Estimation of probabilistic ocean information using machine learning methods applied to the North Atlantic

Start date : Autumn 2023 (3 year contract, funding is already acquired)

Host lab : Laboratoire d'Océanographie Physique et Spatiale, Institut Universitaire Européen de la Mer, Plouzané, France.

Supervisors: Sally Close (sally.close@univ-brest.fr ; UBO), Pierre Tandeo (IMT Atlantique), Guillaume Maze (Ifremer)

Student profile : Candidates should have a masters degree in either physical oceanography, OR in a numerical/mathematical subject that includes a substantial statistical component, OR in both domains. Prior experience of programming in Python will be an advantage. The project encompasses both physical oceanography and statistics, and the recruited person will need a strong interest in both subjects. Funding has been acquired to enable the student to receive further training in oceanography or statistical methods if necessary, according to their prior experience and training needs. The position is open to both French and international applicants.

Summary : The atmosphere plays an important role in determining the ocean's variability, and this is reflected in the signals of ongoing climate change that are detected in numerous ocean variables. However, part of the ocean's variability is not driven by the atmosphere, but instead is random, and generated by the ocean itself through non-linear processes. To quantify the mechanisms and effects of climate change in the ocean-atmosphere system, we need to understand the relative importance of the atmospherically-influenced and "random" parts of the variability. This project will aim to address this need by using novel machine learning methods to describe and predict the random part of the ocean's variability. These techniques will be less energetically-expensive and more flexible than presently-used methods. Once the development phase of the project has been completed, the student will apply the methods that they have developed to the North Atlantic region, and will use their artificially-generated data to study local processes that are critical to the strength of the global ocean circulation.

Background and context for the project :

Ongoing climate change is a critical problem for society, with wide-ranging potential impacts. Physically, this change is associated with an excess of heat in the climate system, and the ocean is the main reservoir in which this excess heat is stored. Through its ability to store, transport and release heat, the ocean will play a key role in determining future climate evolution by modulating atmospheric warming. Quantifying the mechanisms and effects of climate change thus requires that we understand and quantify the effects of ocean-atmosphere interaction.

The detection and attribution of the part of oceanic variability that is due to the atmosphere is complicated by the partly-random nature of the ocean's variability. In response to even a simple, periodic atmospheric forcing, the ocean is able to generate

