

Compte-rendu de la journée thématique LEFE/MANU-PNTS IA & Océan-Atmosphère-Climat 6 février 2019, Rennes

Comité d'organisation : R. Fablet (IMT Atlantique, Lab-STICC), S. Thiria (Sorbonne Univ., LOCEAN), J. Brajard (Sorbonne Univ., LOCEAN), J. Le Sommer (CNRS, IGE), L. Terray (CERFACS, CECI)

Participants à la journée : 120 participants

Liste des laboratoires, instituts et entreprises représentés :

- **Instituts, universités et écoles** : CNRS, IMT Atlantique, INRIA, Ecole des Ponts ParisTech, CNES, Météo-France, CERFACS, SHOM, IRD, Sorbonne Univ., UGA, UR1, UR2, ULCO, ENSIIE, BRGM, Ifremer, UBO, ONERA, CEA, Ecole Polytechnique, GENCI, DGA, IRSTEA, INSA TLSE, ITR Bcom, UBS, UT3
- **Laboratoires du domaine IA** : INRIA Rennes, INRIA Grenoble, IRMAR, IRISA, Lab-STICC, LIP6
- **Laboratoires Océan-Atmosphère-Climat** : CECI, CNRM, LOPS, LEMAR, IGE, LOCEAN, LSCE, LEGOS, LMD, LETG, LATMOS, LISA, LISIC, ECTIA, LA
- **Participants internationaux** : Univ. of Washington, Princeton Univ., CAS, GeorgiaTech Univ.
- **Entreprises** : Scalian, Alyotech, CLS, Eodyn, Magellium, WeatherForce, Quantmetry

Programme de la journée : <https://www.imt-atlantique.fr/fr/conferences/ia-climat-ocean-atmosphere>, joint également en annexe ci-dessous

Soutien à l'organisation de la journée : LEFE/MANU, PNTS, IMT Atlantique

Synthèse des avis des participants (réponses de 20 participants) :

- Domaines de recherche : IA (24%), Modélisation physique (33%), Systèmes d'observation et télédétection (38%), Quantification d'incertitude (5%)
- Evaluation très positive du programme scientifique de la journée, y compris l'équilibre entre les aspects IA et les problématiques du domaine Océan-Atmosphère-Climat (> 90% d'avis "bon" à "excellent")
- Organisation de futures actions IA & Océan-Atmosphère-Climat : intérêt pour une participation à une future édition (100%) sur une base annuelle (60%) ou bisannuelle (40%) organisées sur plusieurs jours (80%) et possiblement à une échelle européenne (54%) plutôt qu'uniquement nationale (46%). A noter également un intérêt marqué pour la participation à des data challenges (56%)

Synthèse des échanges lors de la table-ronde :

- **Participants à la table-ronde** : P. Braconnot (INSU), G. Charpiat (INRIA), P. Dandin (Meteo France), J-P Proux (GENCI)
- **Potentiel & enjeux des approches IA pour les domaines Océan-Atmosphère-Climat**
 - Soutien des instituts et organismes représentés (Ifremer, Météo France, CNRS/INSU) pour le développement de cette interface pluridisciplinaire entre IA & les défis thématiques des domaines Océan-Atmosphère-Climat
 - Pertinence du développement d'approches "IA" exploitant au mieux les modèles, outils et méthodes numériques de l'état de l'art des domaines Océan-Atmosphère-Climat

- Nécessité d'anticiper les exploitations systématiques futures des technologies IA pour les programmes de recherche et des outils opérationnels des domaines Océan-Atmosphère-Climat en termes de compétences et métiers au sein des instituts et laboratoires (e.g., programmes de formation dédiés, cf. ci-dessous)
- **Structuration d'une communauté scientifique à l'interface entre IA & Océan-Atmosphère-Climat**
 - Volonté collective de soutenir l'émergence d'une communauté scientifique à l'interface entre IA & Océan-Atmosphère-Climat
 - Nécessité de développer et soutenir des actions d'animation dédiées (e.g., ateliers, "data challenges", projets collaboratifs...) permettant de croiser ces communautés. Proposition de renouveler l'organisation une journée thématique IA & Océan-Atmosphère-Climat en 2020.
 - Proposition d'actions dédiées de type "data challenge/benchmarking" permettant d'identifier des défis thématiques (possiblement idéalisés) par la mise en disposition (open data) de jeux de données référencées et des outils d'évaluation associés sur le modèle des pratiques générales de la communauté IA (e.g., sites de "benchmarking" en classification d'images)
- **Ressources de calcul et accès aux données**
 - Nécessité du développement de moyens de calcul dédiés (e.g., GPU) permettant d'exploiter et mettre en oeuvre les technologies de l'IA (e.g., apprentissage profond) sur des jeux de données (possiblement de type "big data") des domaines Océan-Atmosphère-Climat
 - Proposition d'une action concertée auprès des moyens de calcul nationaux (GENCI, calculateur Jean Zay) et des instituts et organismes portant l'IA (e.g., INRIA, INS2I) pour identifier les problématiques IA & Océan-Atmosphère-Climat (à la fois dans leurs dimensions de développement algorithmique spécifique (e.g., IA & systèmes dynamiques) et de domaine d'application de l'IA) dans le périmètre d'intérêt du calculateur Jean Zay
 - Nécessité d'interfacer les centres de données Océan-Atmosphère-Climat (e.g., pôle de données du Système Terre) avec les moyens de calcul dédiés, y compris pour favoriser le croisement entre masses de données d'observation et masses de données de simulation numérique
- **Actions de formation**
 - Importance de soutenir l'émergence des activités IA & Océan-Atmosphère-Climat par des actions de formation à différents niveaux (formation initiale Master/ingénieur, formation doctorale, formation continue). Possibilité d'une action concertée PNTS/LEFE dans ce contexte (eg, formation doctorale Data Science and Geoscience, <https://dsg2019.wordpress.com/>)

