Towards transparency and reproducibility in environmental covariates for deep-sea species distribution modeling

THE UNIVERSITY OF RHODE ISLAND

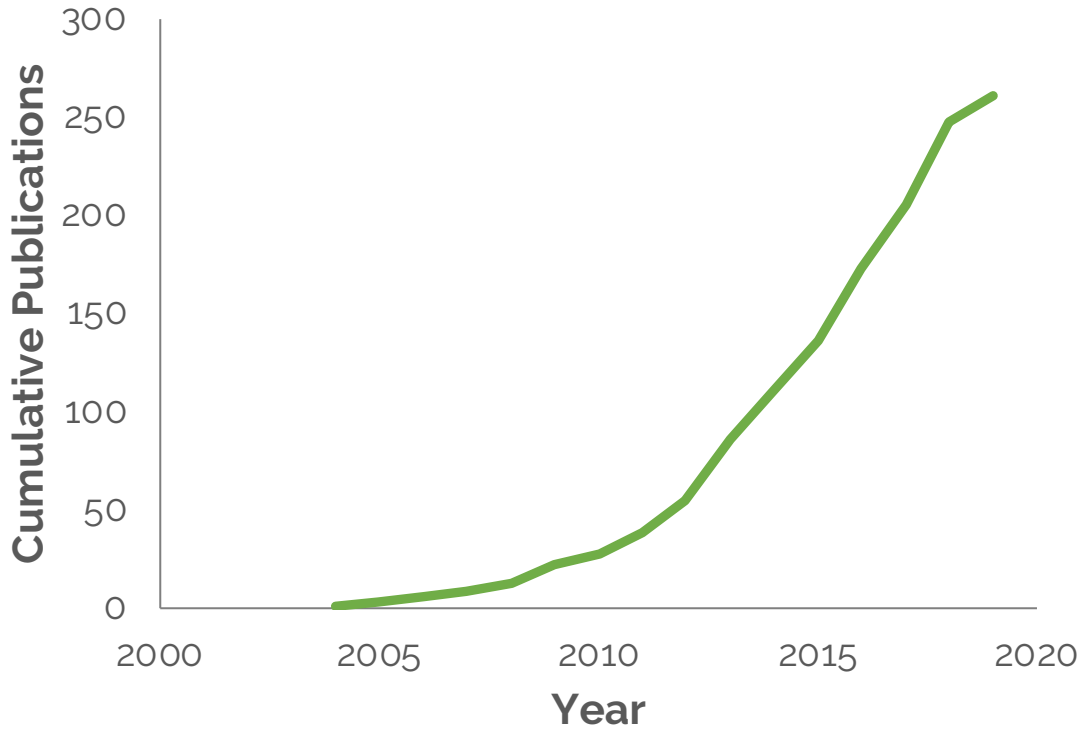
SponGES

Andrew J. Davies

Email: davies@uri.edu

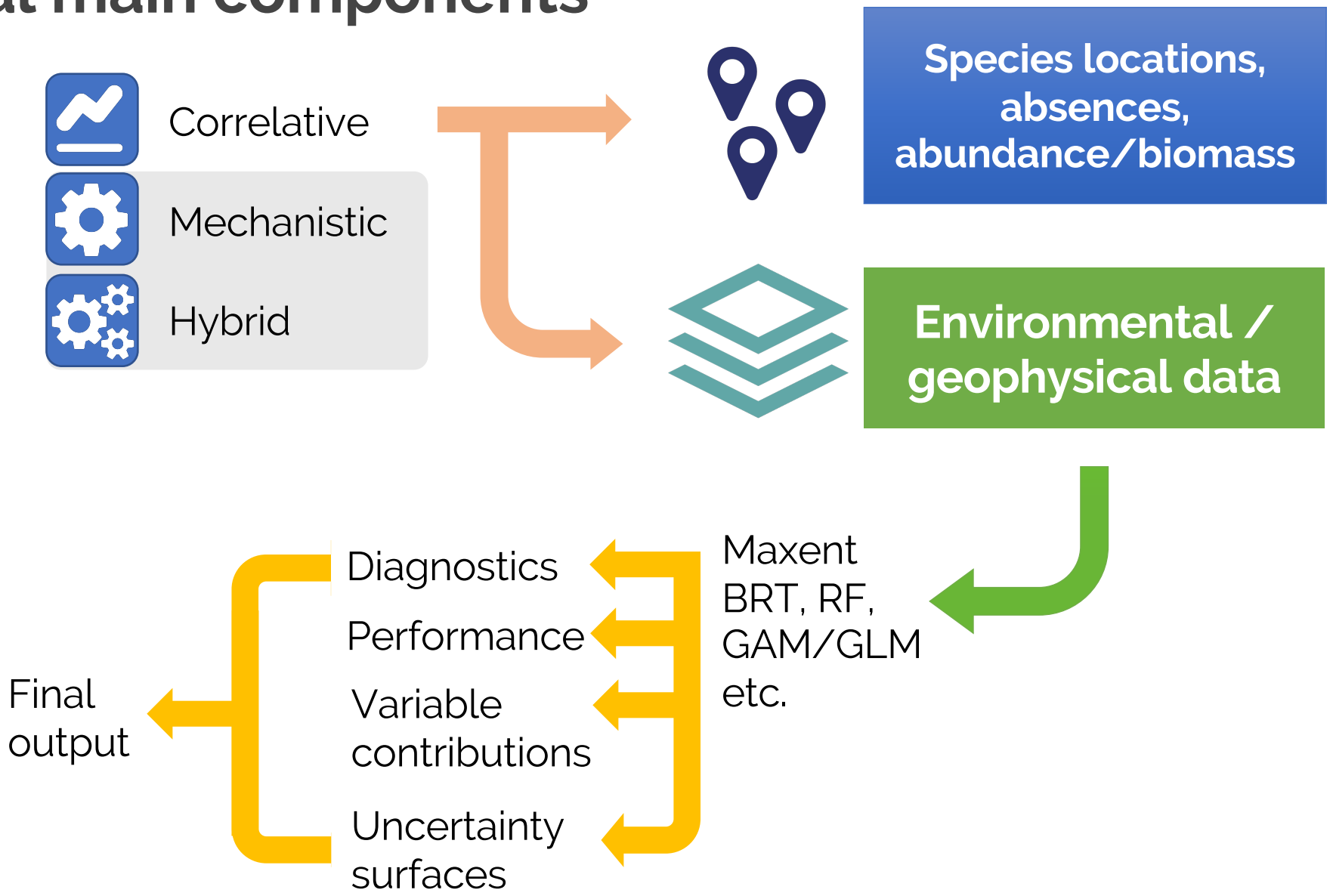
The proliferation of species distribution modeling approaches in marine science

