SEACS: Stochastic modEl-dAta-Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics

Final Report

Introduction

The SEACS project aimed at exploring novel statistical and stochastic methods to address the emulation, reconstruction and forecast of fine-scale upper ocean dynamics. From the recent advances in the theoretical understanding, modeling and simulation of upper ocean dynamics along with the mass of data routinely available to observe the ocean evolution, our key objective was to investigate new tools and methods for the calibration and implementation of novel sound and efficient oceanic dynamical models. In this respect, the emphasis was put on stochastic frameworks to encompass multi-scale/multi-source approaches and benefit from the available observation and simulation massive data.

The addressed scientific questions studied in this project constitute basic research issues at the frontiers of several disciplines. It crosses in particular advanced data analysis approaches, physical oceanography and stochastic representations. To that end the SEACS consortium gathered a set of research groups associated with these different scientific domains.

The SEACS Consortium

The consortium was composed of research groups covering different aspects of oceanic data analysis, physical modeling and processing through different methodological tools. Although they belong to different scientific communities, these groups share common objectives and have a longstanding history of multiple collaborations between each of them. More precisely, The project gathered 6 partners with expertise in statistics, data science, remote sensing and physical oceanography : Inria team Fluminance (PI: E. Memin), Inria/IRMAR team Aspi (PI: V. Mombet), Lab-STICC team TOMS (PI: R. Fablet), IFREMER LOPS laboratory (PIS: B. Chapron and G. Roullet) and LMBA team (PI: P. Ailliot).

Scientific program

The central question addressed by the project concerns our capacity to couple data and models of oceanic flows. This question is at the core of our capacity to forecast and analyze the ocean state. This has obviously huge repercussions in several economical domains such as fishery, transport, energy but also for our future to anticipate the consequences of potential oceanic changing regimes due to global warming. This coupling issue, which is a quite common question in several scientific domains, reaches here a high degree in difficulty due to the huge dimension of the dynamical system at stake, to its nonlinear chaotic nature, to a lack of accuracy at small-scale due to the complexity of the physical phenomenon involved and last, to our inability to quantify properly the associated uncertainty. The nature of the available data does not neither help to greatly facilitate this coupling. The data are of very big dimension as well, sparse either in space or time, strongly incomplete and more and more defined at very high-resolution. This last point, which is certainly a necessity called by numerous applications, may appear at first glance as a great opportunity for our coupling issue. Yet it leads to a strong increase of the difficulty: how to enable a possible discussion when data and models perceive two (very) different realities of the same phenomena?

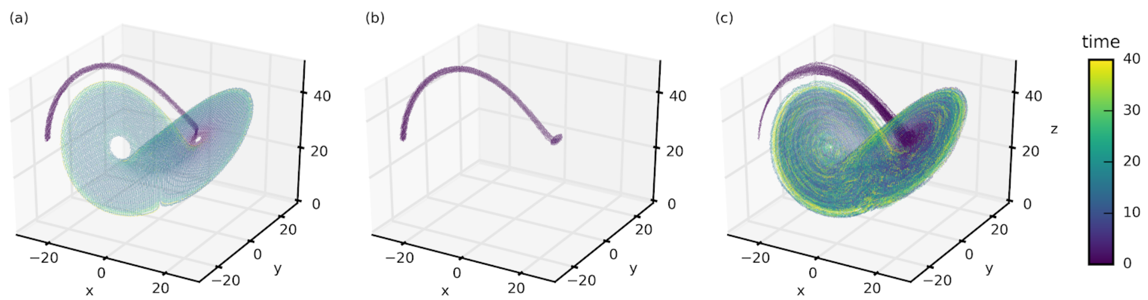
In order to explore new capabilities for this coupling issue the SEACS project focused on new stochastic frameworks to define new fusion strategies between the data and models, but also to propose new dynamical systems defined either from the physics or directly from the data. The project has been in a natural way subdivided in three research themes. In the following report, we focus on the methodological contributions issued from SEACS project according to the two methodological challenges identified in our original proposal :

- Model-driven approaches for stochastic representations of geophysical systems/dynamics (Section 1);
- Data-driven approaches for stochastic representation of geophysical systems/dynamics (Section 2).

A third theme has also naturally emerged to bridge model-driven and data-driven approaches as briefly reviewed below. These methodological contributions have been demonstrated and evaluated on upper ocean dynamics through different case-studies, including for instance space-time interpolation issues for sea surface geophysical tracers (e.g., Fablet et al., 2017, Lopez radcenco et al. 2019) as well as SQG ocean models (e.g., Resseguier et al., 2017abc). We let the reader to these references for further illustration of the thematic contributions.

