Machine Learning for Ocean Prediction: Methods, Applications & Challenges

AI-TT meeting 2026

13 & 14 April 2026

UQAM, Chaufferie, Montréal
Canada

Report

Prepared by

Rachel Furner, Anass El Aouni, Kirsten Wilmer-Becker and session rapporteurs

AI-TT members and guests at UQAM, Chaufferie – April 2026

Enquiries:
Kirsten Wilmer-Becker and Stephanie Cuven
OceanPredict Programme Office Coordinators

Kirsten.Wilmer-Becker@metoffice.gov.uk
scuven@mercator-ocean.fr

Table of content

Executive summary

The first in-person meeting of the OceanPredict Artificial Intelligence Task Team (AI-TT) was held in Montreal in April 2026. Co-chaired by Rachel Furner (ECMWF) and Anass El Aouni (Mercator Ocean International), the event brought together researchers from various backgrounds to assess recent advances in artificial intelligence (AI) for ocean prediction. The workshop provided a forum to share progress, identify challenges, and discuss priorities for future research and operational implementation.

A central theme throughout the workshop was that AI methods are rapidly transitioning from proof-of-concept studies to operationally relevant ocean forecasting applications, mirroring developments previously seen in weather prediction. Participants presented advances across ocean model emulation, data assimilation, data reconstruction, downscaling, uncertainty estimation, forecasting, and benchmarking. At the same time, discussions repeatedly emphasized the importance of maintaining physical consistency, robust evaluation practices, and transparency in AI-based systems.

Major Scientific Themes:

Ocean Emulators and Neural Forecasting

Several presentations demonstrated the growing capability of AI-based ocean emulators to reproduce complex ocean dynamics at a fraction of the computational cost of traditional numerical models.

Mercator Ocean presented a scalable probabilistic ocean emulator capable of generating 100 years of monthly global ocean simulations in approximately 15 minutes, highlighting the potential of AI for climate scenario analysis, digital twins, and ensemble forecasting. Other studies explored machine-learning prediction of sea-surface height and currents in regional domains such as the Gulf of St. Lawrence and the East/Japan Sea. These studies demonstrated that incorporating physical constraints through physics-informed neural networks (PINNs) can significantly improve forecast stability and realism, particularly at longer lead times.

Across these presentations, participants recognized both the promise and the limitations of current approaches. While skill levels are often competitive with traditional systems for selected variables and lead times, challenges remain regarding error growth, physical consistency, and extrapolation beyond training conditions.

AI for Observations, State Reconstruction, and Data Assimilation

A second major focus was the use of AI to reconstruct ocean states from sparse observations and to complement traditional data assimilation methods.

Several presentations demonstrated how neural networks can accurately reconstruct sea-surface height, subsurface temperature and salinity, and biogeochemical variables from incomplete observations. Studies showed that models trained on data from numerical simulations or reanalysis can successfully transfer to real-world observations, often outperforming existing mapping approaches in highly dynamic regions such as the Gulf Stream and Kuroshio.

Particularly noteworthy was the development of uncertainty-aware and transformer-based architectures capable of providing both state estimates and confidence information. These approaches represent an important step toward operational deployment, where quantification of uncertainty is essential for decision-making.

Physics-Informed and Hybrid Approaches

A recurring topic was whether combining machine learning with physical knowledge provides significant benefits compared to purely data-driven approaches.

Studies demonstrated that physics-informed constraints improve the representation of currents, maintain physical balances, reduce spurious behaviour, and enhance long-range forecast skill. Hybrid approaches were also applied successfully to reconstruct ocean oxygen concentrations and biogeochemical fields from physical variables alone, substantially reducing biases in Earth System Models and improving agreement with observations.

The workshop highlighted growing interest in integrating AI directly within modelling frameworks, including parameterizations, data assimilation systems, and differentiable modelling approaches that combine physical equations and machine learning within a unified framework.

Downscaling and Coastal Applications

A number of presentations focused on the challenge of translating coarse-resolution information into detailed coastal and harbour-scale predictions.

Machine-learning approaches demonstrated promising results for statistical downscaling of sea-surface temperature, wave fields, coastal bathymetry evolution, and harbour-scale circulation. AI-based super-resolution techniques were shown to reproduce high-resolution climate and ocean information while dramatically reducing computational costs relative to traditional dynamical downscaling.

Participants highlighted that these applications may represent some of the most immediate opportunities for operational adoption of AI in ocean forecasting.

Key Challenges Identified

Despite significant progress, several common challenges emerged:

- **Physical consistency:** Accurate forecasts do not necessarily imply correct underlying dynamics. Multiple presentations stressed the risk of obtaining the "right answer for the wrong reason."
- **Evaluation standards:** Robust and standardized benchmarking remains essential for comparing diverse AI and numerical forecasting systems fairly.
- **Generalization and extrapolation:** Many models perform well within their training domain but struggle under future climate conditions or unusual events.
- **Uncertainty quantification:** Operational use requires reliable estimates of forecast confidence and model uncertainty.
- **Data limitations:** Uneven observational coverage and incomplete datasets continue to constrain model development and validation.
- **Computational resources:** While AI can dramatically reduce inference costs, training large-scale models efficiently depends on GPU access, which is not widespread.

OceanBench: Toward a Community Benchmark

One of the most significant outcomes of the workshop was the presentation of OceanBench, a new community-driven benchmarking framework for AI and physics-based ocean forecasting systems.

Inspired by the success of WeatherBench in meteorology, OceanBench provides:

- Standardized datasets including observations, analyses, and reanalyses.
- Common evaluation protocols.
- Open-source tools and workflows.
- Leaderboards and model reports.
- Metrics that assess both forecast accuracy and physical realism.

Initial benchmarks show that AI models can already compete with operational numerical systems for some variables, particularly subsurface properties and ocean currents. Importantly, OceanBench introduces process-based diagnostics, including geostrophic currents, mixed-layer depth, and Lagrangian trajectories, to ensure models are evaluated on their physical behaviour and not solely on traditional error statistics.

Participants strongly supported the initiative and emphasized the need for community governance, transparent methodologies, and multiple independent reference datasets.

1. Introduction

1.1. Welcome and Acknowledgments

The AI-TT co-chairs and coordinators express their thanks to the host institution and sponsors for their generous support in making this meeting possible. Special recognition is extended to Gregory Smith, and the local organisers for their dedication and hard work throughout the organization of this event. Thanks also go to Normand Gagnon, Acting director of the Meteorological Research Division at Environment and Climate Change Canada, along with his colleagues their welcoming hospitality and to CMEMS for providing event sponsorship.

This gathering marked the first in-person meeting of the AI-TT under the co-leadership of Rachel Furner (ECMWF) and Anass El Aouni (MOI).

1.2. Meeting objectives

Recent developments in artificial intelligence (AI) capabilities (including neural network approaches, machine learning and deep learning related tools) have demonstrated the potential to provide accurate forecasts in weather and environmental forecasting. The OceanPredict Artificial Intelligence Task Team aims to create a forum to discuss recent developments in the application of AI to ocean forecasting, to share best practices and explore areas for future development.

The mission of the AI-TT workshop, within the OceanPredict framework, was to create a collaborative forum to share progress and practical experience in applying AI to ocean forecasting, while fostering networking and developing future partnerships. The event aimed to facilitate focused discussions on key challenges—such as evaluation methods and assessing the added value of AI—and to identify priorities for future research and development in this field.