- Species distributions
  - Management
    - Conservation
      - Niche/tolerances
        - Forecasting responses to change



ocean AND ("species distribution model" OR "habitat suitability model") from the Scopus database

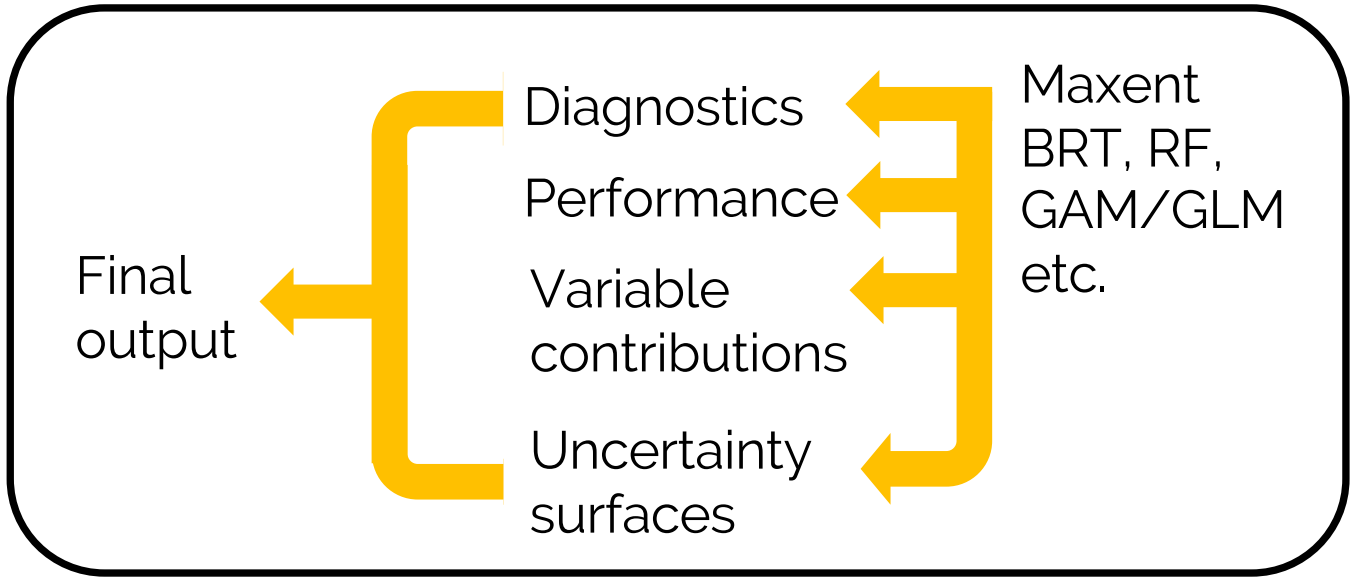
Species distribution models have several main components



Great advancements have been made in two components…

Species locations, absences, abundance/biomass

OBIS, GBIF, regional/national databases, spatial bias maps, improved pseudo-absences etc.



R, Python, Maxent, ensembles, cross validation, n-1, novel validation approaches etc.

… driven by software, innovation and open data..

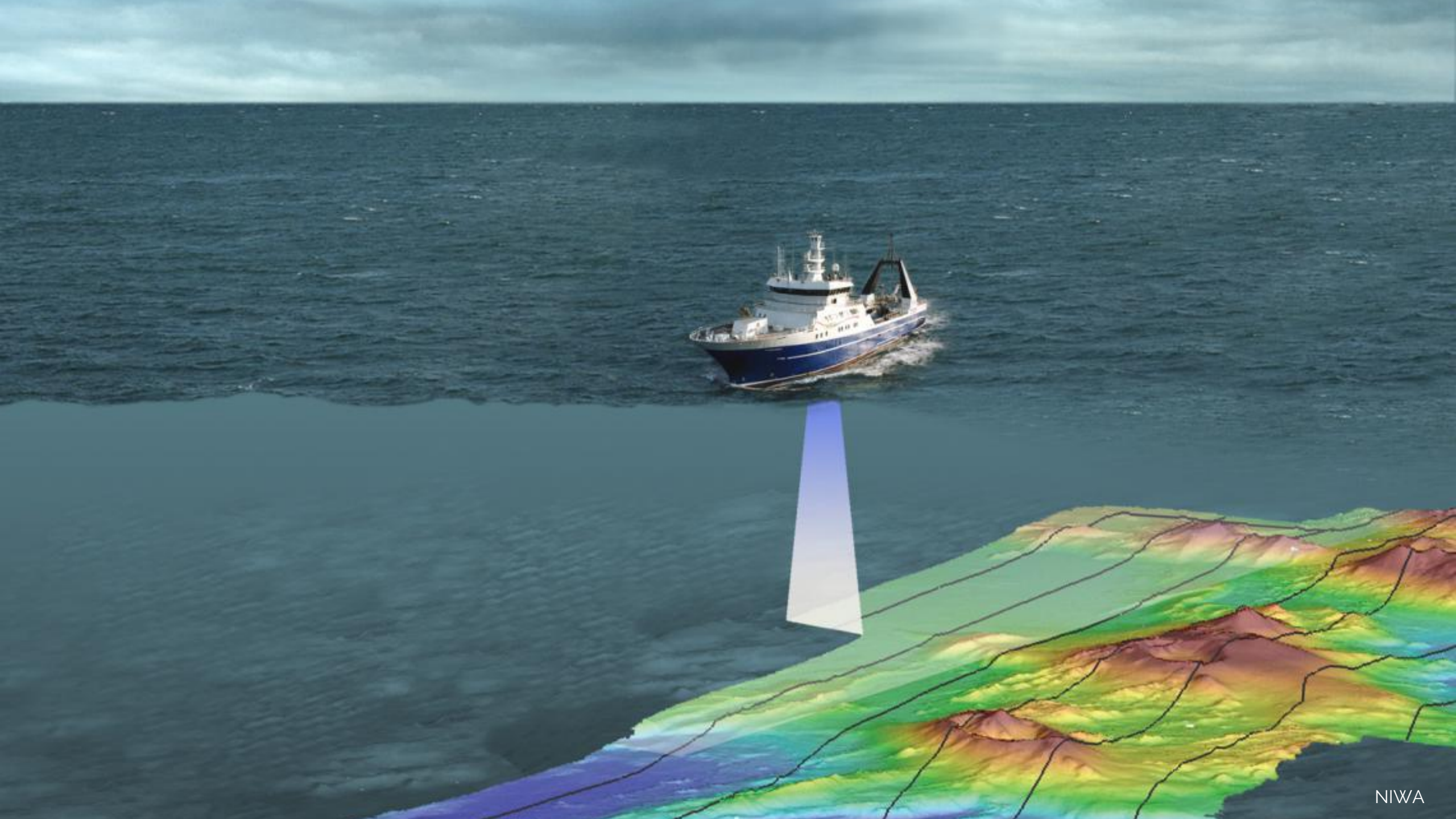
Environmental data
in contrast has somewhat
been overlooked

