Observational campaigns in the age of AI
Introducing ORCAS

Clare Eayrs, Lorenzo Zampieri, and the PCAPS ORCAS Task Team

What observations do AI-based sea ice forecasts need?

Second Observational Campaigns workshop for better forecasts
ECMWF, Reading 3 July 2026

International Science Council
SCOR
Scientific Committee on Oceanic Research

WORLD METEOROLOGICAL ORGANIZATION

WWRP
WORLD WEATHER RESEARCH PROGRAMME

PCAPS
POLAR COUPLED ANALYSIS AND PREDICTION FOR SERVICES

Sea ice remains one of the central challenges in polar forecasting

Heinrich et al., 2024
Weather, Climate, and Society 16, 3; 10.1175/WCAS-D-23-0105.1

Word frequency count examining title, abstract, and keywords from the literature review.

Sea ice is the dominant focus of polar forecasting research and services

88% of publications by year during the PPP decade had an Arctic focus

Yet most polar users do not have access to tailored forecast products

Forecast systems evolve when observations evolve

WAVE 1:

→ **limited measurements**
→ **physical state variables**
→ **models describe sea ice evolution**

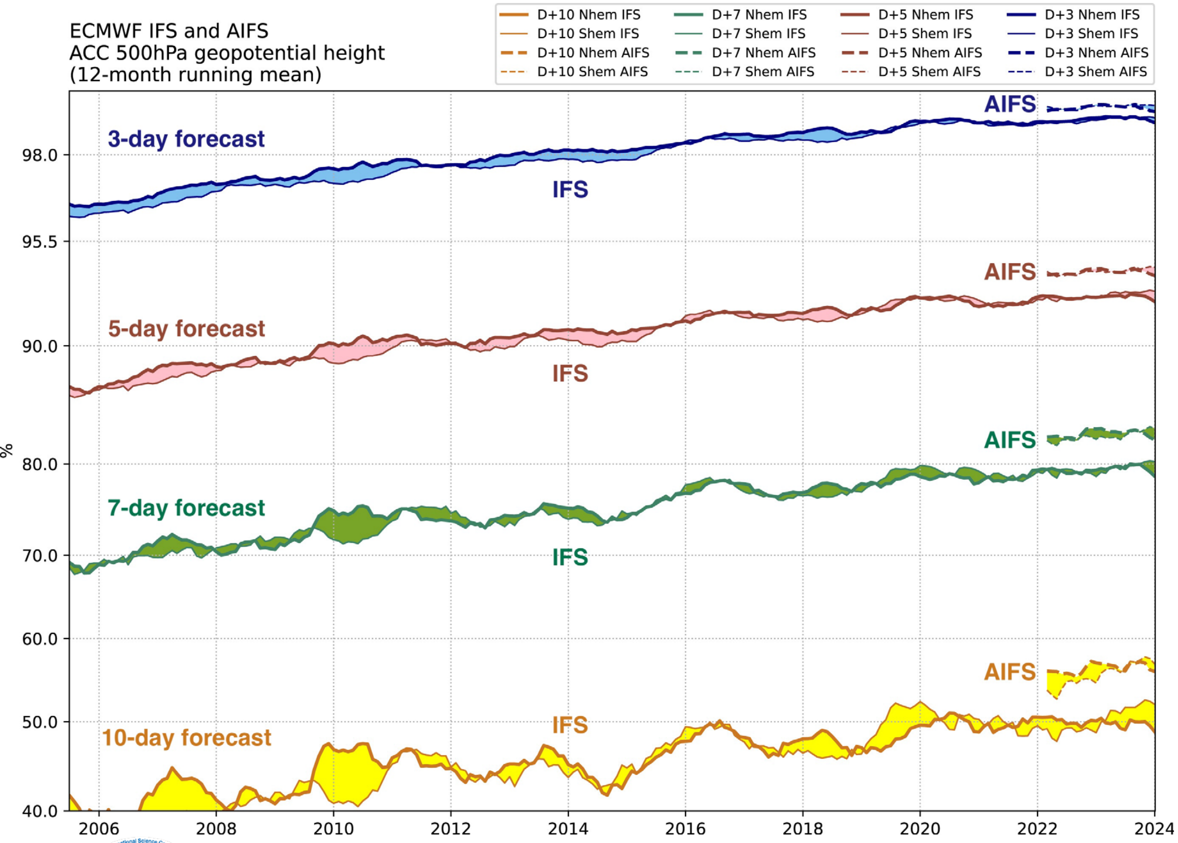
Examples: thickness, compactness/concentration, velocity.