**Programme de la journée thématique PNTS-LEFE/MANU
IA & Océan-Atmosphère-Climat
IMT Atlantique, Campus de Rennes, 6 février 2017**

09h00-9h30	Accueil
09h30-10h00	Introduction de la journée (Animateur: R. Fablet)
10h00-12h45	Session A (Animateur: S. Thiria)
10h00-10h45	Exposé invité : S. Brunton, Prof. University of Washington Data-driven discovery and control of complex systems: Uncovering interpretable models of nonlinear flow physics
10h45-11h05	Pause & Poster (cf. liste des posters)
11h05-11h30	M. Bocquet et al. Data-driven inference of the ordinary differential equation representation of a chaotic dynamical model using data assimilation
11h30-11h55	I. Ayed et al. Learning Partially Observed PDE Dynamics with Neural Networks
11h55-12h20	A. Stegner et al. Deep learning applied to oceanic eddy detection.
12h20-12h45	R. Lguensat et al. Discovering hidden equations in ocean simulation models using Quasi-Geostrophic theory guided machine learning.
12h45-14h00	Repas
14h00-15h40	Session B (Animateur : L. Terray)
14h00-14h45	Exposé invité : V. Balaji, Princeton University Machine learning and the post-Dennard era of climate simulation
14h45-15h10	F. Falasca et al. Dimensionality reduction and network inference for climate data
15h10-15h35	A. Ayet et al. Using analogs to predict cloud cover from satellite images.
15h35-16h00	M. Amiramjadi et al. Using machine learning to estimate subgrid-scale processes from a coarse-grained description of the flow
16h00-16h45	Pause & Poster (cf. liste des posters)
16h45-17h45	Table-Ronde (Animateur : J. Le Sommer) Participants : J. Lambin (CNES), J.-P. Proux (GENCI) A. Bentamy (Ifremer), P. Braconnot (INSU), P. Dandin (Météo France), G. Charpiat (INRIA)
17h45-18h00	Mot de clôture de la journée

Liste des Posters :

- M. Amiramjadi et al. Using machine learning to estimate subgrid-scale processes from a coarse-grained description of the flow
- Charantonis et al. IA pour la reconstitution de profils verticaux dans les océans
- G. Dufour. First attempt to recognize patterns and classify tropospheric ozone retrieved from the IASI satellite instrument.
- K. Grassi et al. Détection d'événements extrêmes à partir de bases de données multivariées. Application à l'observation marine.
- G. Hochard and N. Bousquet (Quanmetry) Instabilités climatiques – Evolutions des tempêtes et classification par machine learning.
- P. Lepetit, C. Mallet, L. Barthès (LATMOS) Estimation des conditions météorologiques d'après photo: état de l'art et perspectives de l'IA.
- N. Longépé et al. Apprentissage profond pour la classification d'image radar (SAR) sur l'océan: apport pour les sciences de l'atmosphère, de l'océan et du climat
- J. Mignot et al. Détermination des modèles d'ensemble climatiques
- J. Monnier et al. Inland Ice-Sheet Bed Topography Estimations from Satellite Surface Measurements and a Dedicated Physical Based Model.
- S. Ouala et al. Learning dynamical model from partial and noisy observations
- R. Paugam et al. Image Segmentation: Fire Front extraction for Rate of Spread Estimation
- D. Ruiz Pino et al. Identification, reconstruction en 3D et quantification du phytoplancton en relation au changement climatique: du microscope électronique à l'accélérateur de particules.
- T. Rieutord et al. Classification de couche limite atmosphérique par apprentissage non-supervisé sur des données de campagnes de mesure.

Liste des Résumés

Exposés invités :

- S. Brunton. Prof. University of Washington. Data-driven discovery and control of complex systems: Uncovering interpretable models of nonlinear flow physics.

Accurate and efficient reduced-order models are essential to understand, predict, estimate, and control complex, multiscale, and nonlinear dynamical systems. These models should ideally be generalizable, interpretable, and based on limited training data. This work develops a general framework to discover the governing equations underlying a dynamical system simply from data measurements, leveraging advances in sparsity-promoting techniques and machine learning. The resulting models are parsimonious, balancing model complexity with descriptive ability while avoiding overfitting. This perspective, combining dynamical systems with machine learning and sparse sensing, is explored with the overarching goal of real-time modeling and closed-loop feedback control of unsteady fluid systems. We will discuss how it is possible to enforce known constraints, such as energy conserving quadratic nonlinearities in incompressible fluids, to “bake in” known physics. Next, we will demonstrate that higher-order nonlinearities can approximate the effect of truncated modes, resulting in more accurate models of lower order than Galerkin projection. This approach is demonstrated on several relevant systems in fluid dynamics with low-dimensional dynamics.

- V. Balaji. Prof. Princeton University. Machine learning and the post-Dennard era of climate simulation.