It is essential, a foundational aspect of the modelling process

Environmental data as a foundation

- Underpins all species distribution models.
- Errors here will introduce uncertainty into outputs
  - Sensor issues
  - Spatial inaccuracies
  - Resampling errors
  - Selection of appropriate scale
- Can introduce unquantifiable uncertainty if not addressed.

Our access to environmental data is increasingly facilitated by novel technology

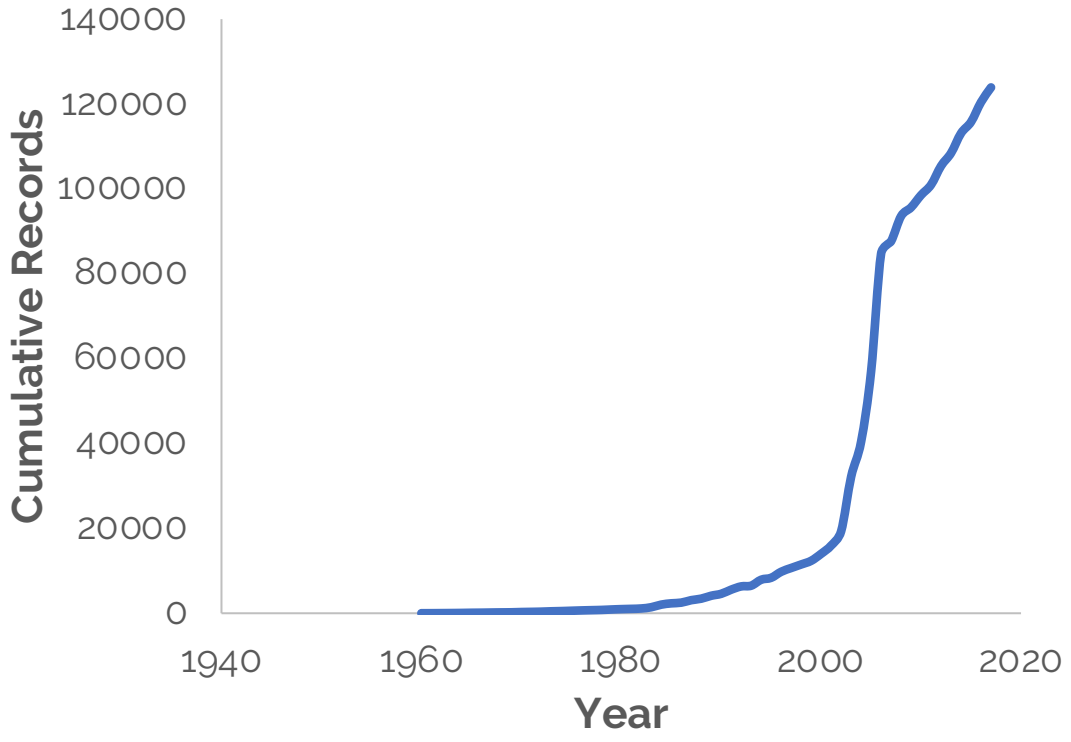

NIWA

NOAA/URI

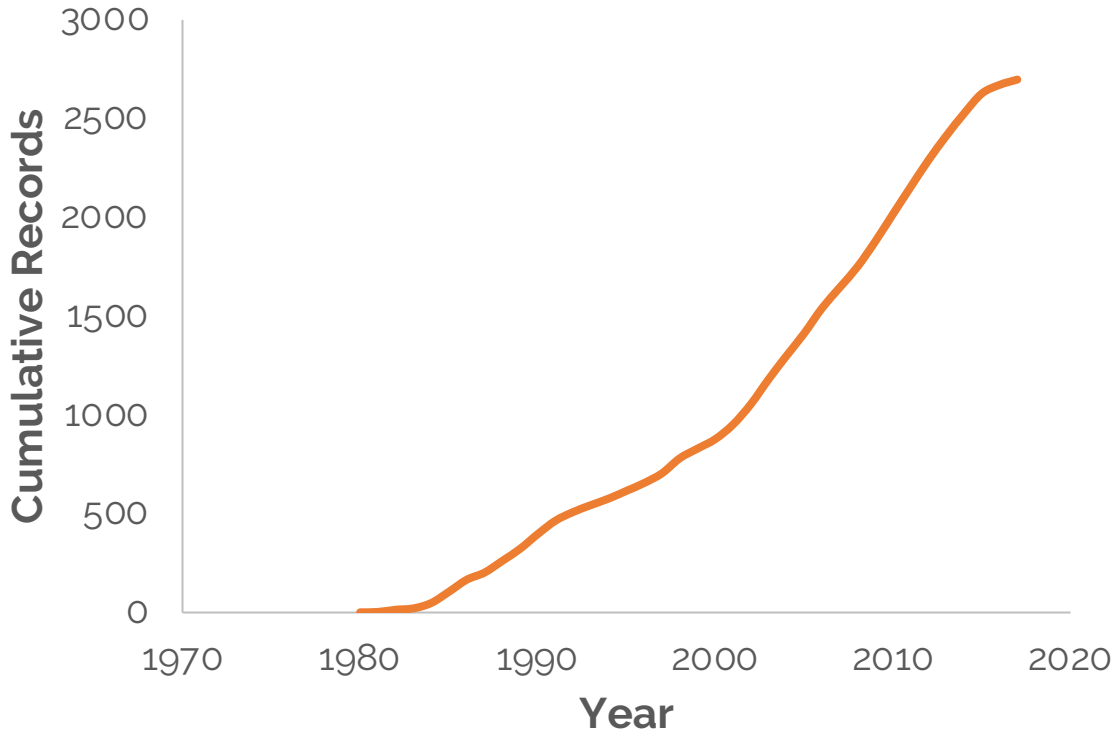
Kongsberg

Marine data is becoming available at almost exponential rates..

“Marine” Records from Pangaea.de



Multibeam data from NOAA NGDC



dbSEABED: Information Integration System for Marine Substrates

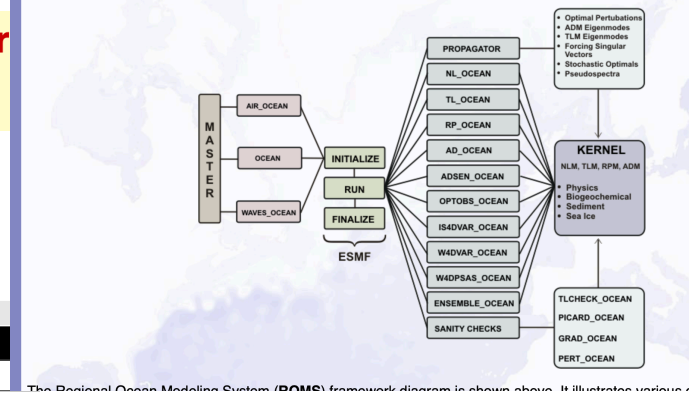
NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

formerly the National Oceanographic Data Center (NODC)... more on NCEI

NOAA Satellite and Information Service

You are here: Home > Ocean Climate Laboratory > OCL Products

Regional Ocean Modeling System (ROMS)

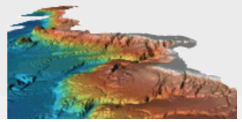


The Regional Ocean Modeling System (**ROMS**) framework diagram is shown above. It illustrates various

NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

NOAA > NESDIS > NCEI (formerly NGDC) > Marine Geology and Geophysics > Bathymetry & Relief

All Bathy/Relief | Coastal DEMs | Fishing | Global



Multibeam Bathymetry

NCEI is the U.S. national archive for multibeam bathymetric data and holds more than 9 million na
from over 2400 cruises and received from sources worldwide. In addition to deepwater data, the M
includes hydrographic multibeam survey data from NOAA's National Ocean Service (NOS).