1.3. Workshop themes

No	Themes
1	Machine Learning Emulators: Design, development, and application of machine learning, including generative approaches, and surrogate models for complex ocean processes
2	Hybrid Approaches: Integration of physics-based ocean models and machine learning techniques to enhance predictive accuracy and efficiency, eg. use of ML components (parameterisations, etc) in state-of-the-art physical ocean models, and combining physics and deep learning within in a single differentiable programming framework
3	Deep Learning for Data Assimilation, and inversion schemes: Innovative uses of deep learning architectures to assimilate diverse oceanographic datasets, including satellite and in-situ observations. Use of ML for ocean state estimation and forecasting.
4	Evaluation Challenges: Strategies and benchmarks for assessing the performance, robustness, and reliability of deep learning-based emulators in operational settings.
5	Technical challenges: Operationalization, Novel architectures, Managing and sharing large datasets, etc.
6	Other relevant ML applications for Ocean prediction (e.g. downscaling applications, ensemble forecasting)

2. Sessions

All workshop session presentations are available in *Appendix B* and on the AI-TT web pages.

The session descriptions in this paragraph were kindly provided by our session chairs and rapporteurs.

Session 1: Ocean emulators

Session chair and rapporteur:

- Rachel Furner (ECMWF)
- Anass El Aouni (MOI)

2.1.1 Toward Scalable and Probabilistic Neural Ocean Forecasting

Presenter: Simon Van Gennip (Mercator Ocean)

The work presented focused on scalable, probabilistic AI-based ocean forecasting, building on recent advances in atmospheric prediction such as GraphCast and GenCast. While AI weather models have rapidly achieved skill comparable to numerical weather prediction, ocean forecasting has lagged because of the ocean's greater complexity, diverse forcings, limited observations, and multi-scale dynamics.

The team has developed an integrated ocean emulator that produces monthly ocean forecasts at 1° resolution with eight vertical layers and can simulate long time periods efficiently. Trained on four H100 GPUs over seven days, the model can generate 100 years of monthly means in around 15 minutes, making it attractive for climate studies, seasonal predictions, and large ensembles of "what-if" scenarios.

Evaluation against reanalysis data showed good performance for sea-surface temperature, sea-ice variability, chlorophyll, and climate indicators such as El Niño. Some biases remain in salinity and sea-surface height trends, but these were viewed as tractable issues. Initial analyses suggest the model preserves important variability on shorter forecast horizons, although longer forecasts tend to lose energy and drift toward a steadier state.

Future work will focus on increasing horizontal and vertical resolution, producing outputs more frequently than monthly, improving probabilistic sampling methods, evaluating subsurface behaviour, and coupling with machine-learning atmospheric models. The approach could ultimately support Digital Twin of the Ocean applications, enabling low-cost exploration of climate intervention, carbon removal, and climate-change scenarios over decadal timescales. Questions from the audience focused on possible training-data contamination, variable selection, and the impact of using mean-square-error loss functions on forecast smoothing.

2.1.2 Application of Deep learning (DL) in the Gulf of St. Lawrence and Estuary

Presenter: François Roy, ECCC

The work presented covered applications of machine learning to emulate sea-surface height (SSH) and ocean currents in the Gulf of St. Lawrence, a challenging environment due to its complex dynamics, tides, stratification, bathymetry, and atmospheric forcing. The study uses a 30-year, 2 km-resolution Canadian operational ice–ocean prediction system dataset for training.

A key aspect of the work is separating ocean flow into barotropic (depth-averaged, largely driven by tides and external boundary forcing) and baroclinic (density-driven and locally influenced by stratification, bathymetry, and wind events) components. This decomposition is intended to support future 3D ocean emulators, where retaining physically meaningful variability is critical.

Initial machine-learning experiments showed promising skill in reproducing sea-surface height and currents but experienced significant degradation after around three days of forecast lead time. Future improvements may include incorporating ocean boundary forcing, combining different machine-learning architectures (e.g., LSTM for SSH and graph-based models for currents), extending lead times to support ensemble forecasting, and using pre-training strategies for 3D current prediction.

Discussion focused on ways to improve performance, including using higher-resolution datasets, assimilating observations such as tide gauges, adding more physical variables and constraints, extending temporal context, and exploring autoregressive approaches. Roy noted that memory limitations currently restrict model complexity, while some forecast artefacts and noisy patterns may be reduced through better model tuning and architecture design. A particularly promising future application is rapid high-resolution downscaling, enabling detailed forecasts without the need to run computationally expensive ocean models.

2.1.3 Prediction of sea surface currents around the Korean peninsula using artificial neural networks

Presenter: Jae-Hun Park, Department of Ocean Sciences, Inha University, Republic of Korea

The work presented focused on AI-based prediction of surface currents and sea-surface height (SSH) in the East/Japan Sea, a deep, semi-enclosed basin surrounded by Korea, Japan, and Russia that exhibits highly complex circulation patterns. Building on earlier work, the team enhanced their model by incorporating Physics-Informed Neural Networks (PINNs), which embed physical constraints into the learning process.

The study focused on predicting daily ocean currents and SSH, using a transformer-based architecture with physics-based loss terms designed to reduce physically unrealistic behaviour. Results showed that the model successfully reproduced both currents and sea-surface height, with the PINN-enhanced version outperforming a purely data-driven approach. In particular, the physics-informed model reduced errors, decreased spurious divergence, and produced more accurate representations of flow structures.

Validation against independent drifter observations demonstrated that the PINN model performed better than both traditional numerical simulations and AI models without physical constraints. While the benefits of the physics-informed approach were limited for very short forecasts, they became increasingly important for longer lead times (up to 10 days), improving forecast stability and physical consistency.

During discussion, participants explored alternative formulations of physical constraints, the role of attention mechanisms within the transformer architecture, and the influence of input data quality. Park noted that adding the physics-based constraints introduced little additional training cost while providing clear gains in long-term forecast skill, highlighting the value of combining machine learning with established ocean dynamics.

Session 2: Observations and Data Assimilation

Session chair and rapporteur:

- Jae-Hun Park (Inha University)
- Vikram Khade (ECCC)

2.2.1 Training end-to-end neural mapping schemes from simulation data for the reconstruction of global-scale sea surface fields

Presenter: Daniel Zhu, IMT Atlantique, France

The objective of this work is to recover/reconstruct the dense global SSH given the sparse and irregular sample provided by satellite altimetry. This work demonstrates that models trained on simulated observations can work well with real data during inference. The model takes in observations and estimates the hidden states. The evaluation is based on the Ocean Data Challenge Global OSE Mapping (by Datlas & CLS), the SLA from a constellation of 6 nadirs in 2019. Three different models are trained – UNet, 4DVarNet-ConvLSTM & 4DVarNet-Unet. The training data is reanalysis GLORYS12 from 2010 to 2017. Two different datasets are used in which the second dataset introduces noise to observations. This work also carries out ensemble inference by adding gaussian noise. The training is deterministic. In the average RMSE, this technique outperforms MIOST. However, it degrades in the Pacific compared to MIOST. This technique makes gains in high variability regions like Gulfstream and Kuroshio.

