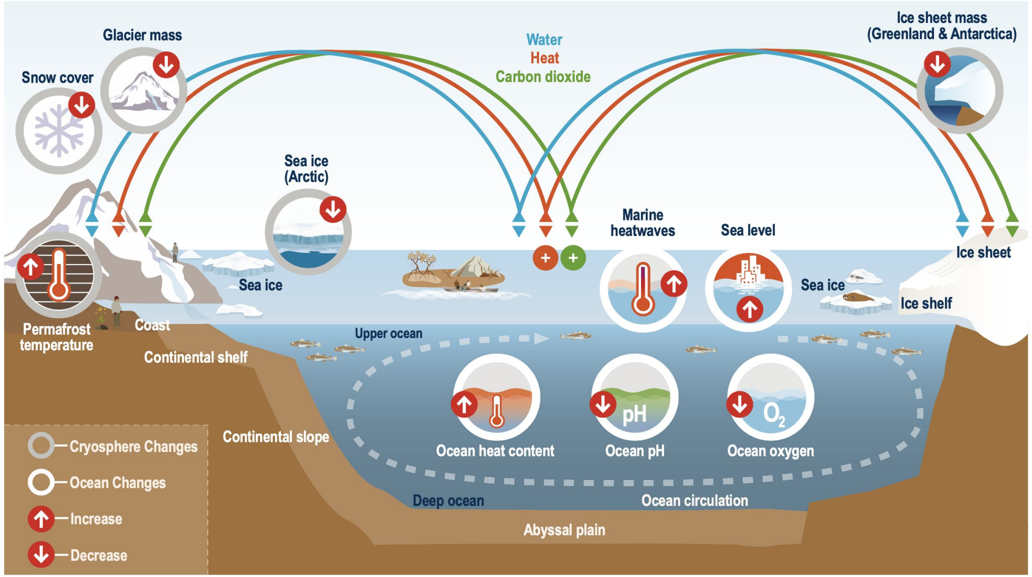
Artificial Intelligence, Machine Learning and Modeling for Understanding the Oceans and Climate Change

Nayat Sánchez Pi, Luis Martí, Andre Abreu, Olivier Bernard, Colomban de Vargas, Damien Eveillard, Alejandro Maass, Pablo A. Marquet, Jacques Sainte-Marie, Julien Salomon, Marc Schoenauer and Michèle Sebag

NEURAL INFORMATION PROCESSING SYSTEMS
Climate Change AI
December 11th, 2020

Climate Change and the Ocean



Source: H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, E. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)]. In: *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* [H.- O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.).

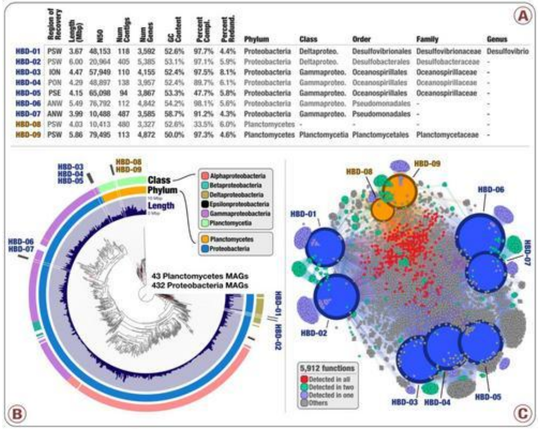
OcéanIA

An opportunity and a challenge to state-of-the-art AI/ML

CHALLENGES

- Small & high dimensional data
- Data Engineering & HPC
- Transfer Learning
- Multiobjective decision making
- Data & AI Governance
- Interpretability Explainable AI
- Causal inference
- Structure and graph-based NN
- Energy-aware Green AI/ML

Small but high-dimensional heterogeneous data



Source: Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T., Rappé, M. S., MacLellan, S. L., Lücker, S., & Eren, A. M. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nature Microbiology, 3(7), 804–813. https://doi.org/10.1038/s41564-018-0176-9

- ➜ Tara expeditions gather lots of data - from a marine biology point of view.
- ➜ Highly heterogeneous: DNA barcodes, images, environment variables.
- ➜ Samples includes many species at the same time.
- ➜ Grouped by sample location, interested in networks and graph-based information.
- ➜ In spite of efforts it is not always consistent and it is always evolving: i.e. new hardware.
- ➜ Data from Tara allows exploration of the relationship between marine ecosystem functioning and biodiversity.