Data Sets & Products

World Ocean Database and World Ocean Atlas Series

Note: The World Ocean Database 2018 expands (0.1

WODselect Online Data Retrieval System
The WODselect interface allows a user to search W
user-specified search criteria. A distribution map ar
will give the user the option to have the data extrac

New WORLD OCEAN DATABASE Quarterly updates and n

SATELLITE GEODESY

Scripps Institution Of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, 92093-0225

HOME
BACKGROUND
PEOPLE
GLOBAL TOPOGRAPHY
RADAR

GLOBAL TOPOGRAPHY

SRTM30_PLUS: SRTM30, COASTAL & RIDGE MULTIBEAM, ESTIMATED TOPOGRAPHY

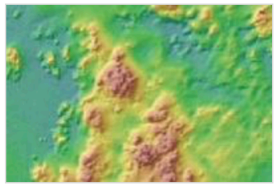
Reference, sounding data: Becker, J. J., D. T. Sandwell, W. H. F. Smith, J. Braud, B. Binder, J. Depner, D. Fabre, J. Factor, S. Ingalls, S-H. Kim, R. Ladner, K. Marks, S. Nelson, A. Pharaoh, R. Trimmer, J. Von Rosenberg, G. Wallace, P. Weatherall., Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS, Marine Geodesy, 32:4, 355-371, 2009.

Quick Contents:

ArcGIS Pricing Map Scene Help

Created November 14, 2012 at 7:18:31 PM -05

Benthic Terrain Modeler



Benthic Terrain Modeler, tools for understanding and classi
environment.

Desktop Application Template by swalbridge

Created: Nov 14, 2012 Updated: Mar 21, 2018 Number of Dov

Description

Modeler (BTM) 3.0

EMODnet BATHYMETRY
Understanding the topography of the European seas

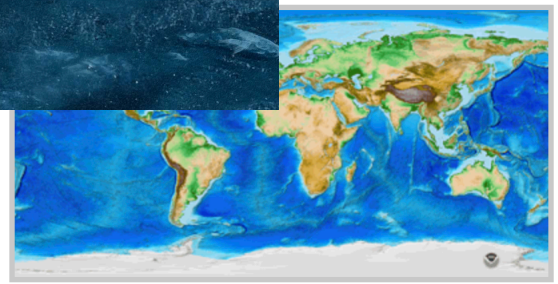
APPROACH METADATA & DATA DATA PRODUCTS NEWS PROMOT

General Bathymetric Chart of the Oceans
GEBCO

GEBCO aims to provide the most authoritative, publicly available bathymetry data sets for the world's oceans.