WAVE 2:

→ **new satellite/field products**
→ **richer parameterizations**
→ **better process constraints**

Examples: ice age, albedo, melt ponds, snow, deformation, floe size.

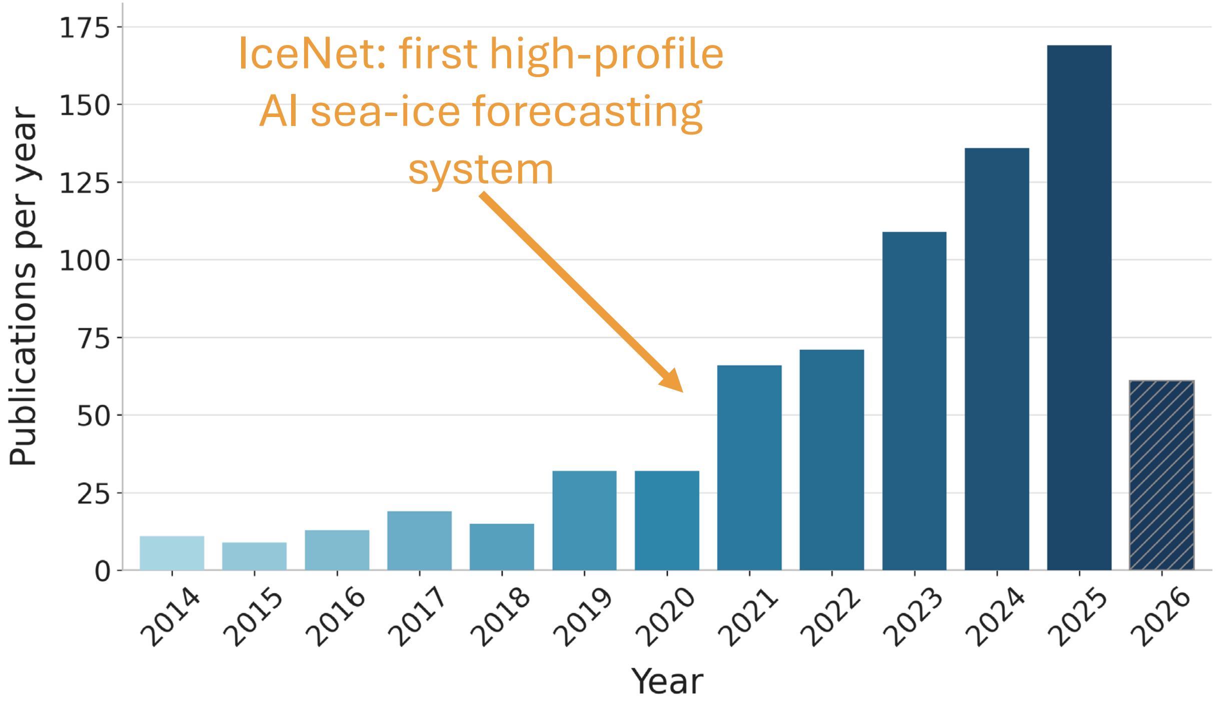
AI systems are built on observations



modified from Bauer (2024) 10.1016/j.jemets.2024.100002

The “AI revolution” is an observational revolution

AI sea-ice prediction is growing rapidly



Yearly count of peer-reviewed publications returned by a Scopus query pairing sea ice prediction terms with AI and machine learning keywords.

AI sea-ice prediction systems are becoming part of the forecasting landscape

AI changes how observations are used

Traditional models

Observations used for:

- Process understanding
- Data assimilation
- Verification

AI models

Need observations for:

- Training
- Validation
- Stress testing
- Physical realism

Observations are embedded throughout the AI workflow

What is ORCAS?

