Program AI4OAC'2023 workshop

Workshop on AI for Ocean, Atmosphere and Climate Dynamics

11-13 April 2023 Brest (France)

Workshop venue for in-person participants: PNBI, Brest (google map)

- Room 207 (3rd floor) for keynote and tutorial sessions
- Rooms 207, 113 and 127 for working groups (2nd and 3rd floors)
- Coffee breaks, lunch and oyster coktail (1st floor)

Webex links below for remote attendees:

- **Tuesday, April 11th, 2023:** link
- **Wednesday, April 12th, 2023:** link
- **Thursday, April 13th, 2023:** link

Tuesday, April 11th, 2023

Video-conference link: https://imt-atlantique.webex.com/imt-atlantique/j.php?MTID=mec4ebf7e2cb36dbe2bfa731d5a30f432

1.30pm-1.50pm	**Welcome coffee & tea**
1.50pm-2.00pm	**Welcome word** (R. Fablet)
2.00pm-4.00pm	**Keynote session on ML for climate model calibration** *Chairs: R. Lguensat and J. Brajard* **R. Roehrig.** Addressing the calibration bottleneck using machine learning: application to the CNRM-CM6-1 climate model (abstract) **K. Haynes**. AI Uncertainty Quantification: An Introduction to Approaches for Creating and Evaluating Uncertainty Estimates Using Neural Networks. (abstract)
4.00pm-4.30pm	**Tea & Coffee time**
4.30pm-5.00pm	**2-minute poster pitches #1 (see list below)**
5.00pm-5.30pm	**Presentation of the proposed working groups' themes**
5.30pm-6.30pm	**Poster session**
6.30pm-8.00pm	**Oyster cocktail**

Wednesday, April 12th, 2023

Video-conference link:
https://imt-atlantique.webex.com/imt-atlantique/j.php?MTID=m7d394d0bbb4752e310b4768fc526d652

9.30am-10.45am	**Tutorial on Diffusion Models.** *F. Rousseau* **(**abstract**)**
10.45am-11.00am	**Tea & Coffee time**
11.00am-12.30pm	**Working group sessions** (in parallel) ● Learning-based Emulators for OAC Processes ● Machine Learning, Data Assimilation, Interpolation and ESM
12.30pm-2.00pm	**Lunch**
2.00pm-4.00pm	**Keynote session on ML for climate data analytics** *Chair: P. Naveau* **L. Raynaud**. About Machine Learning to explore and enhance ensemble weather forecasts : an overview of current activities at CNRM. (abstract) **M. Demangeot**. Insight into spatial extremes and Geostatistics (abstract)
4.00pm-4.30pm	**Tea & Coffee time**
4.30pm-6.00pm	**Working group session** (in parallel) ● Neural Closures for Earth System Models ● Model Calibration and Uncertainty Quantification in ESM ● Deep Learning, Extreme Value Theory and OAC processes

Thursday, April 13th, 2023

Video-conference link:
https://imt-atlantique.webex.com/imt-atlantique/j.php?MTID=m576b5d781145559d53de22b5fc155cbd

9.30am-10.45am	**Tutorial on Implicit Representation Learning.** *J.E. Johnson (abstract)*
10.45am-11.00am	**Tea & Coffee time**
11.00am-12.30pm	**Working group sessions** (in parallel)
12.30pm-2.00pm	**Lunch**
2.00pm-4.00pm	**Keynote session on ML for model discovery** *Chairs: F. Bouchet* **T. Beucler**. Physically and Causally-Informed Machine Learning for Atmospheric Convection. (abstract) **P. Hassanzadeh**. Learning Data-driven Subgrid-scale Parameterizations: Stability, Extrapolation, and Interpretation (abstract)
4.00pm-4.30pm	**Tea & Coffee time**
4.30pm-5.30pm	**Wrap-up session**

Workshop Keynotes

Addressing the calibration bottleneck using machine learning: application to the CNRM-CM6-1 climate model. Romain Roehrig, CNRM, Meteo France.