Download GEBCO's global grid

ETOPO1 Global Relief Model



ETOPO1 is a 1 arc-minute global relief model of Earth's surface that integrates land topography and ocean bathymetry. Built from global and regional data sets, it is available in "Ice Surface" (top of Antarctic and Greenland ice sheets) and "Bedrock" (base of the ice sheets).

ETOPO1 Global Relief Model is used to calculate the Volumes of the World's Oceans and to derive a Hypsographic Curve of Earth's Surface.

click on image for a 8.6 MB

Sounds amazing, so what is the problem?

Let me ask a question...

Can you tell me how “good” your environmental data is?

How “good” is your environmental data?

- Environmental data often undergoes several processes that are not fully reported in the methods.
  - e.g. resampling, filtering.
- Are often just “accepted as is” with little to no validation.
- Many environmental datasets are derived from more than one source, i.e. World Ocean Atlas or a hydrodynamic model, plus a bathymetric grid, which has the potential to compound errors.

80% **used none or inadequate validation for their environmental variables**

100% **did not report sufficient detail for the data to be replicated without contacting authors**

100% **did not provide access to the created data layers**

Survey of recent SDM papers published for deep-sea corals. Full meta-analysis is ongoing, so these stats are only indicative at present.

Three steps towards a “gold” standard for environmental data

1. Ensure adoption of independent validation metrics that are relevant to the modeling exercise being undertaken.
2. Improve transparency in methodologies by providing sufficient detail and/or processing code, in methodologies.
3. Increase sharing and open access of environmental data (as well as other components of models).

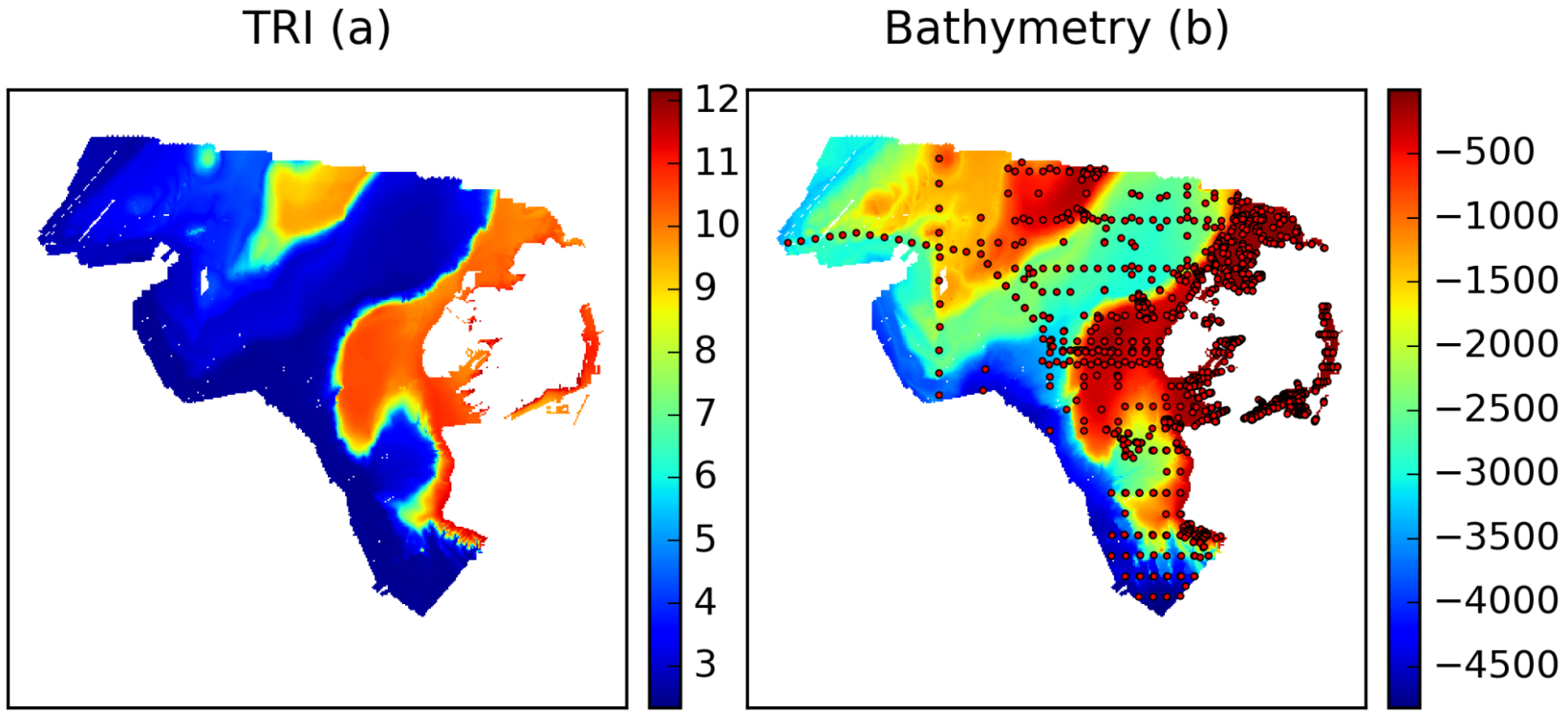
Marine Species Distribution Modeling In the Global Ocean – SCOR WG Proposal https://bit.ly/32VkSy2. Being discussed at SCOR 2019 Annual Meeting (Sept 2019).

Ensure adoption of independent validation metrics that are relevant to the modeling exercise being undertaken.

