, including the latest advances and quality checks in this field of research, and quality assessment guidelines to a large audience. The interpretation of the results and final quality assessment is up to the user, however.

Input description:

Only contains output that passed the quality checks: TRUE

The following parameters are related to the biological sample selection:

- Minimum number of sample per species: 50
- Depth range of the samples: 0 to 200 m
- Time range of the samples: 1950 to 2020

The following parameters are related to the type of data used:

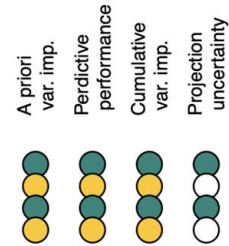
- Type of raw data query from the data access service: `./data/test_input/test_input.csv`
- Type of raw data considered in the analysis (after eventual transformation of the raw data): `presence_only`

The following parameters are related to the algorithm training:

- Number of splits for the cross-validation folds: 3
- Type of cross-validation performed: `lon`
- Hyperparameter grid size: 3
- Compute an ensemble model: TRUE

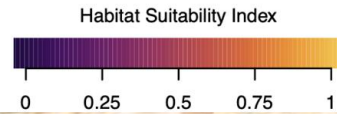
(12 pages per species)

QUALITY CHECK
Dinophysis tripos
ID: 109662

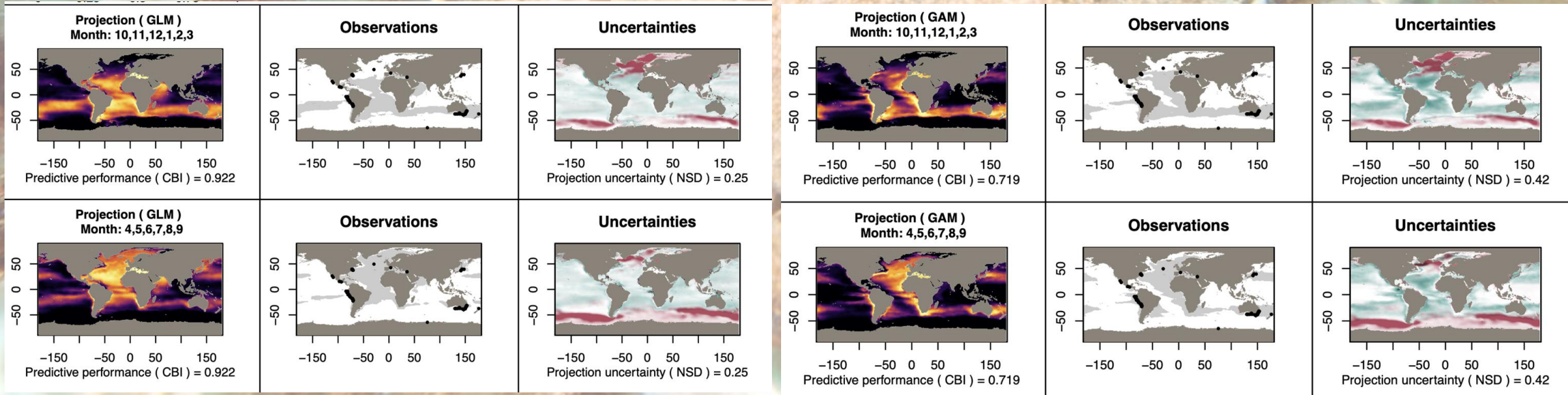
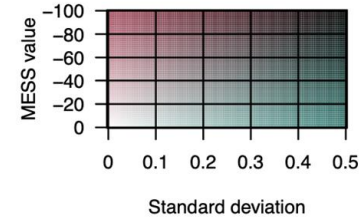


Satisfying for proposal writing
Promising but projection uncertainty is high
Satisfying for proposal writing
Promising but projection uncertainty is high