Conventional computational hardware has reached some physical limits: the phenomenon known as 'Dennard scaling' gave rise to Moore's Law, and many cycles of exponential growth in computing capacity. The consequence is that we now anticipate a computing future of increased concurrency and slower arithmetic. Earth system models, which are weak-scaling and memory-bandwidth-bound, face a particular challenge given their complexity in physical-chemical-biological space, to which mapping single algorithms or approaches is not possible. A particular aspect of such 'multi-scale multi-physics' models that is under-appreciated is that they are built using a combination of local process-level and global system-level observational constraints, for which the calibration process itself remains a substantial computational challenge. In this talk, we examine approaches to Earth system modeling in the post-Dennard era, inspired by the industry trend toward machine learning (ML). ML presents a number of promising pathways, but there remain challenges specific to introducing ML into multi-phase multi-physics modeling. These include, among others: the non-stationary and chaotic nature of climate time series; the presence of climate subsystems where the underlying physical laws are not completely known; and the imperfect calibration process alluded to above. The talk will present ideas and challenges and the future of Earth system models as we prepare for a post-Dennard future.

Communications orales et poster :

- *M. Amiramjadi (Univ. Teheran/LMD), R. Plougonven, A. Hertzog (LMD), Ali R. Mohebalhojeh, Mohammad Mirzaei (Univ. Teheran)*

Using machine learning to estimate subgrid-scale processes from a coarse-grained description of the flow

This survey explore the use of machine learning to estimate a sub-grid process from knowledge of a coarse-grained description of the flow in the troposphere. The process studied is atmospheric gravity waves which need to be parameterized in climate models. In these parameterizations, an essential yet poorly constrained component is the launch spectrum of non-orographic waves in the troposphere or lowermost stratosphere. We make use of the fact that current Numerical Weather Prediction models, with a spatial resolution of only a few tens of kilometers, describe well the variations of the gravity wave field, although it remains only partly unresolved (Jewtoukoff et al 2015). Hence the ERA5 reanalyses at full resolution provide us knowledge of the subgrid- scale process to estimate (the target): gravity waves at 100 hPa, quantified by the variance of vertical velocity or the associated momentum fluxes. We employ a 5-month data-set (May to September 2018), for a mid-latitude region of the Indian Ocean far from land masses and orographic obstacles. Explanatory variables are Low-resolution ERA5 data ($2.5^\circ \times 2.5^\circ$) through the troposphere (from 1000 to 100 hPa) including wind speed, relative vorticity, temperature gradients, and the Fraction of Cloud cover.

As a first exploration of the capacity to reconstruct information on gravity waves from a low-resolution description of the tropospheric flow, we have used Random Forests as a non-parametric statistical model which readily provides information on the relative importance of the different explanatory variables. The statistical model is trained and tested using 10-fold cross-validation. The gravity wave diagnostics are reconstructed from knowledge of the tropospheric flow with promising agreement: correlation coefficients between the target and the reconstructed timeseries are 0.7 or better. We also performed another examination to train the model over the Indian Ocean and test it on another area over Indian and Southern Pacific Ocean. The statistical evaluation shows reasonable results again. This exploration provides insights for the development of gravity wave parameterizations: first, it indicates a lower bound on how much of the gravity wave signal can be reconstructed, given the low-resolution information on the tropospheric flow. Second, it ranks the tropospheric variables by order of relevance, revealing notably the importance of upper-level winds.

· ***I. Ayed, E. De Bézenac, A. Pajot, P. Gallinari (LIP6, Sorbonne Université).
Learning Partially Observed PDE Dynamics with Neural Networks***

Le résumé : Spatio-Temporal processes bear a central importance in many applied scientific fields. Generally, differential equations are used to describe these processes. In this work, we address the problem of learning spatio-temporal dynamics with neural networks when only partial information on the system's state is available. Taking inspiration from the dynamical system approach, we outline a general framework in which complex dynamics generated by families of differential equations can be learned in a principled way. Two models are derived from this framework. We show how the underlying equations fit into our formalism and evaluate our method by comparing with standard baselines.

· ***A. Ayet, IFREMER/LOPS-ENS/LMD, Pierre Tandeo, IMT Atlantique/LabSTICC***

Using analogs to predict cloud cover from satellite images

Accurate forecasting of cloud cover is essential for many applications among which the integration of the solar resource in an electrical grid. Here, we develop an analog method suitable for its up-to-6h probabilistic forecasting. The method does not require calibration to adapt to regional differences in cloud dynamics, and uses only one type of satellite data, covering Europe and Africa. It is thus suited for applications that require a cloud cover forecast for solar energy sources at different locations without ground measurements nor numerical weather predictions.

The method predicts cloud dynamics based on the identification, in a historical database, of cloud patterns similar to the observed one, and assuming that their evolution will be similar. We discuss how the skills of the method in predicting cloud patterns over a given region can be linked to clustering properties of the geophysical data, linked with the regional cloud dynamics. In the present case, we use as a historical database of 5 years of solar irradiance measurements from a geostationary satellite, with a temporal resolution of 1 h and a spatial resolution of 0.05° . Results over one year in different locations in Europe demonstrate the skill of the method in predicting cloud cover.

Reference:

Ayet, A., & Tandeo, P. (2018). Nowcasting solar irradiance using an analog method and geostationary satellite images. *Solar Energy*, 164, 301-315.

M. Bocquet CERE, J. Brajard SU/LOCEAN-NERSC, A. Carrassi NERSC and L. BertinoNERSC

Data-driven inference of the ordinary differential equation representation of a chaotic dynamical model using data assimilation

Recent progress in machine learning has shown how to forecast and, to some extent, learn the dynamics of a model from its output, resorting in particular to neural networks and deep learning techniques. We will show how the same goal can be directly achieved using data assimilation techniques without leveraging on machine learning software libraries, with a view to high-dimensional models. The dynamics of a model are learned from its observation and an ordinary differential equation (ODE) representation of this model is inferred using a recursive nonlinear regression. Because the method is embedded in a Bayesian data assimilation framework, it can learn from partial and noisy observations of a state trajectory of the physical model. Moreover, a space-wise local representation of the ODE system is introduced and is key to deal with high-dimensional models.