- Example of a validation process:

Overall pattern

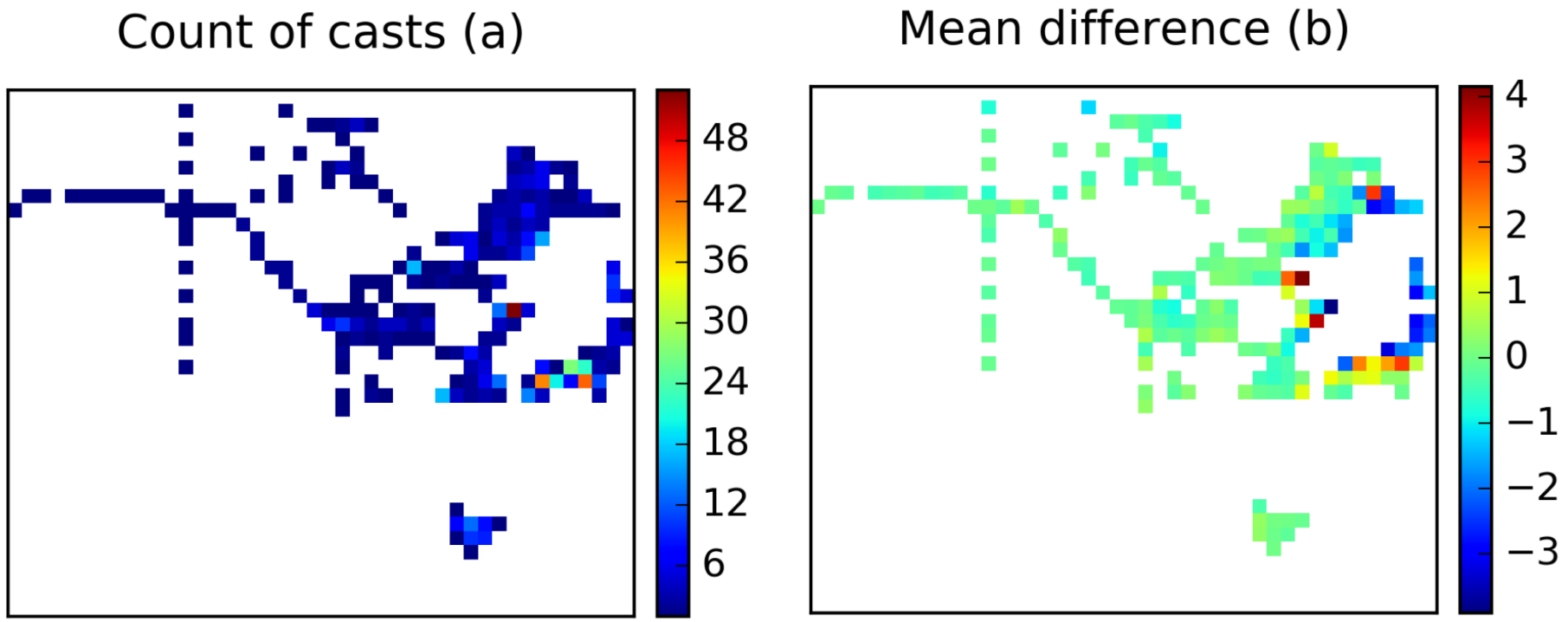
The following map shows: The output TRI variable (a), the depth layer with locations of the CTD points as red points (b). You should look at where these points fall, large clusters, especially in shallower water areas or large expanses of relatively flat bathymetry will generally explain poor model performance.



example for temperature variable from the Irish EEZ

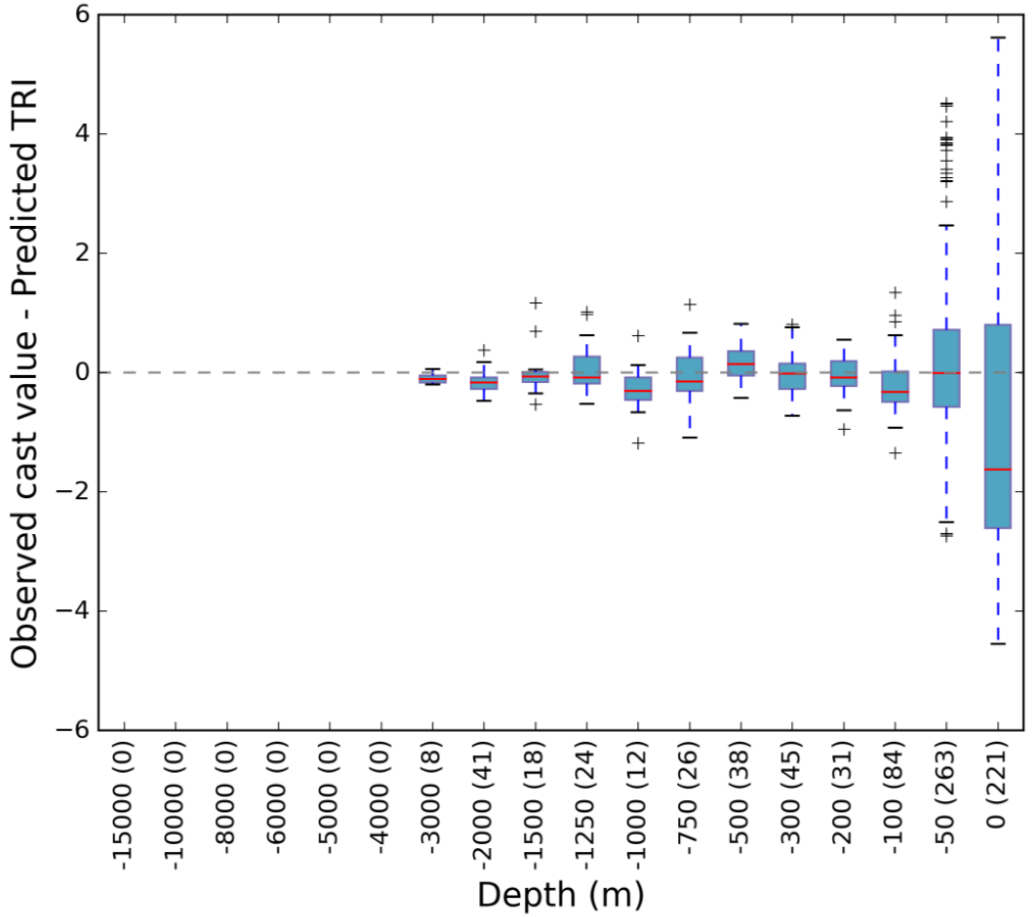
Ensure adoption of independent validation metrics that are relevant to the modeling exercise being undertaken.