2.2.2 A new model-data fusion approach: 2-step training procedure to integrate chlorophyll

Presenter: Teresa Tonelli, OGS, Italy

The objective of this work is to implement an AI based technique in place of a traditional DA to predict biogeochemical variables given physical variables. This work uses a Convolutional Neural Network (CNN) to output the 3D fields of biogeochemical variables on 30 layers. The input to the CNN are fields of physical variables such as temperature. The resolution is 1/12 degrees. This CNN, named MuSt3Net trains over two steps. In the first step it learns to emulate the numerical ocean model and in the second step it learns to integrate BGC-Argo float data. In the first step profiles from BIO tensor are sub-sampled and concatenated with the physical tensors. The training is done to minimize the loss between 3D CNN output and the 3D BIO ground truth. In the second step, the physical and chlorophyll tensors are concatenated. The loss between the model output profiles and the BGC-Argo profiles is minimized. The training data is from CMEMS reanalysis (NEMO-OceanVar) and the chlorophyll data is from CMEMS OGSTM-BFM over the time period of 2019-2021.

After step1, among all the seasons the loss over the fall is minimum (0.051 mg/m3) while that over spring is maximum (0.093 mg/m3). After step2, among all the seasons the loss over the fall is minimum (0.055 mg/m3) while that over spring is maximum (0.099 mg/m3). The profiles of chlorophyll are compared with the float profile. Higher accuracy is seen in summer and fall in spite of challenges in predicting surface blooms. The higher accuracy in Ionian and Levantine seas is due to higher Argo float coverage and lower chlorophyll variability. The Hovmoller plots show that this technique also accurately captures the temporal dynamics of chlorophyll at the basin scale. This technique predicts the 3D domain while preserving the spatial information. However, the availability of observed profiles plays a key role in deciding the quality.

2.2.3 An Uncertainty-Aware Transformer Framework for Satellite-Driven Subsurface Thermohaline Reconstruction

Presenter: Geon Min Lee, Pukyong National University, Republic of Korea

The objective of this work is to reconstruct the subsurface temperature and salinity profile from the incomplete satellite observations. The key contributions of this work include observations-space mapping and uncertainty aware reconstruction. The training samples have SSH, SST, bathymetry and time as the input and the subsurface T/S profiles and its uncertainty as the output. The Neural network, named the Temperature and Salinity Transformer (TST) used in this work consists of convolutional blocks for vertical structure and hybrid attention across variables and depth.

The TST uses a grid-wise strategy. The satellite information from a track point is distributed to the nearest four grid points. The inverse distance weighting interpolation is used to obtain the final profile. TST takes the approach of training a separate model for each grid point. The training takes about 200 minutes for 8300 grid points and the inference takes about 1 minute for 2700 grid points. Close to the surface the profile has high variability and this is called gradient dominated while deeper down it has stable structure and this is called value-dominated. Between these two there is a transition zone. The profile in the transition zone is constructed by smoothly combining the profiles above and below using a weighting function. TST also estimates the uncertainty. This is achieved by outputting the log-variances along with the profiles. A heteroscedastic negative log-likelihood loss is used for training. The northwestern Pacific domain is used in this work. In the first experiment GLORYS surface data interpolated to 0.5 degree resolution as training data. In this experiment GLORYS sub surface data is used as the target. In the second experiment, the satellite surface data are used as the input and in-situ observations such as Argo, KODC and Kuroshio buoy data are used as the target. For both the experiments, the training data is from 2010-2019, the validation in 2020 and test in 2021-2022.

The inference results of the first experiment show that TST accurately reconstructs the T/S. This is quantified by using the RMSE and correlations. The reconstruction maintains the physical consistency as seen from salinity minimum core and potential density structure. The inference results of the second experiment show that the TST outputted calibrated uncertainty closely matches the observed RMSE. This match is improved due to the calibration. The TST has some limitation. The biases present in the GLORYS reanalysis data affect the result of the TST. Also the TST is limited in its ability to capture high-frequency variability since monthly data is used. The future plans to extend this work include extension of TST to global domain and its integration into ocean models for data assimilation.

2.2.4 Short-term end-to-end neural forecasts of sea surface dynamics

Presenter: Daria Botvynko, IMT Atlantique, France

This work presents a deep learning (DL) framework, which forecasts sea surface dynamics given the input of sparse altimetry data. The deep learning framework presented is based on two extant architectures – U-net (17 M parameters) and 4DVarNet (652 K parameters). This framework uses a 14 day temporal window of inputs. It then forecasts the SSH field 7 days into the future. This DL network is trained on simulation data while its performance is assessed on real ocean observation data. The 4DVarnet scheme uses a weighted combination of three losses – MSE loss applied to the state and its spatial gradients and a third term which is a regulariser. The U-net uses the combination of only the first two losses. The input field for training are simulated nadir-like tracks sampled from the GLORYS12 SLA field for the past 14 days. The future 7 days of GLORYS12 SSH fields are the targets. The inference phase the near realtime satellite altimetry data is used. The training period is from 2010-2019. The training time on one A100 GPU is about 2 days for 4DVarNet and about 1 day for Unet.

In the inference, the predicted SLA and SSC fields are assessed by comparing against observational datasets (Saral/AltiKa and velocities of drifters trajectories at 15-m depth.). This assessment is intercompared with SSH outputs of operational ocean forecast GLO12 and state of the art neural ocean emulators. The neural ocean emulators (GloNet, XiHe) use the full field provided by data assimilation as the input.

These models are compared with the DUACS and forecast accuracies are calculated as percentage gains relative to the DUACS score for a given lead time. Overall, all models lose performance relative to their first lead time as forecasts advance, yet 4DVarNet keeping the lowest level of relative errors wrt DUACS, highlighting the ability of this end-to-end neural architecture to remain close to an observations-based short-term forecast even at longer lead-times. Results are inspected in detail in high-variability regions like Gulf Stream and Agulhas. Both 4DVarNet and UNet recover spatial patterns better than GLO12 given the baseline of DUACS. At lead time of 5 days the Unet shows decreased SLA variance whereas this decrease is not observed in 4DVarNet forecasts. The models are also assessed for the explained variance metric with respect to GLO12 for lead-time 0. 4DVarNet performs very well in small scales (65-500 km), in offshore regions. This might be due to the third regularization term in the loss function.

2.2.5 High spatiotemporal resolution ocean temperature reconstruction and forecasting based on artificial intelligence

Presenter: Liying Wan, NMEFC, China

This work introduces the Multi-scale Residual Spatiotemporal Window Ocean Model (MSWO). It is an improved version of the Spatiotemporal Window Ocean Model (SWO) which is based on Convolutional Neural Network architecture with attention. MSWO embed a global model in SWO to resolve multiscale features. This study uses the Pacific ocean domain from the time period 2011-2019. Satellite remote sensing data is used as an input to the MSWO. ARGO in-situ data is used as the target ground truth. The test RMSE shows that MSWO performs better than ConvLSTM, PredRNN, SwinLSTM-B, SwinLSTM-D, SA-ConvLSTM and SimVP. This is true of both the Northern and Southern hemispheres.

Session 3: Emulators

Session chair and rapporteur:

- Charlie Hebert-Pinard (ECCC)
- Daniele Bigoni (CMCC)

The session outlined several applications of machine learning in ocean simulation and provided insight on the dangers of blindly using machine learning models.

2.3.1 Linear Stochastic Emulators of the Ocean Circulation – A Lesson, perhaps, for Machine Learning

Presenter: Andrew Moore, University of California Santa Cruz, US

The session opened with a talk by Andrew Moore (UCSC) on the risks of using machine learning models for dynamical systems without paying enough attention to the dynamics but only to the reconstruction of the fields of interest. Machine learning models leverage correlations between variables to learn relations between features and outputs. Traditionally these correlations have been exploited to build optimal subspaces for the representation of the fields via Empirical Orthogonal Functions, but these fall short in representing any causal relation between input and output. The speaker discussed the ability of Balanced Truncation to overcome these issues. The methodology is likely related to Koopman operator learning.