Model reuse, transfer learning and domain adaptation

General-domain public model

↓

Domain transfer learning

↓

Intra-domain transfer learning

↓

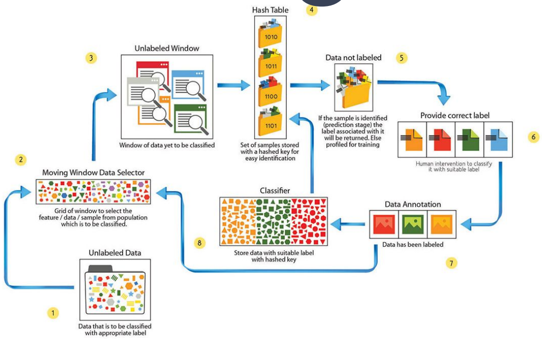
Domain adaptation

A three-step work hypothesis:

- ↳ Is it possible to adapt existing (computer vision or graph) models to out domain?
- ↳ How to transfer a domain model to other biomes, locations or across species?
- ↳ How to cope with variations across species and sensing hardware?

- **Transfer learning** addresses the issue of how to adapt and re-purpose the internal representations of a model that has been trained on a similar problem.
- **Domain adaptation** is the capacity to cope with changes in the environment because of the evolution of the system and/or the need to particularize a general model to a particular instance.

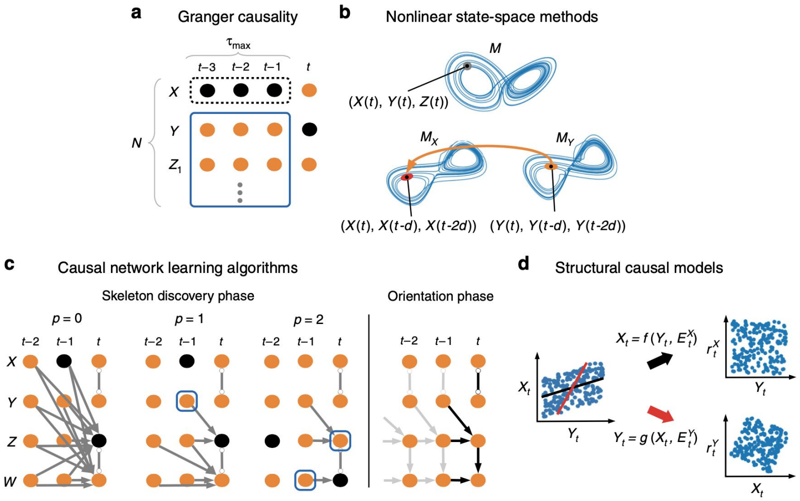
Active, few-shot and multi-task learning



Source: https://brandidea.ai/activeLearning.html

- ➜ Limited data and/or high uncertainty,
- ➜ Direct sampling to the areas of the domain where they are most necessary.
- ➜ Guiding sampling using **active learning.**
- ➜ **Few-shot learning** methods to produce actionable products with minimal data.
- ➜ **multi-source** or **multi-task learning** ensembles training signals of related tasks.
  - ↳ enables the model to generalize better on the main task.
  - ↳ effectively increases the sample size that is being used for training.
  - ↳ biases the model to prefer representations that are useful for other tasks -> transfer learning!