- Test variables throughout modelling extent:



Ensure adoption of independent validation metrics that are relevant to the modeling exercise being undertaken.

- Throughout your depth profile:



Ensure adoption of independent validation metrics that are relevant to the modeling exercise being undertaken.

- Finally, develop comparable validation statistics:

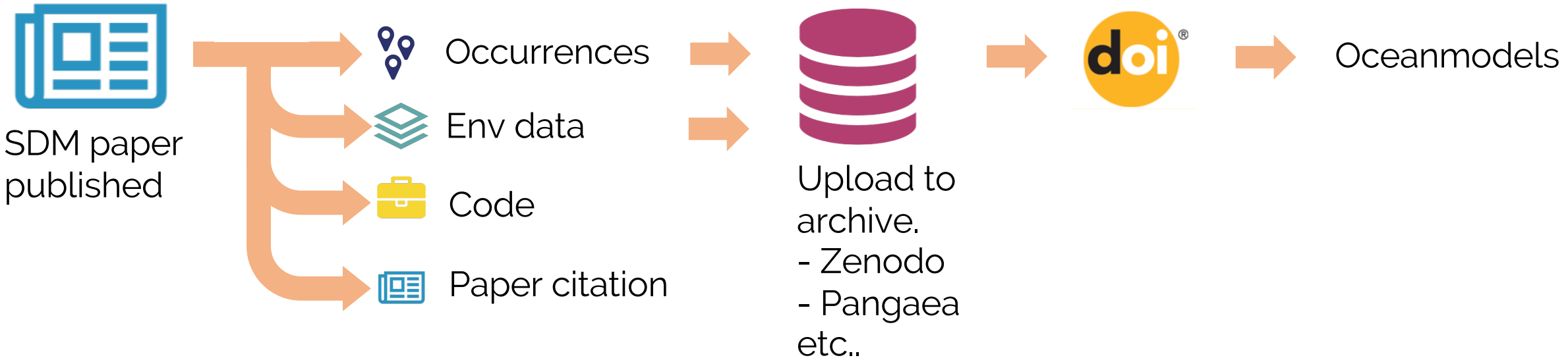
Summary statistics

The statistics provide an insight into the overall performance of the model against validation data that you downloaded from World Ocean database.

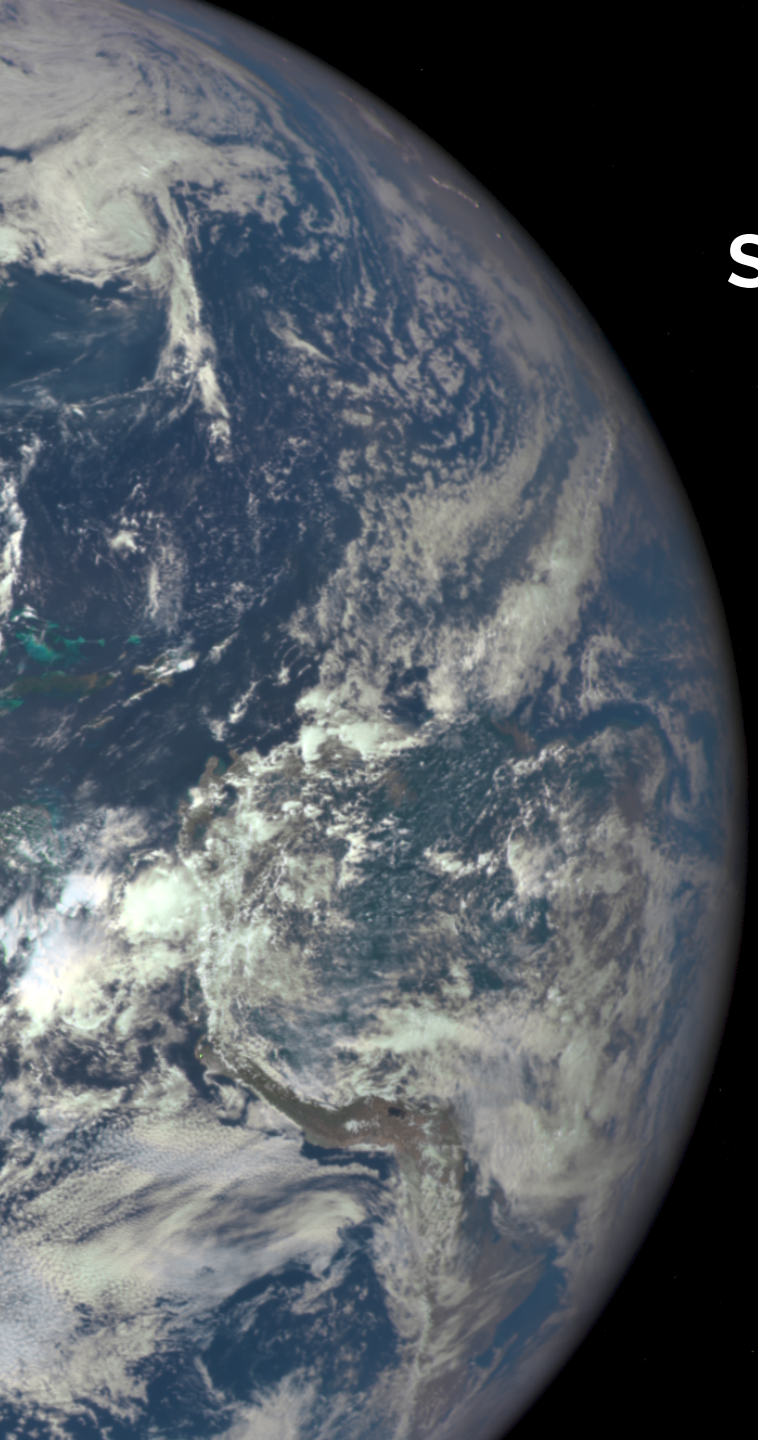
Test	Unfiltered	Filtered	Notes
Samples	1297	811	The number of CTD casts used for validation.
RMSE (Depth)	883.012167985	27.7199574441	Root mean square error, this is overall error of the depth layer.
RMSE (TRI)	3.00182317095	1.70534976079	Root mean square error, this is overall error of the tri layer.
Mean difference (Cast - TRI)	0.757856927052	-0.167957865214	Mean overall difference between validation cast and the modelled TRI layer.
Pearson Correlation (Cast vs TRI)	0.44 (p = 0.0)	0.78 (p = 0.0)	Pearson correlation of the validation casts against modelled TRI per location.