GLM, GAM: OK
MLP, BRT problematic

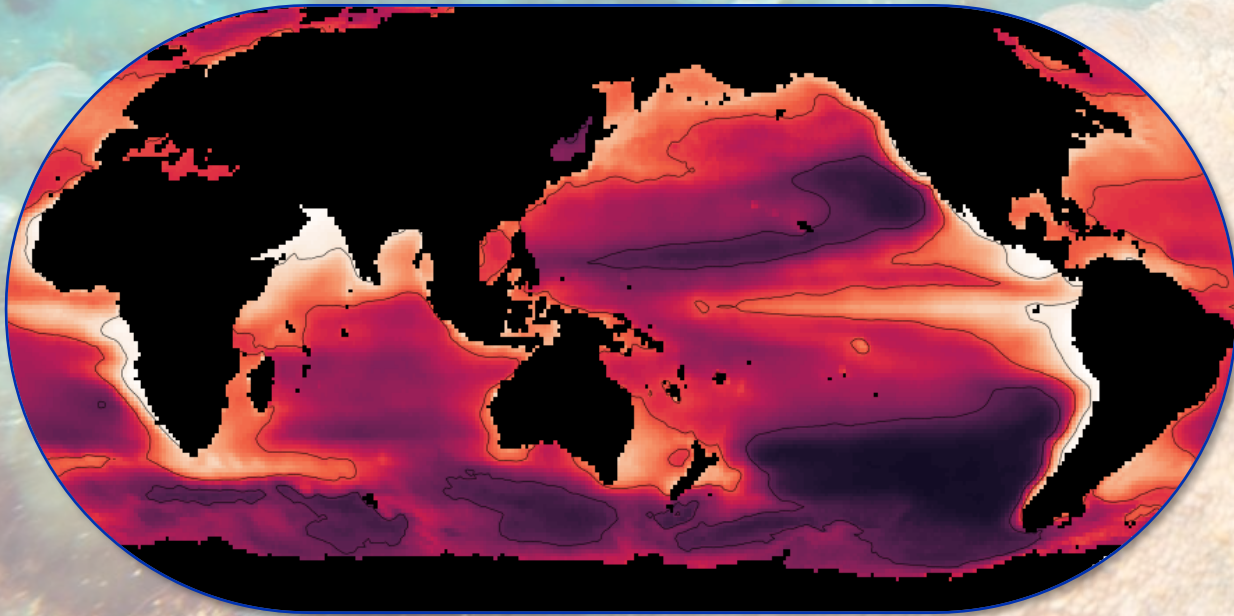


Q75 Habitat Suitability Index
Observation



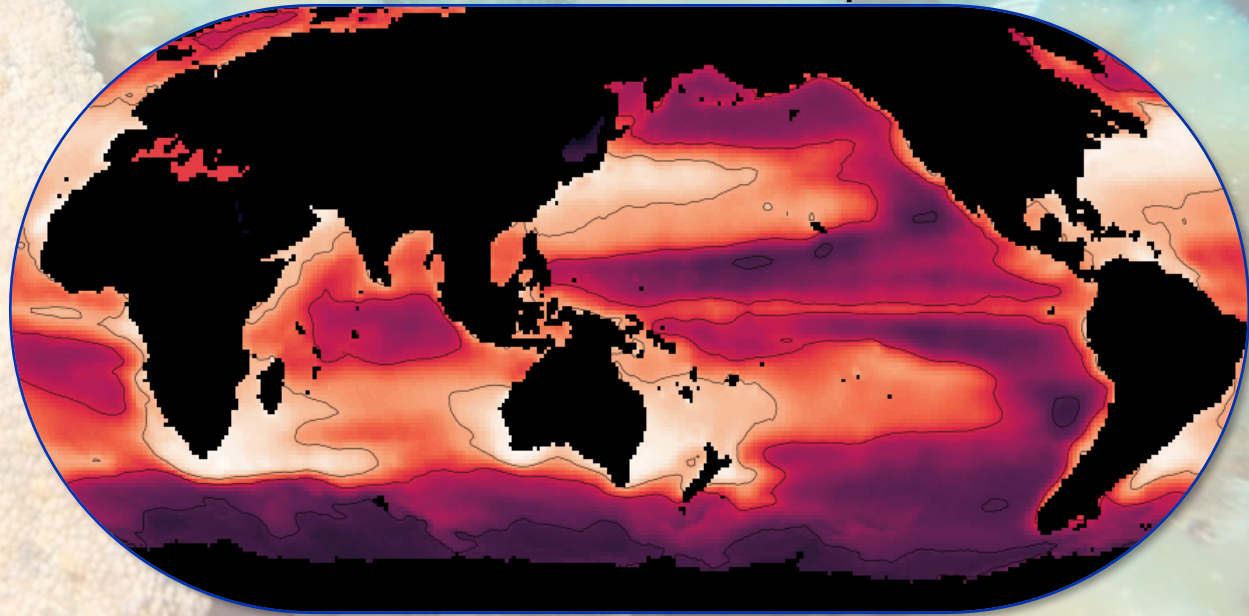
Phytoplankton

Occurrence data – 336

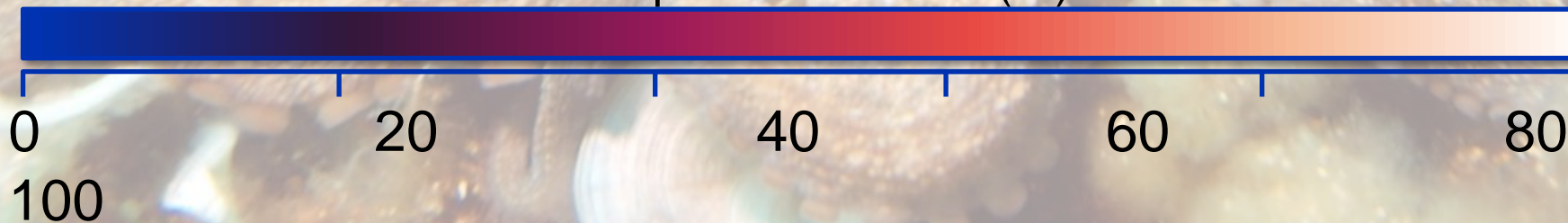


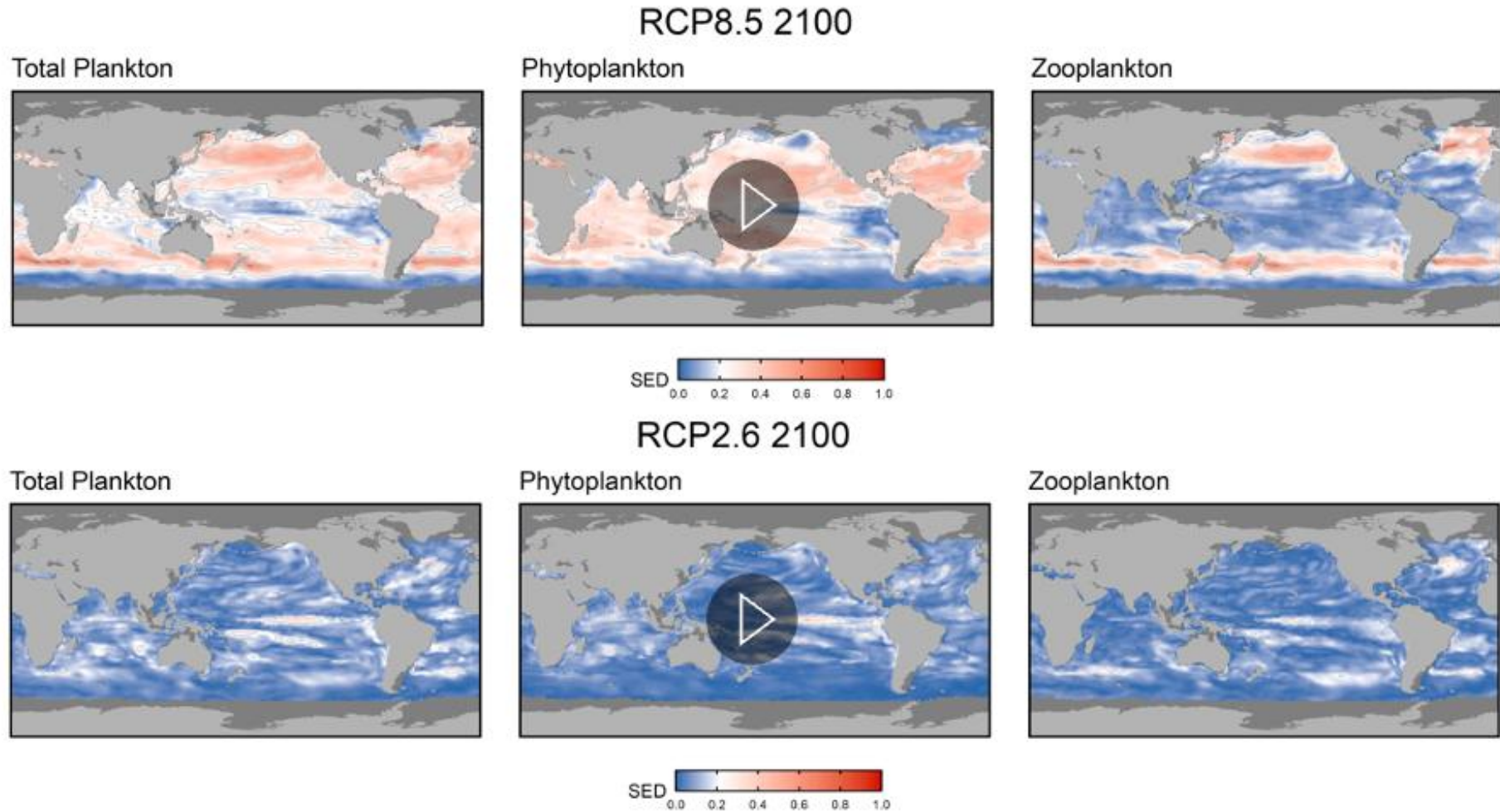
Zooplankton

Occurrence data – 524 species



Species richness (%)





eosc | Blue-Cloud2026



blue-cloud.org



[@bluecloudeu](https://twitter.com/bluecloudeu)



[blue-cloud org](https://www.linkedin.com/company/blue-cloud-org)



Funded by
the European Union