The method is illustrated on several chaotic discrete and continuous models of various dimensions, with or without noisy observations, with the goal to identify or improve the model dynamics, build a surrogate or reduced model, or produce forecasts from mere observations of the physical model.

It has recently been suggested that neural network architectures could be interpreted as dynamical systems. Reciprocally, we show that our ODE representations are reminiscent of

deep learning architectures. Furthermore, numerical analysis considerations on stability shed light on the assets and limitations of the method.

A. Charantonis (ENSIIE, LAMME, LOCEAN), S. Thiria (UVSQ, LOCEAN)
IA pour la reconstitution de profils verticaux dans les océans

La fréquence spatio-temporelle, le nombre de bandes spectrales et la résolution avec lesquelles nous surveillons la surface des océans n'a cessé d'accroître durant l'ère satellitaire. Bien que différents effets (nuages, vagues, orbites satellitaires, cycle diurne...) rendent l'observation constante impossible, la quantité et qualité d'observations récoltées font d'elle un bon exemple du terme "Big Data".

Ceci est en grand contraste avec la fréquence spatio-temporelle d'acquisition de profils verticaux de l'océans, qui sont, eux, limités à des mesures de mouillages permanents, à des campagnes océanographique ponctuelles et à certains drones gliders ou bouées. Bien que la quantité à crête, elle reste largement inférieure à la quantité de données de surface. Elles sont tout de même de grande importance: elle conditionnent un grand nombre de phénomènes cruciaux pour bien conditionner la modélisation climatique.

Relier ces deux types de données est un challenge auquel les technique IA telles que les réseaux de neurones sont bien adaptées. Nous montrerons approches IA développées pour la reconstruction de profils verticaux, et leur applications pour la reconstruction de profils tels que des profils de température, de salinité, de Chlorophyll-a ou de vitesse.

G. Dufour, LISA

First attempt to recognize patterns and classify tropospheric ozone retrieved from the IASI satellite instrument

The IASI instrument, a infrared spectrometer, is aboard the suite of the European meteorological satellites, Metop, operated by EUMETSAT. Three Metop satellite are currently flying: Metop-A since 2006, Metop-B since 2012 and Metop-C since 2018. The Metop satellites are low-earth-orbit satellites with a crossing time at the equator at two fixed local solar times 09:30 LST (descending mode) and 09:30 (ascending mode). The IASI instruments operate in the thermal infrared with a spectral resolution of 0.5 cm⁻¹. The field of view of the instruments is composed of a 2x2 matrix of pixels with a diameter at nadir of 12 km each. IASI scans the atmosphere with a swath width of 2200 km, which allows global coverage twice daily. Ozone profiles and partial columns can be retrieved from the IASI radiances for each pixel, providing global distribution of ozone within the atmosphere. One of our objectives is to understand the drivers of the variability and trends of tropospheric ozone over the anthropised regions of the northern hemisphere. Faced with the large amount of data to analyze, we have started to explore the possibility offered by the unsupervised machine learning algorithms to identify patterns of tropospheric ozone and classified them according to the meteorological and climatic variability, the large scale dynamical processes, and the precursors' emissions. Preliminary results of pattern recognition.

F. Falasca, A. Bracco, A. Nenes, I. Fountalis, Georgia Institute of Technology

Dimensionality reduction and network inference for climate data

We present a new framework for analysing and benchmarking climate model outputs. The framework is built upon δ -MAPS, a recently developed complex network analysis method. Advantages of δ -MAPS over other more traditional methodologies include the possibility of highlighting quantifiable differences across datasets, capturing the magnitude of interactions including lagged relationships, quantifying the modeled internal variability and changes in domain properties and their connections over space and time. We use this methodology to investigate the evolution of connectivity patterns in sea surface temperatures in two reanalysis datasets over the period 1980-2015. We propose a set of four metrics to assess and compare the modeled domains shapes, strengths and connectivity patterns. Ensemble members are then mapped in a reduced metric space allowing for quantification of internal variability and average model error. It is found that network properties are robust to the detrending strategy whenever individual member or ensemble trends are removed.

Grassi K. WeatherForce/Ifremer/LISIC, Poisson-Caillault E.Ifremer, Lefebvre A. LISIC Détection d'événements extrêmes à partir de bases de données multivariées. Application à l'observation marine.

L'observation du milieu marin a considérablement évolué ces dernières années avec un recours à des systèmes automatisés de mesures à haute fréquence. La compréhension générale de la dynamique du phytoplancton (caractéristiques saisonnières et pluriannuelles en réponse à différents types de forçage) commence à être mieux comprise. Néanmoins le rôle et l'impact des événements extrêmes sur cette dynamique ne sont que partiellement abordés faute de méthode numérique adaptée, optimisée. Afin d'extraire la quantité maximale d'information des données, une architecture de classification multi-couche par approche spectrale a été développée. La méthode proposée testée sur un jeu de données simulées et étiquetées manuellement s'avère globalement plus puissante que les méthodes habituellement utilisées. Après validation, la méthode a été appliquée sur un jeu de données issues du système de mesures MAREL Carnot. Alors que les premiers niveaux de classification permettent de retrouver les grandes phases du développement du phytoplancton et permettent même d'identifier automatiquement des défaillances de capteurs, les niveaux supérieurs correspondent davantage à des événements extrêmes (courte durée, amplitude exceptionnelle). Ce genre d'approche laisse présager d'une capacité renforcée à mieux comprendre les relations pressions/impacts donc à améliorer la prédiction de certains événements comme les efflorescences nuisibles du phytoplancton.