2.3.2 A data driven limited area storm surge model

Presenter: Mateusz Matuszak, Norwegian Meteorological Institute, Norway

Mateusz Matuszak (NMI) showed the broad approach to AI taken by the Norwegian Meteorogical Institute, where several emulators are being developed for each of their operational products. Mateusz’s work focused on the construction of Flo, an emulator for storm surges based on NORA-Surge and developed using the ECMWF Anemoi framework. The talk showed encouraging results against observations over the test period, as well as an in-depth analysis of Storm Xaver, showing how the emulator can tackle also extreme event scenarios. In future steps they want to include observations in the model to improve prediction and naturally assimilate data.

2.3.3 A Physics Informed Emulator for Ocean Oxygen (remote presentation)

Presenter: Annalisa Bracco, CMCC, Italy

Annalisa Bracco (CMCC) presented work on the reconstruction of oxygen concentration from observations at global and decadal time scales. The work showed how both neural network architectures and random forecasts are able to reconstruct past data, as well as build climatologies in agreement with WOA18. However, these systems struggle with having predictive skill when extrapolating. The talk closed with the opportunity provided by these models to perform uncertainty quantification and to use this information to pick locations where to deploy measurement instruments.

2.3.4 Application and Verification of the Global Wave Intelligent Forecast Model

Presenter: Fang Hou, NMEFC, China

Fang Hou (NMEFC) outlined the roadmap of the National Marine Environmental Forecasting Center on the development and deployment of marine early warning and forecasting services, both regional and global. This roadmap now encompasses the development of a number of machine learning models based on both reanalysis and forecasting data, of which they measure performance against in-situ observations and comparable products, both overall and during extreme events (typhoons).

Session 4: Downscaling and other ML methods

Session chair and rapporteur:

- Frederic Dupont (ECCC) and
- Corbeil-Letourneau (ECCC)

2.4.1 Development of machine-learning emulators for harbour-scale ocean prediction

Presenter: Michael Dunphy, Institute of Ocean Sciences, Fisheries and Oceans Canada

Across Canada, Nemo-based port models support electronic navigation and drift prediction. The main components adding to complexity of the prediction include dynamics dominated by tides and complex coastlines. The presentation is an overview of potentially useful ML methods that could speed up prediction and decrease the computational cost compared to actual physics driven models. Two explorations were described (1) a super-resolution downscaling method (Residual-Unet) was used for mapping SSS150 hourly snapshots predictions as input onto VH20 model prediction as target, (2) Neural-Operators were used to predict VH20 Barotropic mode from SS500 prediction as input. The base Res-Unet showed limited skill, the most effective Res-Unet variation was one that included 3h lookback inputs. Using de-tided SSH and a model including Tide an additional input didn't improve the performance. The limited receptive field of the convolution kernel is a suspected cause for the poor performance of this approach. The second exploration was based on autoregressive Fourier (FNO) and Laplace Neural operators (LNO). These models were trained on SS500 patches and atmospheric surface forcing (P, u10 and v10) to output barotropic field (ssh, ubar, vbar) snapshot of VH20 as targets. In the case of FNO best performances were obtained using a curriculum rollout, tide clock and a small number of Fourier modes and. In the case of LNO, the results were very similar although the errors were smoother. This discrepancy is possibly associated with Gibb's phenomenon. The FNO seems better at non-tidal Water level and LNO is better at tidal Water level.

2.4.2 From Coarse Models to Coastal Detail: A Deep Learning Approach to AI based Statistical Downscaling in the Adriatic Sea

Presenter: Alessandro De Lorenzis, CMCC, Italy

This study addresses a key limitation in current climate modeling: while Global and Regional Climate Models are essential for understanding climate change, their relatively coarse spatial resolution makes them less effective for analyzing impacts at local scales. To overcome this, the authors propose an AI-based statistical downscaling (SD) pipeline to learn the mapping from coarse-resolution predictor to a high-resolution (SR) SST target, enabling efficient generation of fine-scale daily fields while preserving coastal gradients and mesoscale structure. The Adriatic Sea is used as a test case, with lower-resolution reanalysis data serving as input and higher-resolution dynamical simulations as the target. On the test set, the models show very high pattern correlation and low errors, with near-zero mean bias, with respect to the AdriaClimPlus target. Both dynamical downscaling and statistical downscaling consistently reproduce a robust Adriatic SST warming signal, which intensifies from short- to long-term horizons. The ability of SD to replicate the climate signal even in projection mode confirms the robustness of the ML framework and the effectiveness of the training strategy. This supports the use of SD as a scalable and computationally efficient pathway for generating high-resolution climate projections.

2.4.3 Integrated AI Physics Approaches for Coastal Prediction Across the Open-to-Coastal Ocean Continuum

Presenter: Joanna Staneva, Helmholtz Zentrum HEREON, Germany

Prediction of Coastal dynamics is challenging because of strong nonlinearities, multi-scale dynamics, sparse and heterogeneous observations, uncertainty in forcing and boundary conditions. All these challenging elements are reasons for seeking improvement and integrated systems. One task described in the talk was the reconstruction and the forecasting of bathymetry using deep learning. Convolution combined with LSTM is used to reconstruct and forecast the evolution of it. Another part of the talk was dedicated to ML Downscaling predictions; discussed with the angle of their possible randomness and variability. An ensemble of SRRestNet is proposed to address this issue which was shown to perform well when compared with a deterministic alternative like Multilinear regression. The ensemble is trained and evaluated on global ERA5 dataset to perform spatial SWH downscaling over the Black Sea. A third part of the talk was dedicated to the challenge of coastal risk analysis and Scenario Optimization. Coastal risk evaluation is a multi-compartment system. The proposed approach is to replace, in the ecosystems of models, a costly high-resolution model. In the case shown, the model is a Morphodynamics model, XBeach. A purely AI-model is proposed for the substitution, such a digital twin of coastal systems can enable the exploration of questions of the type ‘What-if’, and enable modelling of multiple scenarios for risk management.

2.4.4 Fix the double penalty in data-driven forecasting by modifying the loss function

Presenter: Christopher Subich, ECCC, Canada

The smoothing of prediction from deterministic AI model is a well know issue. The common wisdom tells that to solve it; one needs to use an ensemble model. This solution is not the only option, a modified MSE loss function can be proposed to mitigate this issue. In fact, the major cause of the smoothing is due to the double penalties’ problems associated with the mean square error. In simple words, prediction wrongly positioned is more costly than predicting nothing. For example, predicting a Hurricane at the wrong location involves a penalty due to a vortex located where no vortex exists and plus the penalties of the vortex not predicted where one must be predicted. The training process leads quickly to a model that predicts a smooth field, limiting extremes. A modified version of MSE loss function, called AMSE, can be proposed to solve the smoothing problem. The mean square error loss function written in terms of spectral amplitude and coherence can be decomposed in two parts, one involving the square of the difference of modes amplitude between target and prediction, the other part, always positive, is proportional to the square root of the mode’s amplitude product of the target and prediction. This second part is responsible for smoothing, modifying the product of amplitude by the max amplitude of target or prediction will be the fix of double penalties problems. With such AMSE loss function, much better predictions can be achieved.