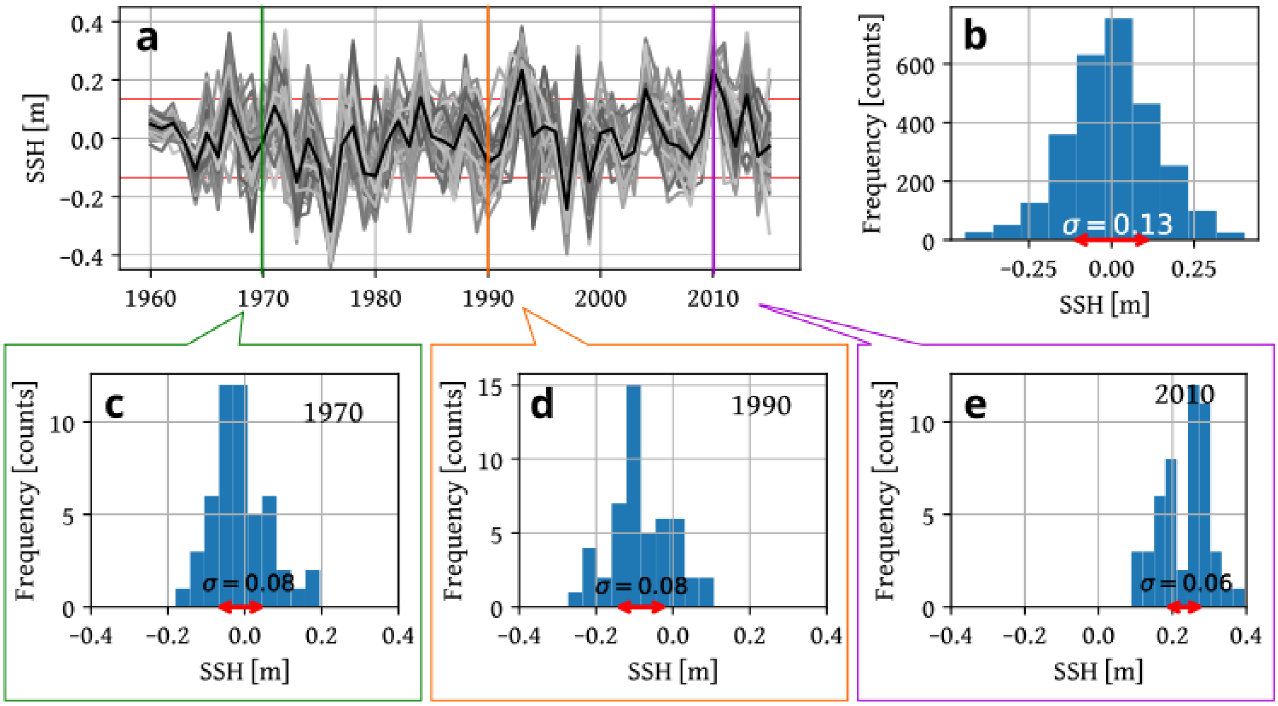

Figure 1: **a:** Time series of the sea surface height (SSH) for 50 ensemble simulations subject to identical atmospheric forcing but different initial conditions (grey) and the ensemble mean (black). **b:** histogram showing the probability distribution for all 50 members over all time steps. **c-e:** time-varying ensemble PDF, composed of the 50 members at a given time step for 1970 **(c)**, 1990 **(d)** and 2010 **(e)**. The green, orange and purple lines in (a) indicate the instants to which the ensemble PDF (c-e) correspond.

chaotic, intrinsic variability at multiple space and time scales. A time-varying probability distribution (probability density function: PDF, Fig 1) then describes the evolution of a chosen quantity, since a range of possible values may be obtained at any given time due to the random nature of the intrinsic contribution. This PDF may be estimated using ensemble model experiments. In ocean "hindcast" experiments, where ocean models are forced by prescribed atmospheric fields, the ensemble mean of a given variable is sometimes taken to approximate the component of the oceanic variability that is due to the influence of the atmosphere, while the PDF spread represents the range of the estimated intrinsic variability.

The time-varying PDF cannot, in theory, be estimated from either observations or a single model experiment alone: an ensemble simulation approach is required. Due to the high cost of both production and analysis, the size of the ensemble is usually determined by practical considerations. This restriction may potentially affect the estimate of the PDF and ensemble statistics. This project will aim to address these issues by developing new, environmentally-sustainable and easy-to-implement solutions to estimate the PDF. The project will contribute to the *Artificial probabilistic information for ocean-climate applications* project (link), which has been funded by the *Agence Nationale de la Recherche*, and which will investigate methods of obtaining probabilistic ocean information using a number of approaches.

Aims and expected contribution of the project :

The primary aims of the project will be to produce estimates of the ensemble PDF directly from a reduced number of simulations, with the ultimate goal of developing a method for application to observational data. The student will also investigate the viability of creating artificial simulations. This will allow us to obtain the information supplied by ensemble simulation at greatly reduced computational cost, and without dependence on model physics. It will also prepare the way for the application of such methods to high resolution data over large regions. The outputs will be used to study the influence of intrinsic variability on deep convection in the North Atlantic subpolar gyre, where we will aim to better understand the mechanisms that determine the onset of deep convection and to identify indicators of ocean susceptibility to this process.

In practice, the student will use an existing large ensemble simulation, and will implement a series of neural networks and analogue methods to generate artificial ensemble statistics using a reduced number of simulations. The initial development work will be performed on small subregions. A variety of flow regimes will be considered, and the size of the regions will then be expanded once the method has been verified on these test regions.

The student will then apply the method to high-resolution model output in the North Atlantic subpolar gyre, with the aim of investigating the effect of intrinsic variability on the onset of deep ocean convection. This process is critical in setting the strength of the Meridional Overturning Circulation, but is difficult to predict, and often poorly represented in models, in part due to the high resolution necessary to resolve relevant small-scale processes. The student will create artificial PDF of sea surface height and the mixed layer depth (a metric of deep convection), and analyse the covariance of the artificial ensemble statistics with relevant atmospheric variables.

Training and prospects :

The supervisory team has complementary expertise in physical oceanography and statistics, and the full support necessary to undertake the project will thus be available to the student. The student will develop skills in both areas, and in computational techniques, including working with supercomputing infrastructure. These are skills that may be relevant to careers in either pure science, or a number of applied industries following the PhD. The student will also obtain core skills in planning research, and in oral and written scientific communication. Funding has been acquired to enable the student to attend further training in either physical oceanography or statistics (e.g. via attendance at a summer school) if necessary. S/he will be based at the Laboratoire d'Océanographie Physique et Spatiale, but collaboration with other groups, both in France and overseas, will be encouraged.

Background reading :

Penduff et al. (2014), Ensembles of eddying ocean simulations for climate. *Clivar Exchanges*

Lguensat et al. (2017), The analog data assimilation. *Monthly Weather Review*

Maze et al. (2017), Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean. *Progr. Oceanogr.*

Close et al. (2020), A means of estimating the intrinsic and atmospherically-forced contributions to sea surface height variability applied to altimetric observations. *Progr. Oceanogr.*

Contact information : for any questions or to discuss the project, please contact Sally Close: sally.close@univ-brest.fr