1- Model-driven approaches for stochastic representations of geophysical systems/dynamics

In this first theme of study, we have proposed a new modeling principle enabling to derive stochastic representation of flow dynamics. This principle is called modeling under location uncertainty and which consists in introducing in the flow dynamics equations uncertainty variable related to the unresolved components of the flow. Such an uncertainty is formalized through the introduction of a random term that enables taking into account large-scale approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also partially known forcing. Rigorously derived from a stochastic version of the Reynolds transport theorem [Mémin 2014], this framework encompasses several meaningful mechanisms for turbulence modeling. It indeed introduces without any supplementary assumption the following pertinent mechanisms for turbulence modeling: (i) a dissipative operator related to the mixing effect of the large-scale components by the small-scale velocity; (ii) a multiplicative noise representing small-scale energy backscattering; and (iii) a modified advection term related to the so-called "turbophoresis" phenomena, attached to the migration of inertial particles in regions of lower turbulent diffusivity. This stochastic modeling framework, built from a notion of stochastic transport, relies on the usual physical conservation laws and follows exactly the same path as usual the deterministic principles. This provides to this framework a great versatility to derive stochastic representation of flows.



Lorenz 63 modem; trajectories of the deterministic system (left), a dissipative large-scale representation which get trapped in the basin of attraction of the equilibrium points (middle) and a stochastic Lorenz system enabling to visit in a faster way the region of the attractor.

In a series of papers, we have shown how such modeling can be applied to provide stochastic representations of the whole variety of the classical geophysical flows dynamics [Resseguier et al 2018]. Numerical simulations and uncertainty quantification have been performed on Quasi Geostrophic approximation (QG) of oceanic models. It has been shown that such models lead to remarkable estimation of the unresolved errors contrary to classical eddy viscosity-based models. The noise brings also an additional degree of freedom in the modeling step and pertinent diagnostic relations and variations of the model can be obtained with different scaling assumptions of the turbulent kinetic energy (i.e. of the noise amplitude). The performances of such systems have been assessed also on an original stochastic representation of the Lorenz 63 derived from the modeling under location uncertainty [Chapron et. al 2018]. In this study it has been shown that the stochastic version enabled to explore in a much faster way the region of the deterministic attractor.
In the wake of these studies we are currently working on techniques to define or calibrate the noise term from data. In that prospect, we intend to explore statistical learning techniques as define in the second axis of the SEACS project.

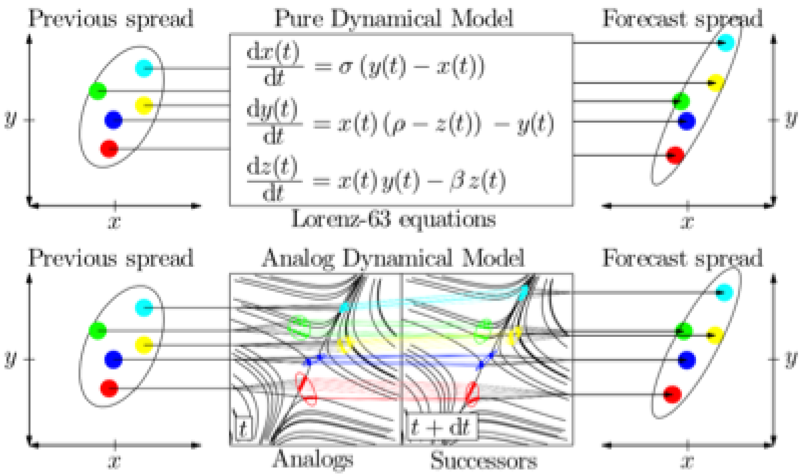
This stochastic modelling principle has been applied in the context of POD-Galerkin methods in order to design computationally efficient reduced order dynamical system. As outlined previously this uncertainty modeling methodology provides a theoretically grounded technique to define an appropriate subgrid tensor as well drift correction terms in the dynamics. The pertinence of this stochastic reduced order model has been successfully assessed on several prototypical flows at different Reynold number. It has been shown to be much more stable than the usual reduced order model construction techniques. Beyond the definition of a stable reduced order model, the modeling under location uncertainty paradigm offers a unique way to analyze from the data of a turbulent flow the action of the small-scale velocity components on the large-scale flow. Regions of prominent turbulent kinetic energy, direction of preferential diffusion, as well as the small-scale induced drift can be identified and analyzed to decipher key players involved in the flow [Resseguier 2107]. Note that these reduced order models can be extended to a full system of stochastic differential equation driving all the temporal modes of the reduced system (and not only the small-scale modes). This full stochastic system has been evaluated at moderate Reynolds number. For this flow the system has shown to provide very good uncertainty quantification properties as well meaningful physical behavior with respect to the simulation of the neutral modes of the dynamics. This study described in [Resseguier 2017] will be soon submitted to a journal paper.