A new WMO WWRP PCAPS and SCOR activity focused on observational requirements for AI-based sea-ice prediction

Why was it created?

AI systems are changing how observations contribute to forecasting, but observational requirements remain poorly understood

What does ORCAS do?

ORCAS connects observationalists and model developers to identify the observations needed for trustworthy AI prediction systems

Building a community around AI and observations

ORCAS

Chair(s)

Clare Eayrs (USA), Lorenzo Zampieri (Germany), Malte Müller (Norway)

Other Full Members

Luisa von Albedyll (Germany), Sandra Barreira (Argentina), David Bromwich (USA), Petra Heil (UK), Zachary Labe (USA), Yafei Nie (China-Beijing), Luciano Ponzi Pezzi (Brazil)

Associate Members

David Clemens-Sewell (USA), Yonghan Choi (South Korea), Wayne De Jager (South Africa), Simon Driscoll (UK), Tobias Finn (France), Lauren Hoffman (USA), François Massonnet (Belgium), Mathieu Plante (Canada), Tian Tian (Denmark)

ORCAS Community Workshop

Workshop title:	*The ORCAS Community Workshop: Connecting Observations and AI Sea Ice Prediction Systems*
Organisers:	ORCAS (Observational Requirements in the Context of AI prediction systems for Sea ice), a SCOR Working Group and PCAPS Task Team.
Dates:	February 10 & 11, 2026 (Two sessions held to accommodate global time zones).

https://doi.org/10.5281/zenodo.19543214

30 mailing list members

We are an open community.

Drop us a message, and you will be involved!

Announcements for more online and in-person workshops will follow.