Mots clefs : série temporelle, haute résolution, extrême, Classification Spectrale Multi-niveau (M-SC), phytoplancton.

G. Hochard, N. Bousquet. Quantmetry Instabilités climatiques – Evolutions des tempêtes et classification par machine learning

Quantmetry se positionne depuis plusieurs années comme pionnier sur des sujets d'IA nécessitant une forte compétence R&D. Depuis 2017, nous collaborons avec un institut de recherche au Canada et un laboratoire aux Etats-Unis (Berkeley Lab) regroupant les meilleurs

experts en matière de changement climatique, afin d'investiguer sur le traitement de données spatio-temporelles et d'apporter une contribution et un soutien à ces équipes de recherche.

Le contexte de ces travaux est la thématique du sujet d'attribution. L'attribution en climatologie est une méthode scientifique permettant de séparer un monde chaud (incluant les gaz à effet de serre et l'impact anthropogénique) d'un monde froid (concrètement, un monde dans lequel la révolution industrielle et les gaz à effet de serre sont absents). Face à l'observation de phénomènes météorologiques extrêmes, la méthodologie d'attribution permet d'émettre des hypothèses sur les causes du phénomène observé : est-il dû à un monde chaud ou à un monde froid?

A partir de données d'observation, nous démontrons la faisabilité d'un classifieur permettant de séparer efficacement des tempêtes et cyclones tropicaux appartenant à deux périodes de temps distinctes (1980-2000 et 2000-2017). Nous montrons ainsi des différences dans les descripteurs de ces tempêtes entre ces deux périodes. Les descripteurs sont créés à partir de variables issues de l'agrégation de séries temporelles.

P. Lepetit, C. Mallet, L. Barthès, LATMOS

Estimation des conditions météorologiques d'après photo: état de l'art et perspectives de l'IA.

L'homme sait discriminer des situations météorologiques sur la base de l'image. La vision artificielle permet-elle de caractériser le temps avec la même acuité? Permet-elle de faire mieux, en estimant, par exemple, des paramètres quantitatifs comme la visibilité ou la température?

L'estimation des conditions météorologiques (ECM), une branche de la compréhension de scène, aborde ces deux questions depuis une quinzaine d'années. A la fin des années 2000, les approches par apprentissage, très flexibles, sont venues compléter la modélisation physique et, suivant une tendance globale, des datasets de plus en plus larges ont été regroupés. A partir de 2016, le deep learning s'est imposé comme un outil incontournable dans le domaine. Dans cet exposé, nous décrivons cette évolution et nous présentons un état de l'art à jour. Les potentielles applications en matière d'observation du temps présent et du climat seront présentées dans un second temps.

R. Lguensat, J. Le Sommer, S. Metref, E. Cosme (UGA, CNRS/IGE) R. Fablet (Lab-STICC)

Discovering hidden equations in ocean simulation models using Quasi-Geostrophic theory guided machine learning

Inferring hidden equations governing dynamical systems from data has always been one of the challenging problems in the interplay between physics and data science. It was just a matter of time before the recent advancements in machine learning and in computational capacities come in hand and spark off a series of works dedicated to address this problem.

In this work we present a Quasi-Geostrophic numerical model coded using differentiable operators thus permitting the use of automatic differentiation libraries (e.g. Tensorflow). This makes the model flexible and suited for parameter optimization, especially using neural networks. We illustrate the relevance of the proposed architecture through an example of a regression problem where we show how can we obtain the parameters of the potential

vorticity equation using only consecutive scenes of Sea Surface Height. This can be of interest for finding the closest QG-like approximation to a given ocean simulation model or help exploring the effect of adding new operators in the potential vorticity equation. The code we provide is suitable for GPU implementation and therefore can allow for faster execution and profit from the quick advancements in GPU development.

We expect that the directions of research we suggest will help in bringing more interest in applied machine learning to ocean numerical modeling.

Nicolas LONGEPE, Chen WANG, Romain HUSSON, Guillaume ERHARD, Alexis MOUCHE, Pierre TANDEO, Justin STOPA

Apprentissage profond pour la classification d'image radar (SAR) sur l'océan: apport pour les sciences de l'atmosphère, de l'océan et du climat.

Dans le cadre du programme européen Copernicus, des milliers d'images radar haute résolution (SAR) sont produites chaque jour représentant une moyenne quotidienne de 3,45 To de Sentinel-1 publiées. Une quantité importante couvre la surface de l'océan, utilisée pour un large éventail d'applications (pollution, sécurité maritime, glace de mer...). Divers phénomènes océaniques sont observés quotidiennement par Sentinel-1. Pour n'en nommer que quelques-uns, les fronts atmosphériques, les fronts océaniques, les cellules de pluie, les micro-cellules convectives, les ondes internes, les ondes de gravité, les nappes biologiques, les remontées d'eau ou les traînées de vent sont des phénomènes qui peuvent intéresser de nombreux utilisateurs. Pourtant, ces phénomènes sont négligés (au mieux, masqués) dans les produits de niveau 2 générées par le segment sol de l'ESA (Agence Spatiale Européenne).

Ces travaux visent à fournir une méthodologie pour classifier et segmenter les images SAR acquises sur l'océan. Une base de données comprenant 36000 imagerie (20 x 20 km) acquises par Sentinel-1 en mode "Wave Mode" est constituée, chaque imagerie étant labellisée manuellement parmi les 10 classes retenues (front océanique, front atmosphérique, houle, zone de vent faible, film biologique, pluie convective, micro-cellule de convection, traînée de vent, glace de mer et iceberg).