Modeling under location uncertainty and the modification it undergoes in the fluid dynamics expression enables as a byproduct to decipher new physical mechanisms. This has been used to explain the mean velocity profile within the whole turbulent layer of canonical wall bounded flows, for which to date no satisfying model exists in a transitional region between the near wall area (called the viscous layer) and a more far-away turbulent region characterized by a logarithmic velocity profile. The stochastic model and its ability to introduce the action of the small-scale inhomogeneity has given rise to a theoretically well-grounded model for this transitional zone. This highlights the driving role played by turbulence inhomogeneity in shaping the large-scale flow. Numerical assessment of the results has been performed for turbulent boundary layer flows, pipe flows and channel flows at various Reynolds numbers [Pinier et. al 2019].

2- Data-driven approaches for stochastic representation of geophysical systems/dynamics

Under this theme, we have investigated two main research directions for the development of data-driven representation of geophysical dynamics. Initially, we focused on analog methods and their use in a data assimilation context. During the second half of the project, we initiated a novel direction towards neural-network representations of geophysical dynamics.

Analog Data Assimilation. As illustrated below, the key idea of analog methods is to replace an explicit dynamical model, typically stated in a discrete state-space framework or ODE/PDE-based one, by an implicit analog dynamical model. The later comes to sample new state dynamics from previously observed ones, with a sampling strategy driven by the similarity between the current state and the samples archived in a reference dataset (See Figure below). A first contribution lies in stating this analog forecasting model as a state-dependent linear Gaussian model, whose parameters are estimated online, and its exploitation in state-of-the-art data assimilation schemes, especially ensemble Kalman filters and particle filters (Lguensat et al., 2017). We have shown that the resulting data-driven model-free assimilation schemes can reach assimilation performance similar to a model-based assimilation with the true dynamical model if the number of samples in the reference dataset, where we look for analogs, is large enough. A second contribution lies in the extension of the analog data assimilation (AnDA) to high-dimensional dynamics (e.g., nD+t fields), as analog methods are very vulnerable to the curse of dimensionality. Using scale-space decomposition techniques and patch-based representations, we have shown that AnDA can scale up to ND+t dynamics and significantly outperform model-based optimal interpolation schemes for the space-time interpolation of geophysical fields with very high missing data rates (Fablet et al., 2017; Lopez radcenco et al., 2019).



Neural network representations. With a view to bridging model-based and data-driven representations, neural networks arise as relevant schemes. More precisely, some neural network architectures can be regarded as numerical solvers for ODEs and PDEs (eg, Fablet et al., 2018). Residual networks are typical examples of such architectures which can be interpreted as Euler explicit integration schemes of an underlying ODE. This analogy provides the basis for exploring neural network (NN) architecture for the identification of governing equations from observation datasets following the pioneering work of Brunton et al. In this context, we may point out different contributions: bilinear NNs associated with RK(4) solver for the identification of second-order polynomial governing equations (Fablet et al., 2018), learning neural ODE representations from irregularly-sampled and/or noisy datasets (Nguyen et al., 2019), learning latent dynamics for partially-observed systems (i.e., when some components of the state of interest are never observed). A second topic of contributions include the exploitation of NN representations for data assimilation issues either based on Kalman-like NN architectures (Ouala et al., 2018) or on energy-based representations (Fablet et al., 2019). These approaches are particularly appealing to address the end-to-end learning of representations of geophysical dynamics from observation datasets, which involve multivariate irregularly-sampled processes.

3- Nonlinear Non-Gaussian data-model coupling strategy

Estimating the parameters of geophysical dynamic models is an important task in Data Assimilation (DA) technique used for forecast initialization and reanalysis. In the past, most parameter estimation strategies were derived by state augmentation, yielding algorithms that are easy to implement but may exhibit convergence difficulties. The Expectation-Maximization (EM) algorithm is considered advantageous because it employs two iterative steps to estimate the model state and the model parameter separately.

We have proposed an original inference algorithm for state-space models with a dynamic equation which has no close form nor any expression as a PDE. This problem is clearly analytically intractable and the proposed solution is an algorithm which combines sequential Monte Carlo method and non-parametric estimation with a stochastic EM optimization algorithm. The state reconstruction is performed by an efficient Particle Smoother which combines a Conditional Particle Filter and Backward simulation. The stochastic EM like algorithm allows to retrieve an estimation of the dynamical model, of the posterior distribution of the state and of the variance of the observation error from a noisy time series (or a time series observed with error in measurements) [Chau 2019].