Atmospheric models, used for either weather or climate applications, encompass so-called parameterizations, which aims at summarizing and quantifying the impact on the resolved model variables of radiative, thermodynamical, or chemical processes, as well as dynamical processes that occur at scales smaller than the computational grid. Though these parameterizations are developed on a physical basis, some simplifications underlying them introduce parameters that need to be properly calibrated to achieve a skillful model. Given the number of parameters (several tens), the possible number of performance metrics used for validation or evaluation, and the computational cost of models, the modelling experts need help to better address this calibration bottleneck. In this work, we experiment the history matching with iterative refocusing framework with the CNRM-CM6-1 climate model to assess whether the model current deficiencies are related to poor model calibration, or if they critically rely on the scientific content of the model. Using a rather small physics perturbed ensemble of short simulations as a learning dataset, Gaussian processes are used to cheaply explore the full space of model parameters and identify the part of it, which provides model configurations compatible with references, given a set of usual performance metrics and the various sources of uncertainty in the whole process. The calibration framework also builds on several waves of true model simulations, to parsimoniously increase the size of the learning dataset only where the surrogate model uncertainty needs to be reduced. We show that several new configurations, in which many CNRM-CM6-1 biases are significantly reduced or even removed (e.g., precipitation over West Africa, regional biases in cloud radiative effect), can be found. Though, some CNRM-CM6-1 biases are truly structural (e.g., biases over eastern sides of tropical ocean basins), calling for further understanding and parameterization development.

AI Uncertainty Quantification: An Introduction to Approaches for Creating and Evaluating Uncertainty Estimates Using Neural Networks. K. Haynes. CIRA, CSU. (Joint work with Ryan Lagerquist, Marie McGrawa, Kate Musgravea, and Imme Ebert-Uphoff**)**

Due to the availability of large amounts of simulation and observation data in ocean-atmosphere-climate science, AI methods can be useful for exploring open questions in these scientific fields. Due to the increasing impacts of climate change, results from this community are vital in informing policy decisions. Given the critical nature of these applications, it is increasingly important to obtain uncertainty estimates along with the results. The computer science community has recently made significant advances in developing methods for neural networks that allow these models to provide uncertainty estimates for both prediction and classification tasks. The goal of this talk is to provide an accessible introduction to AI-based uncertainty quantification, focusing on four key questions: (1) What uncertainties are we tasking our machine-learning (ML) models to estimate? (2) What are some simple approaches that can be implemented in neural network models to create uncertainty estimates? (3) Once we obtain uncertainty estimates, how do we know whether they are any good? (4) How do we best communicate the uncertainty estimates?

This presentation discusses these topics and illustrates several common approaches and evaluation metrics using a real-world atmospheric science application. We hope that this introduction will pave the way for researchers in this community to quickly start incorporating uncertainty estimates in their applications.

About Machine Learning to explore and enhance ensemble weather forecasts : an overview of current activities at CNRM. Laure Raynaud. CNRM, MEteo France.

Ensemble forecasts have become a major tool to anticipe high-impact weather events and their uncertainties. However, their effective use is currently limited by two main aspects. First, their configuration, using kilometer-scale resolution and about twenty members for state-of-the-art systems, is generally not sufficient to accurately predict location, intensity and timing of events. Secondly, summarizing this ensemble information in a relevant and user-oriented way is still a challenge. In recent years, Machine Learning (ML) has been considered as a potential solution to address these two subjetcs. Several post-processing tools relying on feature extraction and dimension reduction have been developed to better explore ensembles. Going toward hybrid Physics/ML forecasts is another avenue of improvement to significantly increase ensemble size and resolution at almost no additional cost.

Insight into spatial extremes and Geostatistics. M. Demangeot. Univ. Montpellier.

Spatial and multivariate extreme value theory helps model and predict the frequency of extreme events in a spatial context like, for instance, extreme precipitations or extreme temperatures. In such framework, the estimation of small occurrence probabilities of rare events is not carried out on the entire sample, but on some relatively small number of largest observations that are considered representative of the tail of the distribution. Until the last decade, the major part of the extreme value theory literature has thus been dedicated to the moderate dimensional setting. Motivated by the growing availability of large databases, the most recent work in this area has focused on connecting the study of multivariate and spatial extremes to statistical learning techniques in order to develop sparsity-based methods. After an overview of the well-established approaches used to study spatial and multivariate extremes, we shall focus on these new procedures and in particular on graphical models for extremes.