ORCAS was created to address three questions

1. How do we evaluate AI sea-ice forecasts?

   Beyond standard skill metrics

2. How do we know they are physically realistic?

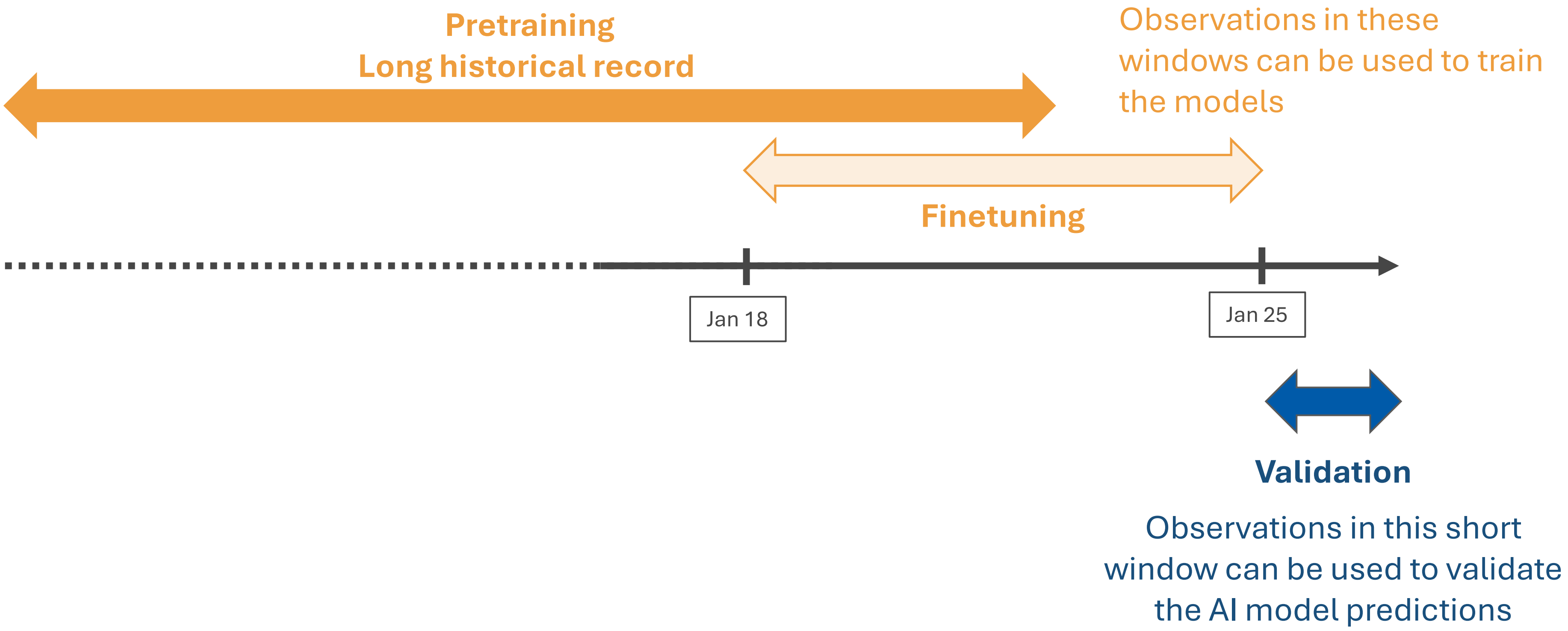
   Testing behaviour, not just accuracy

3. How can observing systems and field campaigns best support the next generation of AI forecasts?

   Training, validation and trust

How do we evaluate AI systems?

Independent observations are increasingly important as AI systems absorb more historical data



AI models predict forecast variables; campaigns often measure processes

MET-AICE
(Palerme et al., 2025)

Sea ice concentration

Ice edge

GenSIM
(Finn et al., 2025)

Sea ice concentration

Sea ice thickness

Sea ice drift

Multiple physically consistent state variables

Campaign observations often measure

Snow depth and snow microstructure

Ice temperature, salinity and brine content

Melt ponds, ridges, floe size and deformation

Turbulent, radiative and atmosphere–ice–ocean fluxes

If a variable is not predicted, it is hard to use it for direct verification

Adding more variables is not automatically beneficial

This creates an observation–model mismatch

How do we know they are physically realistic?

Testing behaviour

Skill is necessary, but not sufficient: AI forecasts must also be physically credible, robust, and useful.

Independent evaluation
Test models on shared cases, common datasets, and observations not used in training.

Physical credibility
Check whether forecasts respect sea-ice processes: drift, thickness evolution, melt, growth, transport, and atmosphere–ice–ocean coupling.

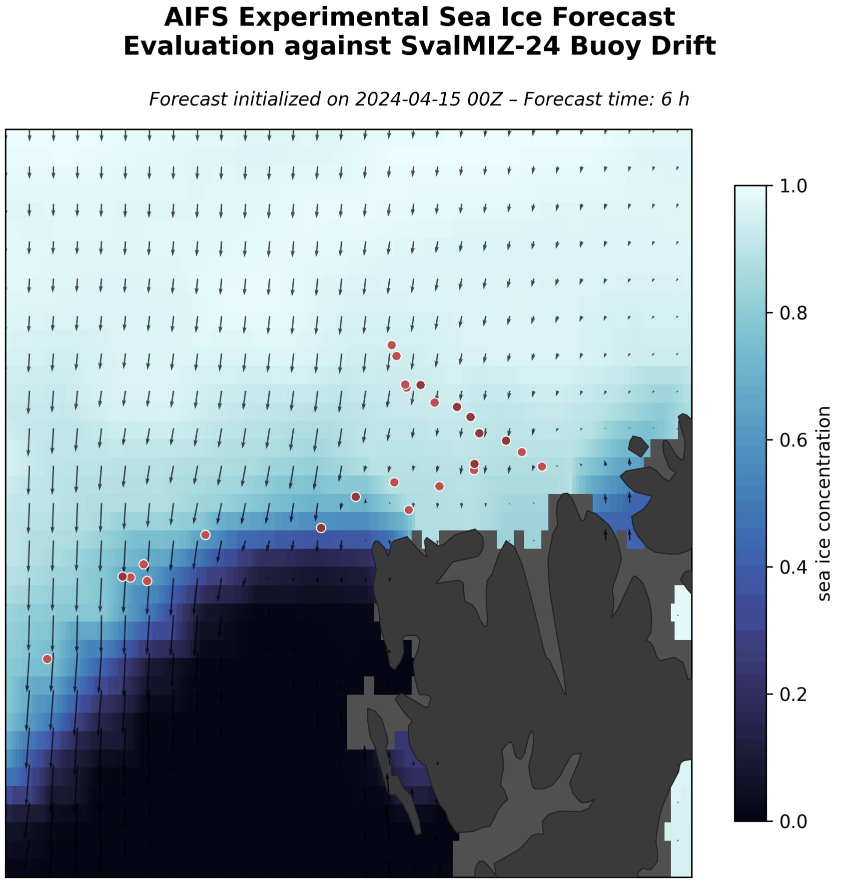
Process-based validation
Use field campaigns and in situ observations to evaluate behaviour that is not captured by simple extent or concentration metrics.