In another work, we proposed a novel ensemble formulation of the Maximization step in EM that allows a direct optimal estimation of physical parameters using iterative methods for linear systems. This departs from current EM formulations that are only capable of dealing with additive model error structures. This contribution shows how the EM technique can be used for dynamics identification problem with a model error parameterized as arbitrary complex form. The proposed technique is used for the identification of stochastic subgrid terms that account for processes unresolved by a geophysical fluid model. This method, along with the augmented state technique, has been evaluated to estimate such subgrid terms through high resolution data. Compared to the augmented state technique, our method is shown to yield considerably more accurate parameters. In addition, in terms of prediction capacity, it leads to smaller generalization error as caused by the overfitting of the trained model on presented data and eventually better forecasts [Yang and Mémin 2018].

4- Publications, animation and outreach

We complement the above scientific description of the main results and contributions issued from SEACS project with a list of: selected publications, involved Phd, postdoc and engineer members, animation and outreach activities, leveraging effects.

Selected publications

B. CHAPRON, P. DÉRIAN, E. MÉMIN, V. RESSEGUIER. Large scale flows under location uncertainty: a
consistent stochastic framework, in Quarterly Journal of the Royal Meteorological Society, vol. 144, no 710, pp. 251-260, 2018.

T.T.T. CHAU, Non-parametric methodologies for reconstruction and estimation in nonlinear state space models, PhD U. Rennes I, Applied Math. 2019.

R. Fablet, J. Verron, B. Mourre, B. Chapron, A. Pascual. Improving mesoscale altimetric data from a multi-tracer convolutional processing of standard satellite-derived products. IEEE Trans. on Geoscience and Remote Sensing, 2017.

R. Fablet, P. Viet, R. Lguensat. Data-driven Methods for Spatio-Temporal Interpolation of Sea Surface Temperature Images. IEEE Trans. on Computational Imaging, 2017.

R. Fablet, P. Viet, R. Lguensat, P.H. Horrein, B. Chapron. Spatio-Temporal Interpolation of Cloudy SST Fields Using Conditional Analog Data Assimilation. Remote Sensing, 10(2):310, 2018.

R. Fablet, L. Drumetz, F. Rousseau. End-to-end learning of energy-based representations for irregularly-sampled signals and images. Preprint, 2019.
L. Gomez-Navarro, R. Fablet, E. Mason, A. Pascual, B. Mourre, E. Cosme, J. Le Sommer. SWOT spatial scales in the Western Mediterranean Sea derived from pseudo-observations and an ad-hoc filtering. Remote Sensing, 2018.
R. Lguensat, P. Tandeo, P. Aillot, R. Fablet. The Analog Data Assimilation. Monthly Weather Review, 2017.
R. Lguensat, P. Viet, M. Sun, G. Chen, F. Tenglin, B. Chapron, R. Fablet. Data-driven Interpolation of Sea Level Anomalies using Analog Data Assimilation. Remote Sensing, 2019.
M. Lopez Radcenco, A. Pascual, L. Gomez-Navarro, A. Aïssa-El-Bey, B. Chapron, L. Gomez-Navarro, R. Fablet. Analog Data Assimilation of Along-track Nadir and Wideswath SWOT Altimetry Observations in the Western Mediterranean Sea. IEEE JSTARS, 2019.
E. MÉMIN. Fluid flow dynamics under location uncertainty, in Geophysical & Astrophysical Fluid Dynamics, vol. 108, no 2, pp. 119-146, 2014.
D. Nguyen, S. Ouala, L. Drumetz and R. Fablet. EM-like Learning Chaotic Dynamics from Noisy and Partial Observations, arxiv.org/abs/1903.10335, 2019.
S. Ouala, R. Fablet, C. Herzet, B. Chapron, A. Pascual, F. Collard, L. Gaultier. Neural-Network-based Kalman Filters for the Spatio-Temporal Interpolation of Satellite-derived Sea Surface Temperature. Remote Sensing, 2018.
S. Ouala (...) and R. Fablet. Learning Latent Dynamics for Partially-Observed Chaotic Systems. arxiv.org/abs/1907.02452, 2019.
B. PINIER, E. MÉMIN, S. LAIZET, R. LEWANDOWSKI, Stochastic flow approach to model velocity profile of wall-bounded flows, Physical Review E, vol. 99, no 6, 063103, 2019.
V. RESSEGUIER, E. MÉMIN, B. CHAPRON. Geophysical flows under location uncertainty, Part I Random
transport and general models, in Geophysical & Astrophysical Fluid Dynamics, vol. 111, no 3, pp. 149-176, 2017.
V. RESSEGUIER, E. MÉMIN, B. CHAPRON. Geophysical flows under location uncertainty, Part II Quasi-geostrophy and efficient ensemble spreading, in Geophysical & Astrophysical Fluid Dynamics, vol. 111, no 3, pp. 177-208, 2017.
V. RESSEGUIER, E. MÉMIN, B. CHAPRON. Geophysical flows under location uncertainty, Part III SQG and
frontal dynamics under strong turbulence conditions, in Geophysical & Astrophysical Fluid Dynamics, vol. 111, no 3, pp. 209-227, 2017.
V. RESSEGUIER, E. MÉMIN, D. HEITZ, B. CHAPRON. Stochastic modelling and diffusion modes for proper
orthogonal decomposition models and small-scale flow analysis, in J. Fluid Mech., vol. 828, 2017.
V. RESSEGUIER, Mixing and fluid dynamics under location uncertainty, PhD U. Rennes I, Applied Math. 2017.
F. Rousseau, L. Drumetz, R. Fablet. Residual Networks as Flows of Diffeomorphisms. JMIV, SI “Mathematical Foundations of Deep Learning in Imaging Science”, 2019.
A. Sánchez-Román, L. Gómez-Navarro, R. Fablet, D. Oro, E. Mason, J. M. Arcos, S. Ruiz & A. Pascual. Rafting behaviour of seabirds as a proxy to describe surface ocean currents in the Balearic Sea. Scientific Reports, 9(17775), 2019.
Y. YANG, E. MÉMIN. Estimation of physical parameters under location uncertainty using an Ensemble^2-
Expectation-Maximization algorithm, in Quaterly journal of the royal meteorological society, vol. 145, no 719, pp 418-433, 2018.