Cette base de données sert d'entrée à des méthodes d'apprentissage profond type réseaux de neurones convolutifs (CNN). Différents modèles (Inception, Xception, ResNet50, VGG ...) sont alors testés. Un modèle Inception v3 est finalement choisi, puis appliqué à l'ensemble des données acquises par S1 en mode "Wave Mode" depuis 2014, i.e plus de 4 millions d'images couvrant les océans Pacifique, Atlantique, Indien et Austral. L'analyse de ces classifications permet de produire des cartes mensuels d'occurrence de ces phénomènes à l'échelle globale. Les transitions spatiales dans la basse atmosphère entre les traînées de vent et des conditions plus convectives de la couche limite sont aussi étudiées.

Enfin, le potentiel et les limites de ces travaux pour les sciences de l'atmosphère, de l'océan et du climat sont élaborés.

J. Mignot, C. Mejia, C. Sorrow, A. Sylla, S. Thiria, M. Crépon (LOCEAN)

Détermination des modèles d'ensemble climatiques

Les modèles climatiques de type CMIP sont des gros modèles numériques qui présentent tous des défauts dus à la représentation différente des processus physiques et de la discrétisation

numérique. Un moyen de pallier ceci est de réaliser des modèles d'ensemble, qui sont des combinaisons des ces différents modèles. Nous proposons dans cette étude une méthodologie sélectionner ces différents modèles dans la réalisation du modèle d'ensemble. Pour cela on détermine à l'aide d'une classification neuronale (Cartes topologiques) couplée à une analyse factorielle des correspondances, les modèles qui représentent au mieux selon une certaine métrique les paramètres climatiques observés (Réanalyse) dans une zone géographique donnée. On peut ainsi déterminer un modèle climatique d'ensemble pour une zone géographique donnée. On présente les résultats d'une étude faite pour estimer un modèle d'ensemble optimal dans une zone couvrant l'upwelling Sénégal-Mauritanien à partir de 45 modèles climatiques CMIP5 en prenant comme estimateur la variabilité saisonnière de la température de surface et comme observations les données HadISST. On est capable de donner une hiérarchie des différents modèles CMIP5 selon ce critère et pour cette zone.

***J. Monnier and J. Zhu. Mathematics Institute of Toulouse (IMT) & INSA, France.
Inland Ice-Sheet Bed Topography Estimations from Satellite Surface Measurements and a Dedicated Physical Based Model***

Bed topographies beneath inland ice-sheets, Antarctica in particular, are estimated by Kriging between (very) sparse airborne measurements or from gravity field inversions. The resulting uncertainties may be large. The present study proposes a new physical-based inversion method to estimate these inland bed topographies i.e. in relatively slow flow areas. The inverse problem is particularly challenging since the measured "surface signatures" integrate a mix of the bottom features (bed elevation and friction-slip amount) and the internal deformation (including varying temperature vertical profiles).

The first key step of the present inversions is the derivation of a "Reduced Uncertainty" formulation of the Shallow Ice Approximation (SIA) model taking into account temperature variations. The resulting RU-SIA model contains the satellite surface measurements (altimetry and InSAR) as coefficients and presents a single uncertain (multi-physics) parameter γ . The next key steps are advanced Variational Data Assimilation (VDA) formulations combined with a stochastic extension of γ based on the trend observed in the in-situ measurements (along the flight tracks).

Numerical estimations obtained for large inland Antarctica areas are analysed into details. Employed datasets are multi-missions ones, containing altimetry, InSAR, airborne radars measurements.

References

- [1] J. Monnier, J. Zhu. "A reduced uncertainty shallow ice flow model and bed topography estimations from surface data ". Submitted.
- [2] J. Monnier, J. Zhu. "Physical-based bed topography inversions inland Antarctica". In prep.
- [3] J. Monnier, P.-E. des Bosc. "Inference of the Bottom Properties in Shallow Ice Approximation Models". *Inverse Problems*, vol. 33 (9) 2017.

S. Ouala, V.-D. Nguyen, L. Drumetz, B. Chapron, A. Pascual, F. Collard, L. Gaultier, R. Fablet. Lab-STICC/Ifremer/IMEDEA/ODL

Learning dynamical model from partial and noisy observations

The constantly increasing wealth of simulation and observation data on geophysical dynamics make more and more appealing data-driven strategies as new means to address key issues in ocean and atmosphere science, including for instance forecasting and assimilation issues [1]. In this respect, recent studies have investigated data-driven strategies to identify governing equations from data using different machine learning frameworks, especially sparse regression models [2] and neural networks [3]. Among others, such data-driven representations of dynamical operators have shown to be of key interest for data assimilation issues, especially the spatio-temporal interpolation of geophysical fields from ocean remote sensing data [4]. The availability of representative training datasets is a strong requirement for the development of such approaches. When considering observation datasets (e.g., satellite-derived data or in-situ observations), the question whether one may learn such data-driven representations from noisy and partial observation data naturally arises. For instance, regarding sea surface dynamics, beyond observation noise patterns, satellite sensors also involve irregular space-time sampling patterns due to their intrinsic characteristics or their sensitivity to the atmospheric conditions. In this work, we investigate these issues. We show that the effectiveness of previously proposed learning-based methods [2, 3, 5] is strongly affected when applied to noisy and partial observation datasets. Within a neural-network-based framework, we address the data-driven identification of governing equations as the joint learning of a dynamical models and the assimilation of the hidden states from noisy and partial observations. We restate state-of-the-art assimilation schemes (e.g., Ensemble Kalman Filters, Variational schemes) as neural network architectures, such that the identification of the dynamical operator comes to the minimization of a data assimilation cost, rather than a simple forecasting error. From numerical experiments on Lorenz-63 dynamics and sea surface dynamics, we demonstrate the relevance of the proposed approaches for the data-driven identification of dynamical operators, both in terms of forecasting performance and long-term topological patterns, within respect to the state-of-the-art. These experiments also illustrate that different assimilation-inspired neural-network-based architectures may be of interest depending on the noise levels.