Causality and Explainable AI

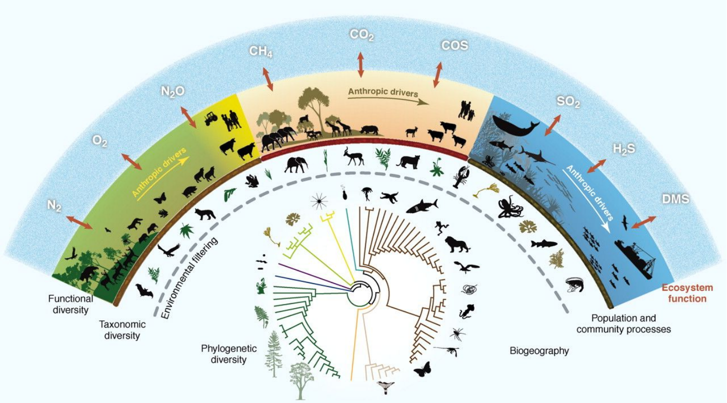


Source: Runge, J., Bathiany, S., Bollt, E. et al. Inferring causation from time series in Earth system sciences. Nat Commun 10, 2553 (2019). https://doi.org/10.1038/s41467-019-10105-3.

- ➜ Produce explainable models, while maintaining performance (prediction accuracy),
- ➜ research support tools that combine explainability and causality for new scientific discoveries and theories by making surrogate human-readable models, and
- ➜ enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

Essential for policy making.

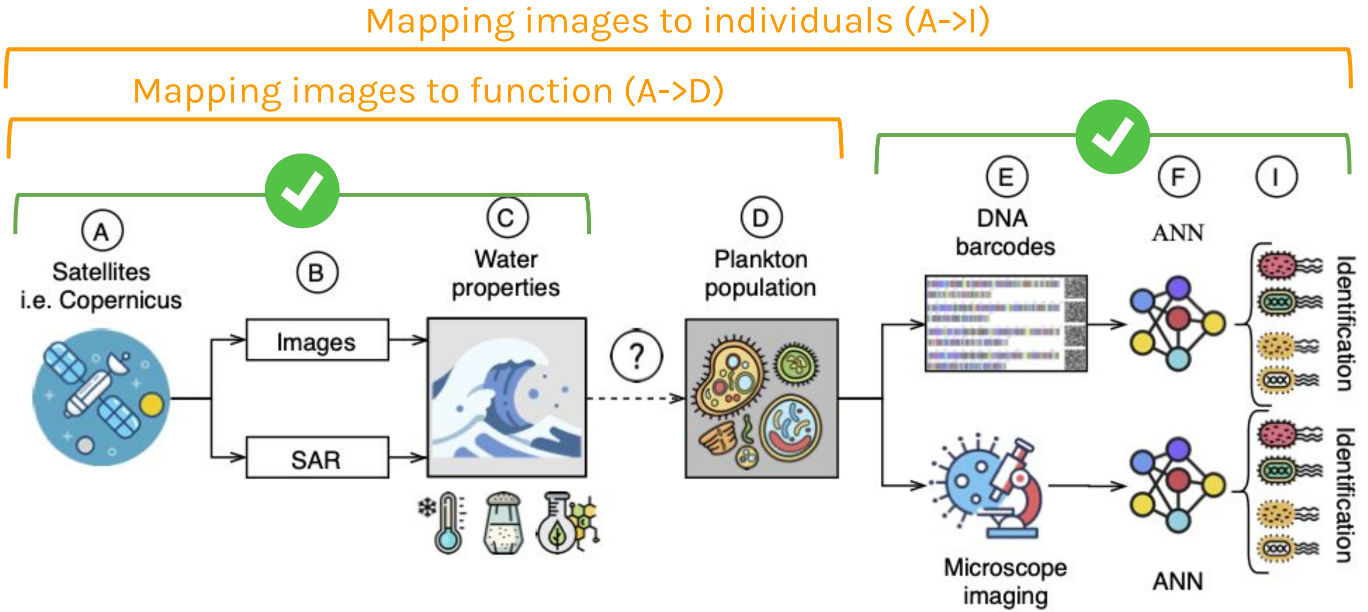
Biodiversity and ecosystem functioning



Source: Naeem, S., Duffy, J. E., & Zavaleta, E. (2012). The Functions of Biological Diversity in an Age of Extinction. *Science*, 336(6087), 1401 LP – 1406. https://doi.org/10.1126/science.1215855

- ➜ Biodiversity supports functions like primary productivity and carbon fixation and sequestration, etc.
- ➜ Understanding this is fundamental: science and policy making.
- ➜ Data from Tara allows exploration of the relationship between marine ecosystem functioning and biodiversity.
- ➜ How variations on biodiversity impact those functions?
- ➜ How changes in temperature (or other variables) impact biodiversity and functions?
- ➜ Understand causality and circular causality among different levels of biodiversity, ecosystem functioning.