Which observations will make AI sea-ice prediction trustworthy?

Uncertainty and robustness
Quantify when forecasts are reliable, when they fail, and whether models transfer across regions, seasons, and changing climate conditions.

Transparency and interpretability
Understand which inputs drive the forecast and whether the model is using physically meaningful information.

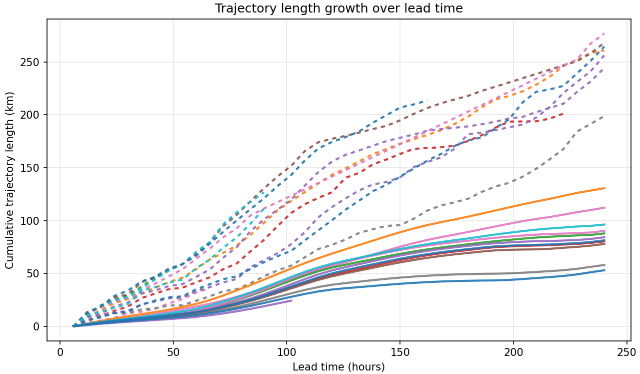
Process observations reveal model behaviour



Lorenzo Zampieri, ECMWF

SvalMIZ-24 (2024)

Wave buoys and drifting observations in the Arctic marginal ice zone



Observed buoy trajectory

Forecast trajectory

WORLD METEOROLOGICAL ORGANIZATION

WWRP WORLD WEATHER RESEARCH PROGRAMME

PCAPS POLAR COUPLED ANALYSIS AND PREDICTION FOR SERVICES

How can observing systems and field campaigns best support the next generation of AI forecasts?

SCOR International Science Council Scientific Committee on Oceanic Research

WORLD METEOROLOGICAL ORGANIZATION

WWRP WORLD WEATHER RESEARCH PROGRAMME

PCAPS POLAR COUPLED ANALYSIS AND PREDICTION FOR SERVICES

What observations should future campaigns prioritise?

Training

- ✓ Sea ice and ocean reanalysis
- ✓ Atmospheric reanalysis
- ✓ Remote sensing datasets
- ? What role for campaigns?

Forecast Production

- ✓ Sea ice and ocean analysis
- ✓ Atmospheric forecasts
- ✓ Remote sensing datasets
- ? What role for campaigns?

Evaluation

- ✓ Remote sensing datasets
- ⭐ Campaign and in situ observations

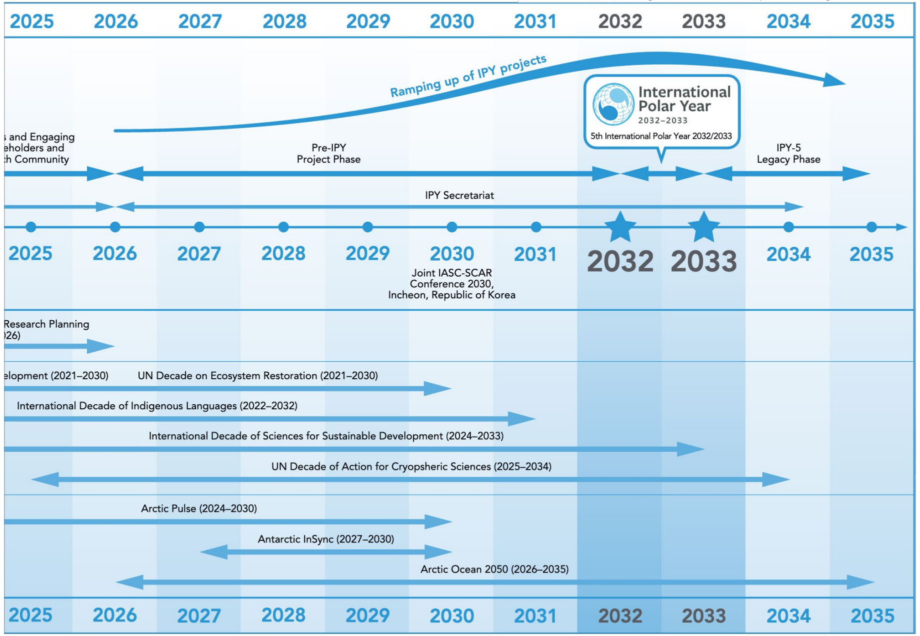
Where do in situ observations & campaign data add the greatest value for AI systems?