Keywords. Data-Driven Models, Identification of governing equations, Forecasting, Data assimilation, Neural Networks, Missing data.

References

- [1] Pierre Tandeo, Pierre Ailliot, Bertrand Chapron, Redouane Lguensat, and Ronan Fablet, “The analog data assimilation: application to 20 years of altimetric data,” in International Workshop on Climate Informatics, Boulder, United States, sep 2015, pp. 1 – 2.
- [2] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” Proceedings of the National Academy of Sciences, vol. 113, no. 15, pp. 3932–3937, Apr. 2016.
- [3] Ronan Fablet, Said Ouala, and Cedric Herzet, “Bilinear residual Neural Network for the identification and forecasting of dynamical systems,” SciRate, dec 2017.
- [4] Said Ouala, Ronan Fablet, Cédric Herzet, Bertrand Chapron, Ananda Pascual, Fabrice Collard, and Lucile Gaultier, “Neural network based kalman filters for the spatio-temporal

interpolation of satellite-derived sea surface temperature,” Remote Sensing, vol. 10, no. 12, pp. 1864, Nov 2018.

[5] Redouane Lguensat, Pierre Tandeo, Pierre Ailliot, Manuel Pulido, and Ronan Fablet, “The Analog Data Assimilation,” Monthly Weather Review, aug 2017.

· ***R. Paugam, M. Rochoux, N. Cazard, C. Lapeyre (CERFACS), W. Mell (US FS), M. Wooster (KCL)***

Image Segmentation: Fire Front extraction for Rate of Spread Estimation

Keywords: fire behavior; airborne observation; thermal Camera; Prescribed Burn; Deep Learning

To mitigate wildfire effects, coupled fire-atmospheres systems are developed to resolve simultaneously the plume updraft/smoke dispersion, the propagation of the fire front, and their mutual interactions. They are design to simulate landscape-scale (>100m) propagating fire and rely on parameterized fire model. Fires are parameterized as front lines with associated Rate Of Spread (ROS) and sensible heat flux both predicted according to empirical model based on local orography and atmospheric variables (ie wind speed, humidity).

To help the development and validation of coupled fire-atmosphere system, the fire modelling community need validation datasets of fire scenarios where fire induced winds influence fire front behavior (eg monitoring of front acceleration induced by merging). Because of the high variability of fire behavior, potential complex front geometry structures and fast spread, high temporal and spatial resolution observation are required to build such dataset.

Infra Red (IR) thermal cameras operated from helicopter can be used to monitor landscape scale propagating fire at high resolution. To extract valuable information from those observations (eg front position, ROS), image processing task need to be applied on the raw radiance field collected by the camera and can be divided into 3 steps: (a) orthorectification which consists in warping the raw images on a fix coordinate system grid to correct for camera lens distortion and perspective effects induced by camera tilt and terrain, (b) segmentation which reside in delineating the fire front location out of the orthorectified radiance field, and finally (c) the computation of fire behavior metrics (e.g. computation of the ROS from the consecutive location of the front location).

In this presentation we will focus on step (b) and show how deep learning method can be used to draw fire front perimeters on more than 1000 IR images processed independently. In particular, we will show the iterative learning methodology we setup that relies on user interaction and a UNET neural network fed initially with a set of 12 manually annotated images.

· ***Thomas Rieutord, Alexandre Paci, Alain Dabas, CNRM***

Classification de couche limite atmosphérique par apprentissage non-supervisé sur des données de campagnes de mesure

La couche limite atmosphérique est la plus basse partie de l’atmosphère : celle influencée par le sol à l’échelle de l’ordre de l’heure. Mais comme le sol est sujet à de grande variations au cours du cycle diurne, la couche limite atmosphérique présente aussi des caractéristiques très différentes. La nuit, le sol froid induit une stratification stable, alors que le jour, le réchauffement par le soleil induit une stratification convective. En sus, d’autres phénomènes

sont à l'œuvre, par exemple les écoulements locaux en terrain complexe ou les nuages au sommet de la couche limite, qui rendent les choses encore plus complexes. Par conséquent, il est fréquent de distinguer plusieurs types de couche limite (parmi lesquelles « stable » et « convective ») afin de mieux décrire ce qui s'y passe.