Understanding plankton communities using AI, ML, and vision



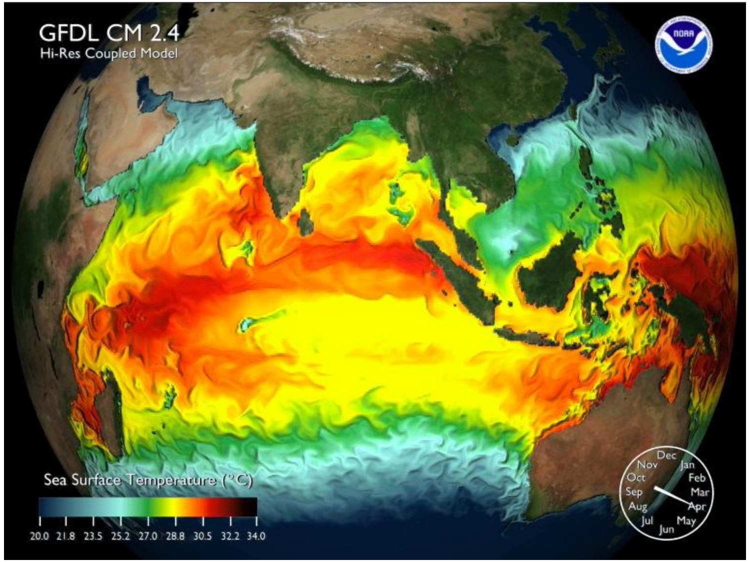
Anomaly detection and explainable AI for automatic plankton ID

Source: Wikipedia

- ➜ Identifying plankton as a supervised problem: already addressed.
- ➜ Estimated that more than 70,000 species unknown.
  - ↳ unsupervised or active learning approaches.
- ➜ Tara sampling includes high-res microscope.
- ➜ Identify **unknown** or **out of context** species.
- ➜ Why an organism represents an interesting specimen?
- ➜ Methods involved:
  - ↳ **transfer learning** and **domain adaptation**,
  - ↳ **(un/self)supervised object detection** and **segmentation**
  - ↳ **causal inference** to understand context.
  - ↳ **explainable AI**: i. e. hint what parts of the organism that determining its selection.

Field experts want a research tool not a black-box.

Integrating model-based (i.e. PDEs) and data-based (i.e. ML) approaches



Source: https://www.gfdl.noaa.gov/hires_indian_sst/

Ocean Circulation Models studying dynamics:

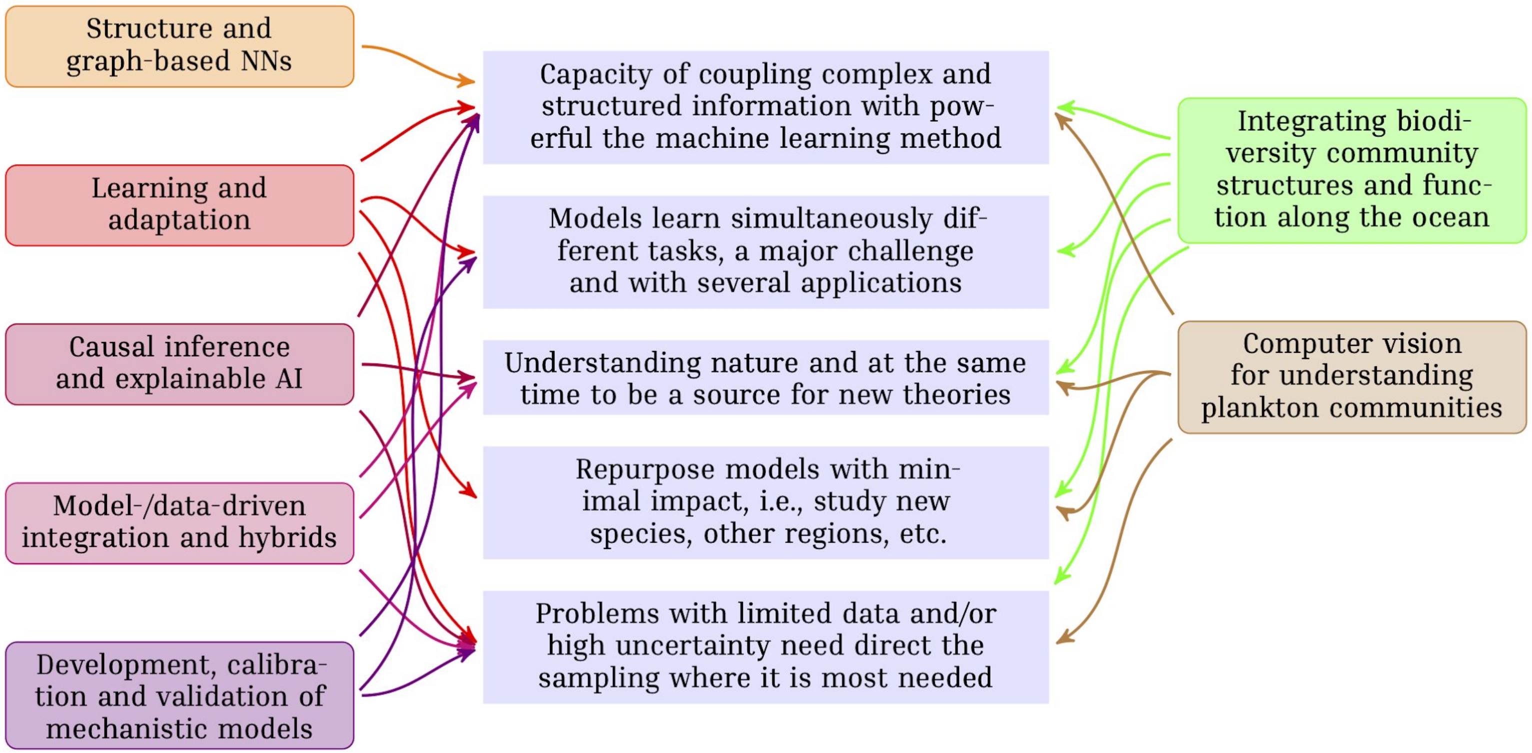
➜ high-res models are computationally inviable,

➜ current resolution of models is not sufficient,

➜ requiere large viscosity and diffusion coefficients smooths out features such as jets and mesoscale eddies.

➜ oceanic turbulence at small scales, which play an important role.

➜ Planned actions:
- ↳ Learning PDEs from Data
- ↳ Understanding learning dynamics
- ↳ Hybrid models: Combining PDE solvers and DNNs (improved explicability).


Structure and graph-based NNs
Learning and adaptation
Causal inference and explainable AI
Model-/data-driven integration and hybrids
Development, calibration and validation of mechanistic models
Capacity of coupling complex and structured information with powerful the machine learning method
Models learn simultaneously different tasks, a major challenge and with several applications
Understanding nature and at the same time to be a source for new theories
Repurpose models with minimal impact, i.e., study new species, other regions, etc.
Problems with limited data and/or high uncertainty need direct the sampling where it is most needed
Integrating biodiversity community structures and function along the ocean
Computer vision for understanding plankton communities