Physically and Causally-Informed Machine Learning for Atmospheric Convection. Tom Beucler. Univ. Lausanne. (Joint work with Fernando Iglesias-Suarez, Veronika Eyring, Pierre Gentine, Michael Pritchard, Jakob Runge)
Data-driven algorithms, in particular neural networks, can emulate the effects of unresolved processes in coarse-resolution Earth system models (ESMs) if trained on high-resolution simulation or observational data. However, they can (1) make large generalization errors when evaluated in conditions they were not trained on; and (2) trigger instabilities when coupled back to ESMs. First, we propose to physically rescale the inputs and outputs of neural networks to help them generalize to unseen climates. Applied to the offline parameterization of subgrid-scale thermodynamics (convection and radiation) in three distinct climate models, we show that rescaled or "climate-invariant" neural networks make accurate predictions in test climates that are 8K warmer than their training climates. Second, we propose to eliminate spurious causal relations between inputs and outputs by using a recently developed causal discovery framework (PCMCI). For each output, we run PCMCI on the inputs time series to identify the reduced set of inputs that have the strongest causal relationship with the output. Preliminary results show that we can reach similar levels of accuracy by training one neural network per output with the reduced set of inputs; stability implications when coupled back to the ESM are explored. Overall, our results suggest that explicitly incorporating physical knowledge into data-driven models of Earth system processes may improve their ability to generalize across climate regimes, while quantifying causal associations to select the optimal set of inputs may improve their consistency and stability.

Learning Data-driven Subgrid-scale Parameterizations: Stability, Extrapolation, and Interpretation P. Hassanzadeh. Rice Univ.
The Earth system involve a variety of nonlinearly interacting physical processes spanning a broad range of spatial and temporal scales. To make simulations of the Earth system accurate while computationally tractable, processes with scales smaller than the typical grid size of general circulation models (GCMs) have to be parameterized. Recently, there has been substantial interest (and progress) in using machine learning (ML) to develop data-driven subgrid-scale (SGS) parameterizations for a number of key processes in the atmosphere, ocean, and other components of the Earth system. However, for these data-driven SGS parameterizations to be useful and reliable in practice, a number of major challenges have to be addressed. These include: 1) instabilities arising from the coupling of data-driven SGS parameterizations to coarse-resolution solvers, 2) learning in the small-data regime, 3) interpretability, and 4) extrapolation to different parameters and forcings. Using several setups of 2D turbulence, as well as two-layer quasi-geostrophic turbulence and Rayleigh-Benard convection as test cases, we introduce methods to address (1)-(4). These methods are based on combining physics and recent theoretical and applied advances in ML. For example, we will use backscattering analysis to shed light on the source of some of the instabilities and incorporate physical constraints to enable learning in the small-data regime. We will further introduce a novel framework based on spectral analysis of the neural network to interpret the learned physics and will show how transfer learning enables extrapolation to flows with very different physical characteristics. We will also briefly mention some of the advances in supervised and semi-supervised learning of the SGS models, as well as the use of equation-discovery techniques. In the end, we will discuss scaling up these methods to more complex systems and real-world applications, e.g., for SGS modeling of atmospheric gravity waves.

Workshop Tutorials

Introduction to diffusion models. François Rousseau. IMT Atlantique.

Diffusion models have emerged as a powerful new family of deep generative models. They have achieved remarkable performance in many applications, including image synthesis and video generation. In this presentation, we will provide an overview of the rapidly growing body of work on diffusion models, looking at key areas such as efficient sampling or likelihood estimation. Time will be devoted to a hands-on session to illustrate the use of diffusion models on toy examples and images.

Introduction to Implicit representation learning. J. Emmanuel Johnson. CNRS/UGA.

Neural Fields (NerFs) are an emerging class of coordinate-based neural networks. There has been many developments in the last few years for applying NerFs to data like images. In this tutorial, I will introduce NerFs from the geoscience perspective and highlight some potential advantages to using these methods. I will demonstrate some concrete work on sea surface height interpolation and highlight some of the problems (and potential solutions) I faced when applying this class of methods to spatiotemporal data.