2.4.5 Variational autoencoder-based clustering for geophysical fluid circulations with small sample size

Presenter: Kunihiro Aoki, Meteorological Research Institute, Japan Meteorological Agency

Kyucho phenomenon is a sporadic strong current event along Japan's coast, with rare occurrences. Classification and recognition of such a dangerous meteorological event is an important task that ML can address. A variational autoencoder is presented as a method suitable for such a task, however the largely unbalanced dataset of Kyucho and non-Kyucho events is an issue. Over 60 years, only 197 events were usable for training ML methods. Such small datasets are very limited for properly training a deep learning method. A data-augmentation method is presented to circumvent this limitation. The data-augmentation approach is a noise injection method along the main directions that are obtained by principal component decomposition. The resulting classification following an appropriate training of the VAE, leads to identification of potential precursors of Kyucho formations. Specific circulation fields structures have been identified based on 4 distinctive classification modes of Kyucho generation.

3. Keynote presentation on “OceanBench: A Benchmark for Data-Driven Global Ocean Forecasting systems”

Presenter: Anass El Aouni, MOI

Overview

Anass El Aouni (Mercator Ocean International) presented OceanBench, a community-driven benchmarking framework designed to evaluate both machine-learning and physics-based ocean forecasting systems. The initiative is inspired by the success of WeatherBench, which played a major role in accelerating AI weather forecasting by providing common datasets, evaluation protocols, and leaderboards. With the rapid emergence of AI ocean forecasting models since 2024, OceanBench aims to provide a similar shared infrastructure for the ocean community.

Motivation

The presentation began by highlighting the explosive growth of AI in geosciences over the past five to six years. Weather forecasting has benefited enormously from community benchmarks, open datasets, and standardized evaluation procedures, enabling collaboration between domain scientists and machine-learning researchers. Ocean forecasting is now following a similar trajectory, with increasingly sophisticated AI models trained on high-resolution ocean datasets. OceanBench was developed to support this transition by providing a common platform for fair and reproducible model assessment.

OceanBench Framework

OceanBench consists of several integrated components:

- **Reference datasets**, including reanalyses, analyses, and observations.
- **Benchmark models**, including both numerical and AI-based forecasting systems.
- An open-source **GitHub repository** implementing evaluation workflows and metrics.
- A **website and leaderboard** displaying benchmark results and model reports.
- A growing **computing infrastructure** hosted within Mercator Ocean for automated evaluation.

The framework is intended to be community-driven, allowing researchers to evaluate their own models using a common set of standards.

Data and Evaluation Setup

All benchmarked models are initialized from the same ocean state and driven by the same atmospheric forcing to ensure fair comparison. The primary reference datasets include:

- GLORYS reanalysis
- GLO12 analysis
- In situ and satellite observations

Training data can be accessed through OceanBench APIs or existing marine data services. The benchmark currently focuses on short-range global ocean forecasts.

Models Benchmarked

Three AI models and one operational physics-based model were highlighted:

1. **GLO12** – Mercator Ocean’s operational numerical forecasting system used as a baseline.

2. **GLO-Net** – Mercator's neural-network forecast model designed to capture both large-scale and small-scale ocean dynamics through multi-scale architectures and latent-space learning.
3. **XiHe (Wang et al)** – A transformer-based model operating at native high resolution and using multiple specialized networks for different forecast lead times.
4. **WinHai** – A newer transformer-based model using Swin Transformer v2 architecture with local attention mechanisms to improve efficiency and multi-scale representation.

The models differ significantly in architecture, resolution, training strategy, and forecasting methodology.

Evaluation Methodology

OceanBench performs two complementary styles of evaluation:

1. Point-wise Evaluation

Traditional forecast verification comparing model outputs against:

- Observations (Class-4 style verification)
- Reanalysis datasets
- Analysis products

Metrics such as RMSE are computed across variables, depths, locations, and forecast lead times.

2. Process-Based Evaluation

Beyond simple error statistics, OceanBench assesses physical realism by examining derived quantities that are not explicitly trained:

- Geostrophic currents
- Mixed layer depth
- Lagrangian particle trajectories
- Ocean dynamics consistency

These diagnostics help determine whether models reproduce physically meaningful behaviour rather than simply matching observed fields.

Key Results

Several noteworthy findings emerged from the benchmark:

- AI models performed **competitively with operational numerical systems** when compared directly with observations.
- Machine-learning models often showed particularly strong performance in **subsurface variables and ocean current prediction**.
- Surface-layer variables remained more challenging, where numerical models frequently retained an advantage.
- Compared with analyses and reanalyses, numerical models generally performed best overall, but some AI models showed reduced forecast error growth at longer lead times.
- Direct forecast models such as XiHe exhibited less accumulation of autoregressive error than recursive forecasting approaches.
- Process-based diagnostics revealed differences in physical consistency that were not apparent from simple RMSE scores alone.

Commented [1]: @anas.elaouni@gmail.com I assume this is meant to be XiHe (Wang et al), unless I'm missing something....
Assigned to anas.elaouni@gmail.com

The results emphasized the importance of evaluating both forecast accuracy and physical realism.

Demonstration of OceanBench

A major portion of the workshop was devoted to a live demonstration. Anass showed how users can:

- Install OceanBench through Python or command-line interfaces.
- Access benchmark datasets.
- Define new model configurations.
- Generate automated evaluation reports.

The system produces comprehensive notebooks and reports containing:

- Error statistics versus observations and reanalysis.
- Mixed-layer-depth diagnostics.
- Geostrophic current evaluations.
- Lagrangian trajectory comparisons.
- Spectral analyses.
- Visualization products.

A key design goal is that users should only need to specify the path to their forecast outputs, with OceanBench handling the remainder of the workflow.

New and Experimental Features

The team is developing additional diagnostics, including:

- Mesoscale eddy detection and tracking.
- Spectral energy analysis.
- Scale-dependent verification metrics.
- Regional benchmarking capability.
- Improved visualization and interactive dashboards.

The framework already supports regional subsetting, enabling the same evaluation pipeline to be applied to user-defined domains rather than only global products.

Future Development

Planned future releases include:

- Additional benchmark years (2023 and 2025).
- Inclusion of more numerical forecasting systems.
- New AI models such as Fuji Ocean and other recently published architectures.
- Ensemble and probabilistic forecast benchmarks.
- Lower-resolution tracks for users with limited computing resources.
- Seasonal forecasting benchmarks.
- Sea-ice and biogeochemical variables.
- Expanded physical diagnostics, including eddy and spectral metrics.

Discussion and Community Feedback

The discussion emphasized several important themes:

- The need for **multiple independent reanalyses** to avoid bias toward any single reference product.
- Careful documentation of verification methodologies and metric definitions.
- Consistent treatment of masking, grids, and conventions.
- The importance of community governance for adding new diagnostics and datasets.

Participants welcomed the framework's openness and extensibility, while also stressing that transparency in evaluation procedures is essential for ensuring fair comparisons.

Main Takeaway

OceanBench represents the first comprehensive, community-oriented benchmarking platform for AI and numerical ocean forecasting. By combining standardized datasets, operational evaluation pipelines, physical diagnostics, and open-source tools, it aims to provide the ocean forecasting community with a common framework analogous to WeatherBench and to accelerate the development, comparison, and validation of next-generation ocean prediction systems.

4. Breakout Sessions

Breakout sessions were held using Slido to collect immediate responses to several questions. Discussions were held in small groups, with thoughts and comments entered in real-time into slido, allowing groups to see each other's responses, and discuss and 'react' to these. A summary of the discussion and responses follows. The full responses are given in appendix D, with a summary of the two sessions here.