AWARDS

V. Resseguier received the best PhD award 2018 in apply mathematics of the GAMNI/SMAI

SEACS PhDs, Postdocs & Engineers

- B. Boussidi (2013-2016) (PhD ARED/FP IMT Atl.)
- R. Lguensat (2014-2017) (PhD FP IMT Atl.)
- M. Lopez Radcenco (2015-2018) (PhD Cominlabs)
- T.T.T.. Chau (2015-2018) (PhD Lebesgue)
- P. Derian (2017-2018) (Postdoc Cominlabs)
- Phi Viet Nguyen (2016-2017) (Engineer, Teralab grant)
- C. Gonzalez (2016-2017) (Postdoc ESA)
- V. Resseguier (2015-2018) (PhD Ifremer/inria)
- S. Ouala (2017-2020) (PhD Bretel/ODL)
- W. Bauer(2018-2019) (Postdoc Region/FP Inria)
- Long Li (2017-2020) (PhD IRMAR/U. Rennes I)

Animation & visiting scientists

- Ocean Remote Sensing Synergy Summer School (each year since 2015)
- Workshop on Stochastic Weather Generators (Vannes, 2016)
- National Symposium on Data Assimilation (2018)
- Doctoral course on Data Science for Geoscience (2018-2019)
- 1st Workshop on AI for Geosphysical dynamics (Brest, 2018)
- 2nd Workshop on AI for Geosphysical dynamics (Brest, 2019)
- over 40 short incoming visits (from a few days to a few months): eg,
    - Prof. D. Giannakis, New York Univ.
    - Prof. S. Gotwald, Univ. of Sydney
    - Prof. C. Wikle, Uinv. of Missoury
    - Prof. S. Brunton, Univ. of Washington
    - Dr. A. Pascual, Sen. Scientist, IMEDEA

Leveraging and structuring effects

- ESA postdoc fellowship (C. Gonzalez, 2016-2017)
- Teralab grant (2016-2018, computational resources + 18m engineer)
- H2020 Eurosea (2020-)
- OSTST MANATEE (co-PI, R. Fablet, CNES, 2017-2020)
- R&T CNES ANDA (PI: R. Fablet, 2019)
- ANR BOOST SWOT (PI. E. Cosme, 2018-2021)
- LEFE/MANU IA-OAC (PI. R. Fablet, 2019-2021)
- Isblue theme « Observing Systems » (co-PIs B. Chapron & R. Fablet)
- New collaborations: SHOM (Brest), IGE (Grenoble), IMEDEA (Spain), Univ. of Washington (USA), Imp. College (UK)
- Industrial partnerships (eg, PhD fundings): ACRI-ST, ITGA, CSTB, OceanDataLab, CLS, Eodyn, OceanNext
- “Brittany council” postdoc scholarship SeaStorm (W. Bauer)