Ce travail vise à utiliser des méthodes d'apprentissage non-supervisé pour faire une telle classification. A la lumière des récentes avancées dans d'autres domaines, de multiples améliorations sont attendues de ces méthodes : des algorithmes retournant la même classification qu'un expert humain, des cas complexes mieux gérés qu'avec un arbre de décision fait à la main et une meilleure flexibilité quand les instruments et la position changent. Cette présentation s'attachera seulement à la première de ces attentes. Elle se concentrera sur la classification hiérarchique ascendante : avec une métrique et une stratégie de connexion, cet algorithme retourne des classes imbriquées, hiérarchisées. Un cas d'étude détaillé explorera la sensibilité de l'algorithme aux différentes configurations possibles (plusieurs métriques et plusieurs stratégies de connexion ont été testées) ainsi qu'aux changements dans les données (résolution, bruit...). La performance des différentes options seront comparées à la fois par une évaluation humaine et des scores numériques (indice de Calinski-Harabaz et score de silhouette). Les données proviennent de deux instruments enclins à être déployés largement : un radiomètre micro-ondes (mesurant des profils de température) et un télémètre à nuage (mesurant des profils de rétrodiffusion par les aérosols). Ils opéraient pendant la campagne de mesure Passy-2015, consacrée à la dynamique atmosphérique et la qualité de l'air dans une vallée alpine en hiver. Le cas d'étude montre que, malgré une phase de réglage qui ne doit pas être négligée, cette méthode donne une information pertinente à plusieurs échelles qui mériterait d'être testée sur des jeux de données plus large. Elle permet de définir les types de couche limite « d'après les données » plutôt que « d'après la théorie ».

D. Ruiz Pino (LOCEAN), S. Gardin (CR2P), L. Lemelle (LGL-TPE)

Identification, reconstruction en 3D et quantification du phytoplancton en relation au changement climatique: du microscope électronique à l'accélérateur de particules.

Un acteur majeur et peut être une des solutions au changement climatique actuel nous est donné par la captation du CO₂ par le phytoplancton (algues unicellulaires) qui constitue la forêt cachée des océans. Cet organisme lorsqu'il est fossilisé est le meilleur outil pour la reconstruction des paléoclimats et des paléo-environnements.

Identifier ces organismes micrométriques, impose aujourd'hui un travail d'obtention de millions d'images très précises, obtenues en Microscopie Electronique à balayage (MEB) qui permettent de reconnaître et de documenter leur biodiversité et leur abondance. Néanmoins, cette identification dans la majorité de cas, nécessite aujourd'hui encore, l'inspection de chacune des images par l'œil des experts. Le temps nécessaire pour scruter les images, les erreurs, les incertitudes, et les problèmes de non-répétabilité pourrait être réduit grâce à un travail joint entre biologistes de l'océan capable de reconnaître chacune des espèces et de spécialistes de méthodes de reconnaissance de forme et de réseaux de neurones. Si des démarches pour identifier certaines de ces algues ont vu le jour dans la littérature, de nombreux problèmes restent à résoudre et au jour d'aujourd'hui il n'existe pas de méthode capable de déterminer automatiquement la très grande biodiversité du phytoplancton marin.

Quantifier à échelle de la cellule, le contenu d'éléments clé comme le carbone, mais aussi en différents composants, tel que le CO₂, le soufre et le fer émis par les volcans et/ou par l'activité anthropique qui peuvent affecter la croissance cellulaire et/ou la fabrication de test (partie dure de ces organismes, composé de Carbonate de calcium ou de Silice), reste un de plus grands challenges actuels. Cette quantification implique la connaissance en 3D de structures complexes, parfois fractales et un voyage à l'intérieur d'organismes parmi les plus petits de notre planète. Aujourd'hui, de nombreuses distributions élémentaires de ces organismes microscopiques peuvent être imagées sur les lignes synchrotron de nano-imagerie de la fluorescence X (accélérateur de particules). Le traitement d'images de structures fines (MEB) et d'images compositionnelles en vue de quantifier des masses élémentaires totales gagnerait à être résolu avec l'utilisation de l'intelligence artificielle. Il doit en effet combiner, en plus de l'extrême biodiversité, des problèmes d'ordre mathématique, géométrique, informatique appliqués à un grand nombre d'images. Ces reconnaissances et reconstructions de l'infiniment petit dans l'océan apporteraient également des avancées importantes dans la connaissance de la biodiversité mais aussi de la productivité de 70% de la surface de notre planète, avec des implications certaines pour les ressources marines vivantes.

A. Stegner, B. Le Vu (LMD, CNRS, Ecole Polytechnique); O. Schwander, P. Gallinari (LIP6, SU)

Deep learning applied to oceanic eddy detection

The increases of the spatial resolution of numerical models and remote sensing observations both revealed the prevalence of eddies throughout the oceans. These structures are able to trap and transport heat, mass, momentum, and biogeochemical properties from their regions of formation to remote areas. Eddies have therefore a significant impact on the surface circulation at both local and regional scales. In order to investigate a large number of coherent structures for long periods (several years) the development of automatic eddy detection and tracking algorithms is now an "hot spot" of research in oceanography.

Many objective criterions could be used to detect and identify eddies (Mkhini et al. 2014, LeVu et al. 2017). However, the remote sensing observations (sea surface height, temperature or color...) are often corrupted by clouds (visible imagery) or by the coarse distribution of altimetric tracks. The combination of distinct remote sensing data-set (SSH, SST, CHL...) could restore the lack of information that may occurs on a single data set. To reproduce the human capabilities of pattern recognition artificial neural network is now an efficient and powerful tool.

The first step for such deep learning approach is to provide a very large number of "images of eddy signature" to train the artificial neural network. We will present a reliability index, that we developed recently, to select only the detected eddies with a high-level of accuracy on the AVISO/DUACS altimetric data-sets. Therefore, we will guarantee that the eddy location and size are correct and we will be able to select the correct eddy area on the visible images (SST, CHL) without any a priori on the eddy signature.

In a second step a classification and identification procedure, using deep learning methods, to detect oceanic eddy from velocities and temperatures at the sea surface was tested on one year of numerical model output (NEMO-MED12) in the Mediterranean Sea.