Breakout 1: Expectations and requirements for AI in Ocean prediction, and how can the task team best support this

Breakout session 1 was focused on the broad role of AI in ocean prediction over the coming years. With 3 focal questions:

- What do you believe will be the most beneficial applications of AI in ocean prediction, considering both short-term and long-term potential?
- What are your biggest concerns regarding AI and ocean prediction?
- What objectives should the task team pursue over the next five years?

Participants agreed that, in the short term, the most valuable role for AI is to enhance existing ocean prediction systems rather than replace them. High-impact applications include improving uncertainty estimation through fast ensemble methods, post-processing and bias correction using high-resolution and in situ data, and addressing expensive elements of data assimilation such as error covariance modelling. AI also shows clear potential for improving marine hazard prediction, particularly for events i.e. storm surges and harmful algal blooms.

There was also a focus on higher-resolution and more integrated modelling. AI is expected to play a key role in modelling coastal processes, through downscaling and/or reducing computational costs through hybrid AI–physical models, i.e. improving parameterisations of poorly resolved processes such as ocean mixing. There was interest in extending these approaches to coupled systems, including biogeochemistry, and in developing decision-support tools for risk assessment and operational applications. A consistent message was that AI should complement and augment physically based approaches.

On the second prompt, concerns were raised around trust, data limitations, and the preservation of domain expertise. Many participants highlighted the risk of losing physical understanding if AI is adopted uncritically, alongside the challenge of limited and uneven observational data—especially in coastal regions. The "black box" nature of many AI models raises issues of explainability and transparency, which in turn affect user confidence. In addition, current validation and verification practices are seen as insufficient, with a need for more rigorous and standardised approaches to assess performance and ensure physical consistency.

Other risks include overfitting due to limited training data, difficulties in transferring models across regions or regimes, and the computational and environmental costs of large-scale AI systems. There is also a strong emphasis on the need to maintain and expand observational networks, as ML systems ultimately depend on the quality and coverage of underlying data.

When discussing the role of the Ocean Predict Task Team, participants emphasised the importance of coordinated community efforts over the next five years. Priorities included enabling non-ML experts to access and use ML tools, developing shared benchmarking and intercomparison platforms, improving data sharing and standardisation, and establishing best practices for use of ML. Strengthening collaboration between ML specialists and ocean scientists was seen as essential, as was investing in training and capacity building across different user groups.

Breakout 2: Benchmarking, assessment and validation of ML models

Breakout session 2 focused on benchmarking, assessment and validation of machine learning models, with three guiding questions:

- Should the assessment of machine learning models differ from the evaluation of physical models, and if so, how?
- What are the most effective ways to compare machine learning and physics-based models in terms of useful metrics, studies and assessment criteria?
- How can we prevent machine learning models from exploiting our designated metrics or case studies to "game the system"?

Participants agreed that assessing machine learning models requires a broader approach than simply applying traditional verification metrics. While there was recognition that many existing evaluation methods remain valuable, there was a strong consensus that statistical measures such as RMSE and bias alone are insufficient. Instead, evaluation should consider whether models are fit for their intended purpose, represent the underlying physics realistically, and provide meaningful estimates of uncertainty. Many responses emphasised the need to assess physical consistency, including conservation properties, realistic energy budgets, representation of key ocean processes, continuity across space and time, and performance under previously unseen conditions. Explainability, reproducibility and physical plausibility were identified as important aspects of model evaluation, particularly where machine learning models are intended to support scientific understanding rather than simply maximise predictive skill.

When discussing comparisons between machine learning and physics-based models, participants highlighted the importance of moving beyond generic verification metrics towards application-driven assessments. There was strong support for evaluating models using downstream products and user-relevant outcomes, such as trajectories, frontal locations, mixed-layer depth and geostrophic currents, alongside their ability to support operational forecasting and decision making. Several participants noted that assessment criteria should depend on the scientific question being addressed and the region or application of interest. Benchmark datasets, common evaluation frameworks and well posed challenges were widely viewed as essential for enabling fair comparisons between different modelling approaches, while stress-testing models under challenging scenarios and comparing their behaviour using common reference datasets were also identified as valuable approaches.

The discussion on preventing machine learning models from "gaming the system" centred on ensuring that benchmark performance reflects genuine scientific capability rather than optimisation for a limited set of metrics. Participants stressed the importance of maintaining strict independence between training, validation and test datasets, including separation in both space and time, and ensuring that validation periods are never reused during model development. There was broad support for incorporating process-based and physics-informed assessments alongside statistical scores, evaluating performance on challenging and extreme events, and testing models under conditions outside their training distribution. Participants also highlighted the need for improved observational and reference datasets, shared best practices and community guidelines, drawing where appropriate on experience from software engineering and data assimilation. Overall, there was strong agreement that robust benchmarking should provide confidence that machine learning models are scientifically credible, transferable and capable of delivering reliable performance in real-world ocean prediction applications.

5. Overall workshop conclusion

The workshop demonstrated that AI is becoming an increasingly important component of ocean prediction science. Significant advances are being made in forecasting, state reconstruction, downscaling, environmental monitoring, and Earth-system emulation. AI methods already offer substantial computational advantages and, in many cases, skill comparable to traditional approaches.

However, the meeting also highlighted a clear consensus that future progress requires more than predictive accuracy alone. Physical consistency, interpretability, uncertainty estimation, and rigorous evaluation must remain central research priorities. The emergence of OceanBench provides a critical foundation for community-wide assessment and comparison of future systems.

Overall, participants concluded that the most promising path forward lies not in replacing physics-based modelling, but in developing hybrid AI–physics approaches that combine the strengths of both paradigms to improve the accuracy, efficiency, and scalability of future ocean forecasting systems.

Appendices

Appendix A: Meeting agenda overview

13th April - Day 1

09:15 Arrival and registration (Coffee on arrival)

09:40 Welcome, Gregory Smith and Rachel Furner

Block 1 talks: Ocean Emulators

Chair: Rachel Furner (ECMWF), Rapporteur: Anass El Aouni (MOI)

09:50 Toward Scalable and Probabilistic Neural Ocean Forecasting, ***Simon Van Gennip*** *(Mercator Ocean International)*

10:10 Application of Deep learning (DL) in the Gulf of St. Lawrence and Estuary, ***Francois Roy*** *(ECCC)*

10:30 AI-driven Nanming Model for Three-dimension Ocean Variable Forecasting, ***Xiaoyan Li*** *(Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai)*

10:50 Prediction of sea surface currents around the Korean peninsula using artificial neural networks, ***Jae-Hun Park*** *(Department of Ocean Sciences, Inha University)*

11:15 Coffee

11:45 **Discussion session 1: Expectations and requirements for AI in Ocean prediction, and how can the task team best support this**

Chair: Gregory Smith (ECCC), Rapporteur: Rachel Furner (ECMWF)

12:45 Lunch - self pay at local restaurants

Posters

14:00 Poster flash talks: 2 mins from each author to introduce their poster.