Themes for Working groups

Working group on Neural Closures for Earth System Models (ESM). Facilitators: F. Bouchet and F. Sevellec

This working group will focus on the design of sub-grid closures for atmosphere and ocean models. A key challenge in climate science is to reduce uncertainties in projections and the study of possible abrupt climate changes. In physics based models, these uncertainties are primarily caused by the parameterization of sub-grid parameterization, for instance clouds, boundary layer exchanges, subgriscale turbulence, which are based on heuristic assumptions. During the last decades a hope has risen to use the large datasets coming either from short-term high-resolution simulations which are now possible, or the fast increasing amount of observation datasets, in order to learn some key parameterizations directly from data. The aim of this working group will be to discuss the state of the art of machine learning for subgridscale parameterization, to exchange about the key literature, to discuss the current lines of research, the expected breakthrough and possible pitfalls of this approach. Some key questions to be discussed: which datasets, how to foresee the implementation of machine learning based parameterizations, which level of modularity for the implementation, how to validate and test the new paramerization, is there a need for a community or national scale organization?

Working group on Model Calibration and Uncertainty Quantification in ESM. Facilitators: R. Lguensat and R. Roehrig

Working group on Learning-based Emulators for OAC Processes. Facilitators: F. Rousseau, L. Raynaud and E. Martinez

The learning of complex AOC dynamics directly from data, considered as practically impossible a few years ago, is about to become a reality. Recent publications by Big Tech companies on purely data-driven weather forecasts have shown impressive results, with performances similar to state-of-the art NWP models for some variables. Such rapid progress raise a number of questions : should we try to reproduce such results, on our own or by collaborating with these companies, what about learning on higher resolution data, how should the community get organized to advance this subject ?

Working group on Deep Learning, Extreme Value Theory and OAC processes. Facilitators: P. Naveau and M. Demangeot

Working group on Machine Learning, Data Assimilation, Interpolation and ESM. Facilitators: M. Beauchamp, G. Tissot and J. Brajard

A growing interest has emerged in bridging interpolation methods, data assimilation (DA) schemes and machine learning frameworks to address limitations and challenges of data assimilation problems in earth systems. This includes both the integration of learning-based components in state-of-the-art DA and interpolation schemes as well as the design of DA-inspired learning-based schemes to address inverse problems and uncertainty quantification for dynamical processes. The goal of this working group will be to review recent advances on these topics and discuss expected breakthroughs and possible pitfalls compared to the state-of-the-art and operational DA systems. Most current demonstrations involve relatively simple case-studies (e.g., Lorenz systems, one-layer QG flows). The exploitation of learning-based schemes in operational systems probably may then benefit from the design of intermediate-complexity setups more representative of earth systems in terms of dimensionality, modeling uncertainties and biases, observing systems...

2-minute short presentations and Poster Session

Manon Langlet. ULCO. High-throughput imaging to observe in situ sinking behavior of marine particles: a new methodology for carbon fluxes assessment?

Daria Botvynko. ENIB/Lab-STICC Deep Learning for Lagrangian drift sumulation at the sea surface.

Robin Marcille. FEM. Ultra-short-term probabilistic forecasting of offshore wind speed for offshore wind maintenance operations.

Maxime Beauchamp. IMT Atlantique/Lab-STICC. Ensemble-based 4DVarNet uncertainty quantification for the reconstruction of Sea Surface Height dynamics

Anthony Frion. IMT Atlantique/Lab-STICC. Self-supervised learning of Sentinel-2 reflectance time series

Matteo Zambra. IMT Atlantique/Lab-STICC. Reconstruction of spatial wind speed fields with trainable variational data assimilation methods.

Quentin Febvre. IMT Atlantique/Lab-STICC. Scale aware neural calibration of SWOT data.

Perrine Bauchot. ENSTA Bretagne/Lab-STICC. Monitoring Regime Shifts in Chaotic Dynamics with Learning Variational Data Assimilation.

Jean Littaye. CNRS/Lab-STICC/LEMAR. Learning of carbon cycle models through emerging data from new devices and observing platforms.