buoys • field campaigns • ship measurements

Antarctica InSync and IPY5 provide a unique opportunity

Antarctica InSync (2027–2030)

IPY5 (2032–2033)



Questions ORCAS is addressing

- Which variables provide the greatest value?
- What scales matter most?
- Which sampling strategies reveal model behaviour?
- How should datasets be curated?

SCOR · WORLD METEOROLOGICAL ORGANIZATION · WWRP · PCAPS

The AI forecasting landscape is evolving rapidly

Learn the evolution over time

Predict the future in one step

Predict one day at a time

Future AI systems may use observations differently

Traditional pathway

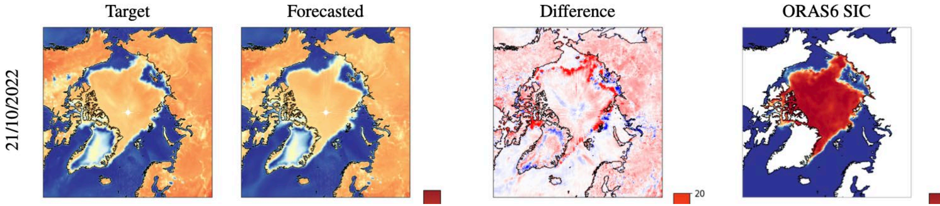
Observations → Reanalysis → Prediction

Emerging pathway

Observations → Direct learning approaches

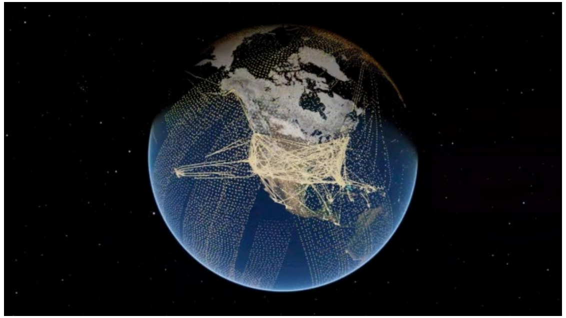
Observing systems should remain useful as AI architectures continue to evolve.

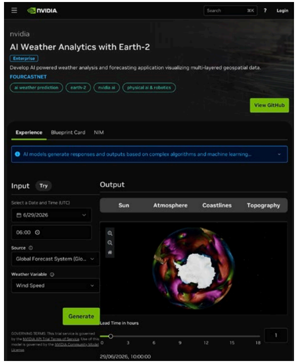
Example: GraphDOP forecasts rapid Arctic sea-ice freeze-up directly from AMSR-2 brightness temperatures



From Boucher et al., 2025, *Learning Coupled Earth System Dynamics with GraphDOP https://doi.org/10.48550/arXiv.2510.20416*

Example: new approaches from NVIDIA's Earth-2 initiative





Sea ice thickness as a test case

ORCAS is using sea-ice thickness as a pilot variable to explore how observational datasets can best support AI prediction systems.

Open questions:

Are multiple datasets better than a merged one?

Is a gridded product better, or is working with swath observations the way to go?

Can AI models handle less refined data?

Can AI models handle multiple missions?

Can AI models handle campaign observations?

Does training on thickness information benefit the prediction of other observations?

Join ORCAS

clare.eayrs@nyu.edu

Sign up for the community mailing list

Contribute to the literature review

Join our online events

Stay posted for information on in-person workshops