14:15 Poster viewing and coffee

Chair: Anass El Aouni (MOI)

Filling the Ocean's Gaps: a Self-Supervised Neural Network for Argo Profiles Data Augmentation, ***Teresa Tonelli,*** *(OGS, University of Trieste)*

Assimilation of sea surface temperature in the Mediterranean Sea using ML-based operators, ***Daniele Bigoni*** *(CMCC Foundation, Italy)*

Detection and Discrimination of Marine Oil Spills and Look-alike Phenomena in Synthetic Aperture Radar Imagery, ***Xudong Huang*** *(Department of Oceanography, Dalhousie University)*

Machine Learning Forecast Correction for Geophysical Fields, ***Jacopo Dall'Aglio*** *(University of Bologna)*

Bridging Scales in Mediterranean Biogeochemical Prediction: A High-Order Ensemble Assimilation Coupled with AI-Driven Downscaling, ***Simone Spada*** *(OGS)*

Data-driven ocean modelling at ECMWF, ***Rachel Furner*** *(ECMWF)*

Block 2 talks: Observations and Data Assimilation

Chair: Jae-Hun Park (Inha University), Rapporteur: Vikram Khade (ECCC)

15:15 Training end-to-end neural mapping schemes from simulation data for the reconstruction of global-scale sea surface fields, ***Daniel Zhu*** *(IMT Atlantique)*

15:35 Short-term neural forecasts of ocean dynamics from sparse satellite observations, ***Daria Botvynko*** *(IMT Atlantique)*

15:55 Two-Phase CNN for Model Data Fusion: Predicting 3D Chlorophyll-a in the Mediterranean Sea, ***Teresa Tonelli*** *(OGS, University of Trieste)*

16:15 A Deep-Learning Observation Operator for Subsurface Thermohaline Reconstruction from Satellite Surface Observations, ***Geon Min Lee*** *(Pukyong National University, Republic of Korea)*

16:35 AI perform on high resolution three-dimensional ocean forecasting: remote sensing data driven becomes a new possibility, ***Liying Wan*** *(National Marine Environmental Forecasting Center) (recorded presentation)*

17:00 Close of day 1

18:30 Optional (self-pay) dinner at **Trois Brasseurs on St. Denis**

14th April - Day 2

Block 3 talks: Emulators

Chair: Charlie Hebert-Pinard (ECCC), Rapporteur: Daniele Bigoni (CMCC)

09:00 Linear Stochastic Emulators of the Ocean Circulation - A Lesson, perhaps, for Machine Learning, ***Andrew Moore*** *(University of California Santa Cruz)*

09:20 A data driven limited area storm surge model, ***Mateusz Matuszak*** *(Norwegian Meteorological Institute)*

09:40 A Physics Informed Emulator for Ocean Oxygen, ***Annalisa Bracco*** *(CMCC)* *(remote)*

10:00 Application and Verification of the Global Wave Intelligent Forecast Model, ***Fang Hou*** *(National Marine Environmental Forecasting Center)* *(recorded presentation)*

10:20 Group photo

10:30 Coffee

Keynote

Chair: Fraser Davidson (ECCC), Rapporteur: Kristian Mogensen (ECMWF)

11:15 OceanBench: A Benchmark for Data-Driven Global Ocean Forecasting systems, *Anass El Aouni (Mercator Ocean International)*

12:15 Lunch - self pay at local restaurants

Block 4 talks: Downscaling, and other ML applications

Chair: Frederic Dupont (ECCC), Rapporteur: Simon Corbeil-Letourneau (ECCC)

13:20 Development of machine-learning emulators for harbour-scale ocean prediction, ***Michael Dunphy*** *(Institute of Ocean Sciences, Fisheries and Oceans Canada)*

13:50 From Coarse Models to Coastal Detail: A Deep Learning Approach to AI based Statistical Downscaling in the Adriatic Sea, ***Alessandro De Lorenzis*** *(CMCC Foundation - Euro-Mediterranean Center on Climate Change)* *(remote)*

14:10 Integrated AI_Physics Approaches for Coastal Prediction Across the Open-to-Coastal Ocean Continuum, ***Joanna Staneva*** *(Helmholtz Zentrum HEREON)*

14:30 Fix the double penalty in data-driven forecasting by modifying the loss function, ***Christopher Subich*** *(Environment & Climate Change Canada)*

14:50 Variational autoencoder-based clustering for geophysical fluid circulations with small sample size, ***Kunihiro Aoki*** *(Meteorological Research Institute, Japan Meteorological Agency)*

15:15 Coffee

15:50 **Discussion session 2: Benchmarking, assessment and validation of ML models**

Chair: Rachel Furner (ECMWF), Rapporteur: Simon Van Gennip (MOI)

16:50 Workshop wrap up

17:00 Close

Appendix B: Oral and poster presentations (all sessions)

Keynote presentation

Chair: Fraser Davidson (ECCC), Rapporteur: Kristian Mogensen (ECMWF)

OceanBench: A Benchmark for Data-Driven Global Ocean Forecasting systems	Anass El Aouni	Mercator Ocean International	Recording

Block 1 talks: Ocean Emulators

Chair: Rachel Furner (ECMWF), Rapporteur: Anass El Aouni (MOI)

Toward Scalable and Probabilistic Neural Ocean Forecasting (not avialable)	Simon van Gennip	Mercator Ocean International, France	Recording
Application of Deep learning (DL) in the Gulf of St. Lawrence and Estuary	Francois Roy	ECCC, Canada	Recording
Prediction of sea surface currents around the Korean peninsula using artificial neural networks	Jae-Hun Park	Department of Ocean Sciences, Inha University, Republic of Korea	**Recording**

Block 2 talks: Observations and Data Assimilation

Chair: Jae-Hun Park (Inha University), Rapporteur: Vikram Khade (ECCC)

Training end-to-end neural mapping schemes from simulation data for the reconstruction of global-scale sea surface fields	Daniel Zhu	IMT Atlantique, France	**Recording**

Short-term neural forecasts of ocean dynamics from sparse satellite observations	Daria Botvynko	IMT Atlantique, France	Recording
Two-Phase CNN for Model Data Fusion: Predicting 3D Chlorophyll-a in the Mediterranean Sea	Teresa Tonelli	OGS, University of Trieste, Italy	Recording
A Deep-Learning Observation Operator for Subsurface Thermohaline Reconstruction from Satellite Surface Observations	Geon Min Lee	Pukyong National University, Republic of Korea	**Recording**
AI perform on high resolution three-dimensional ocean forecasting: remote sensing data driven becomes a new possibility	Liying Wan	NMEFC, China	**Recording**

Block 3 talks: Emulators

Chair: Charlie Hebert-Pinard (ECCC), Rapporteur: Daniele Bigoni (CMCC)

Linear Stochastic Emulators of the Ocean Circulation – A Lesson, perhaps, for Machine Learning	Andrew Moore	University of California Santa Cruz, US	**Recording**
A data driven limited area storm surge model	Mateusz Matuszak	Norwegian Meteorological Institute, Norway	**Recording**
A Physics Informed Emulator for Ocean Oxygen *(remote presentation)*	Annalisa Bracco	CMCC, Italy	**Recording**
Application and Verification of the Global Wave Intelligent Forecast Model	Fang Hou	NMEFC, China	**Recording**

Block 4 talks: Downscaling, and other ML applications

Chair: Frederic Dupont (ECCC), Rapporteur: Simon Corbeil-Letourneau (ECCC)

Development of machine-learning emulators for harbour-scale ocean prediction	Michael Dunphy	Institute of Ocean Sciences, Fisheries and Oceans Canada	**Recording**
From Coarse Models to Coastal Detail: A Deep Learning Approach to AI based Statistical Downscaling in the Adriatic Sea *(remote presentation)*	Alessandro De Lorenzis	CMCC, Italy	Recording
Integrated AI_Physics Approaches for Coastal Prediction Across the Open-to-Coastal Ocean Continuum	Joanna Staneva	Helmholtz Zentrum HEREON, Germany	**Recording**
Fix the double penalty in data-driven forecasting by modifying the loss function	Christopher Subich	ECCC. Canada	**Recording**
Variational autoencoder-based clustering for geophysical fluid circulations with small sample size	Kunihiro Aoki	Meteorological Research Institute, Japan Meteorological Agency	**Recording**