Summary

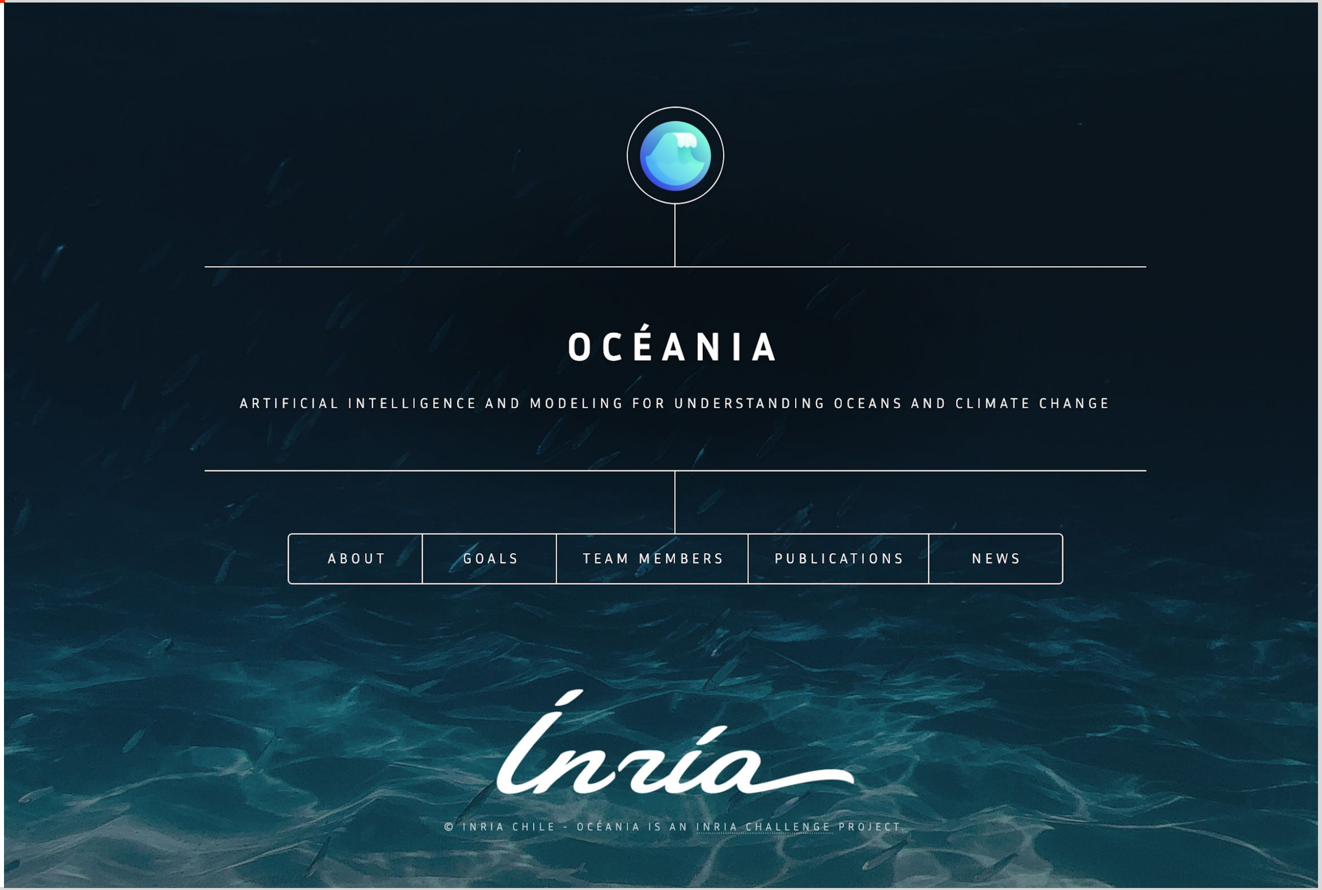
- Oceans and climate change are intimately related:
  - Carbon capture, impact of change of temperatures, etc.
- Oceans are the last 'unknown':
  - Understanding the role of oceans in climate change is not only important but also challenging for modern AI, ML and applied math.
- A way to address current hot topics like explainability, bias, etc. on a different domain.
- Open data sources can be crossed.
- **We are hiring!**

Inria Challenge

OcéanIA:

Artificial Intelligence and Modelling for Understanding Climate Change

https://oceania.inria.cl/



Thank you! Merci ! Obrigada! ¡Gracias!

Follow us: www.inria.cl and oceania.inria.cl