Improve transparency and increase sharing and open access of environmental data (as well as other components of models).

- Development of a new online meta-resource: oceanmodels.org*



*not yet released



Summary

- To advance species distribution models in the deep ocean:
  - Develop and apply a coherent **validation framework** for produced environmental data, and to establish a minimum standard validation protocol.
  - Follow **Open Access** and **Open Data** standards, with full releases of various programs and data.
- If you are interested in the developing oceanmodels.org framework and/or the M-SDM GO working group, drop me an email or talk to me in the last hours of the meeting.

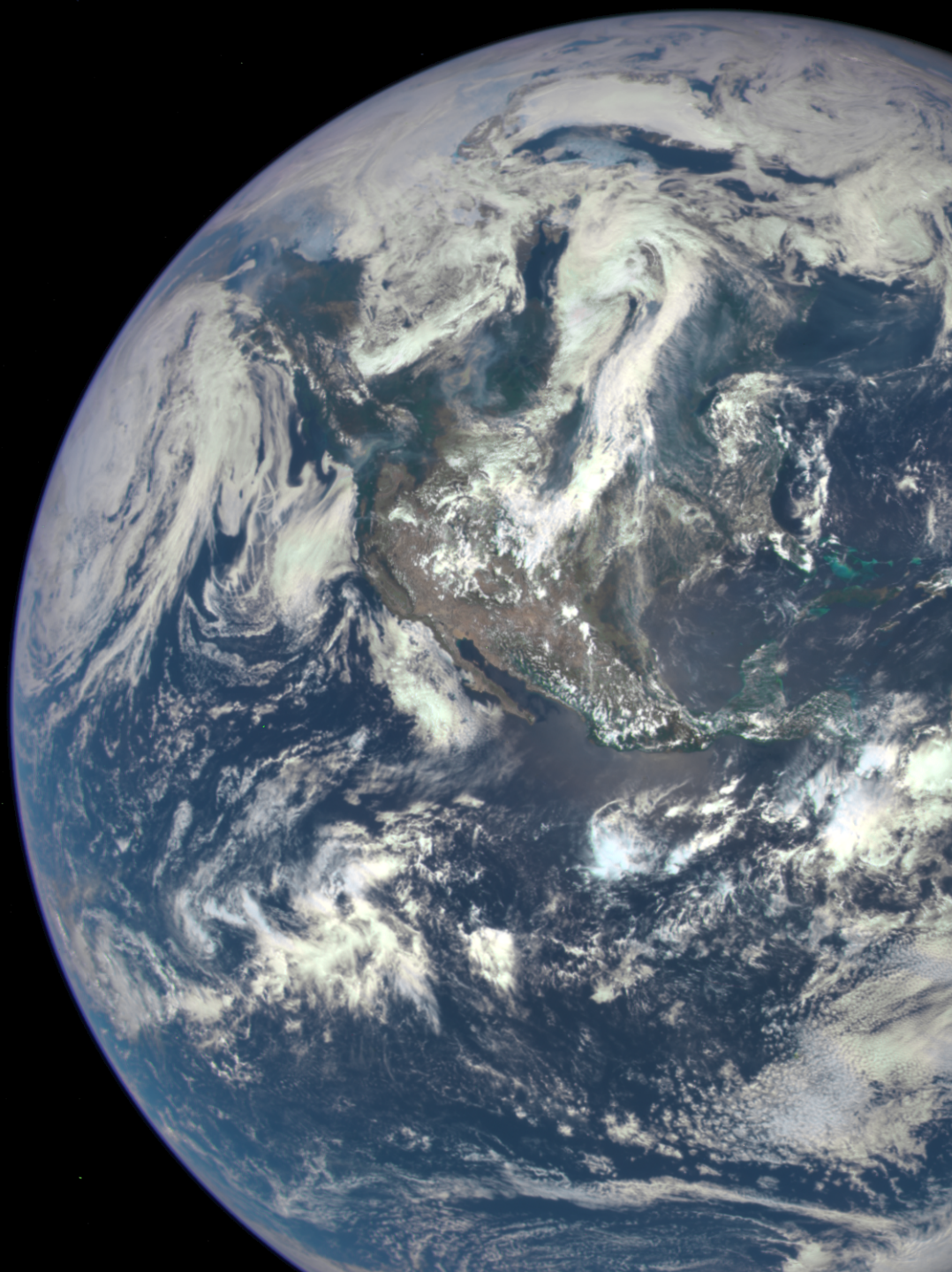
Finally

- We have a relevant session at the AGU Ocean Sciences Meeting 2020, which will take place on 16-21 February in San Diego, California.

    - Session ID: 85248
    - Session Title: **Biological hotspots in the deep sea: natural variability and adaptation to changing oceans**
    - Topic Area: Ocean Biology and Biogeochemistry

- **Abstract submissions are open will close 11 September 2019.**

- Session Chairs: Furu Mienis, Hans Tore Rapp, Jasper de Goeij, Andy Davies



Towards transparency and reproducibility in environmental covariates for deep-sea species distribution modeling

THE UNIVERSITY OF RHODE ISLAND

SponGES

Email: davies@uri.edu

This research has been performed in the scope of the SponGES project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 679849.