Posters

Chair: Anass El Aouni (MOI)

Filling the Ocean's Gaps: a Self-Supervised Neural Network for Argo Profiles Data Augmentation	Teresa Tonelli	OGS, University of Trieste, Italy	Flash talk: **Pdf** and **recording**
Assimilation of sea surface temperature in the Mediterranean Sea using ML-based operators	Daniele Bigoni	CMCC, Italy	Flash talk: Pdf and **recording**
Bridging Scales in Mediterranean Biogeochemical Prediction: A High-Order Ensemble	Simone Spada	OGS, Italy	Flash talk: **Pdf** and **recording**

Assimilation Coupled with AI-Driven Downscaling			
Data-driven ocean modelling at ECMWF	Rachel Furner	ECMWF	Flash talk: **Pdf** and **recording**
Physics-Informed Neural Data Assimilation for High-Resolution: Coastal SPM Reconstruction from Model and Satellite data	Wei Chen	Helmholtz-Zentrum Hereon	No flash talk or recording available
Toward an AI-enhanced hydro-morphodynamic model for nature-based solutions in coastal erosion mitigation *(abstract only)*	Nour Dammak	Helmholtz-Zentrum Hereon	No flash talk or recording available

Appendix C: Meeting attendees

(we might be missing names of some online participants as we do not hold all the information)

	First name	Last name	Affiliation
1	Saima	Aijaz	Bureau of Meteorology
2	Ahmadreza	Alleosfour	DFO/MPO
3	El Aouni	Anass	Mercator Ocean International
4	Kunihiro	Aoki	MRI-JMA
5	Federica	Benassi	University of Bologna
6	Daniele	Bigoni	CMCC Foundation
7	Bing	Yuan	Helmholtz Zentrum HEREON
8	2nd Lt.	Blanka	Brazilian Navy
9	Daria	Botvynko	IMT Atlantique
10	Annalisa	Bracco	CMCC Foundation
11	Simon-Philippe	Breton	Environment and Climate Change Canada
12	Eric	Chassignet	Florida State University
13	Wei	Chen	Helmholtz Zentrum HEREON
14	Kamel	Chikhar	Environment and Climate Change Canada
15	Byoung-Ju	Choi	Chonnam National University
16	Emanuela	Clementi	CMCC Foundation
17	Simon	Corbeil-Letourneau	Environment and Climate Change Canada
18	Jacopo	Dall'Aglio	University of Bologna
19	Nour	Dammak	Helmholtz Zentrum HEREON
20	Joshua	DaRosa	The MITRE Corporation
21	Fraser	Davidson	Environment and Climate Change Canada
22	Alessandro	De Lorenzis	CMCC Foundation
23	Pierre	De Mey-Frémaux	LEGOS
24	Prasanth	Divakaran	Bureau of Meteorology
25	Craig	Donlon	ESA
26	Marie	Drevillon	Mercator Ocean International
27	Yann	Drillet	Mercator Ocean International
28	Michael	Dunphy	Fisheries and Oceans Canada
29	Frederic	Dupont	Environment and Climate Change Canada
30	Anass	El Aouni	Mercator Ocean International
31	Mikhail	Entel	Bureau of Meteorology
32	Ivan	Federico	CMCC Foundation
33	Luiz Claudio	Fonseca	Brazilian Navy
34	David	Ford	Met Office
35	Yosuke	Fujii	JMA/MRI
36	Rachel	Furner	ECMWF
37	Isabelle	Gaboury	Fisheries and Oceans Canada

38	Yao	Gahounzo	Florida State University
39	Audrey-Anne	Gauthier	Environment and Climate Change Canada
40	Stephanie	Guinehut	Mercator Ocean International
41	Yukie	Hata	Environment and Climate Change Canada
42	Charlie	Hebert-Pinard	Environment and Climate Change Canada
43	Fabrice	Hernandez	IRD/LEGOS
44	Dante	Horemans	Virginia Institute of Marine Science
45	Fang	Hou	National Marine Environmental Forecasting Center
46	Xudong	Huang	Dalhousie University
47	Mo Rokibul	Islam	Environment and Climate Change Canada
48	Ho-Jeong	Ju	Inha University
49	Sarah	Keeley	Met Office
50	Vikram	Khade	MSC, ECCC
51	Nam-Hoon	Kim	KIOST
52	Robert	King	Met Office
53	Jae-Il	Kwon	KIOST
54	Yvonnick	Le Clainche	Fisheries and Oceans Canada
55	Pierre Yves	Le Traon	Mercator Ocean International
56	Geon-Min	Lee	Pukyong National University
57	Xiaoyan	Li	Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai)
58	Yaru	Li	Plymouth Marine Laboratory
59	Bataa	Lkhagvasuren	Chonnam National University
60	Robert	Ma	DFO/MPO
61	Nadim	Mahmud	Oceanprediction DCC
62	Simona	Masina	CMCC Foundation
63	Mateusz	Matuszak	MET Norway
64	Kristian	Mogensen	ECMWF
65	Andrew	Moore	University of California Santa Cruz
66	Sophia	Moreton	Met Office
67	Olmo	Zavala-Romero	Florida State University
68	Jae-Hun	Park	Inha University
69	Patricia	Pernica	Fisheries and Oceans Canada
70	Vinícius	Pessanha	Brazilian Navy
71	Andrew	Peterson	Environment and Climate Change Canada
72	Keshav	Raja	Florida State University
73	Sthitapragya	Ray	Florida State University
74	Charly	Regnier	Mercator Ocean International
75	Hal	Ritchie	Environment and Climate Change Canada
76	Francois	Roy	Environment and Climate Change Canada

77	Leonardo	Saccotelli	CMCC Foundation
78	Gregory	Smith	Environment and Climate Change Canada
79	Simone	Spada	National Institute of Oceanography and Applied Geophysics - OGS
80	Emil	Stanev	Helmholtz Zentrum HEREON
81	Joanna	Staneva	Helmholtz Zentrum HEREON
82	Andrea	Storto	CNR
83	Christopher	Subich	Environment and Climate Change Canada
84	Yuhei	Takaya	MRI-JMA
85	Beniamino	Tartufoli	University of Bologna
86	Anna	Teruzzi	OGS (National Institute of Oceanography and Applied Geophysics)
87	Charles-Emmanuel	Testut	Mercator Ocean International
88	Ricardo	Todling	NASA/GSFC
89	Hendrik	Tolman	NOAA / National Weather Service
90	Teresa	Tonelli	OGS (National Institute of Oceanography and Applied Geophysics)
91	Ricardo	Torres	Plymouth Marine Laboratory
92	Joshua	Turner	Environment and Climate Change Canada
93	Simon	Van Gennip	Mercator Ocean International
94	Liying	Wan	National Marine Environmental Forecasting Center
95	Anthony	Weaver	CERFACS
96	Kirsten	Wilmer-Becker	Met Office
97	Chunxue	Yang	CNR-ISMAR
98	Irem	Yildiz	Helmholtz Zentrum HEREON
99	Jorge Eduardo	Velasco Zavala	Florida State University
100	Daniel	Zhu	IMT Atlantique