# SPECIAL PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

| | |
|---|---|
| **Project Title:** | Stochastic Coastal/Regional Uncertainty Modelling 2: consistency, reliability, probabilistic forecasting, and contribution to CMEMS ensemble data assimilation |
| **Computer Project Account:** | spgrver2 |
| **Start Year - End Year :** | 2018 - 2020 |
| **Principal Investigator(s)** | Vassilios D. Vervatis (1), Pierre De Mey-Frémaux (2) |
| **Affiliation/Address:** | (1) National Kapodistrian University of Athens (UoA) ..<br>(2) Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS) |
| **Other Researchers (Name/Affiliation):** | Sarantis Sofianos (1), Nadia Ayoub (2), Bénédicte Lemieux-Dudon (2) |

The following should cover the entire project duration.

## Summary of project objectives

(10 lines max)

The ECMWF-SP resources were used in a joint project named SCRUM2, within the CMEMS Service Evolution under Lot 5: "cross-cutting developments on observation, assimilation and product quality improvements". The proposed work was built upon the previous SP project SCRUM and was based on stochastic modelling of ocean physics and biogeochemistry in coastal and regional data assimilation systems, including methods suitable to assess the reliability of Ensembles.

## Summary of problems encountered

(If you encountered any problems of a more technical nature, please describe them here.)

We did not encounter any problems in the previous year. We are thankful to the ECMWF technical and scientific advisory committee for their help during the project, solving promptly all issues.

## Experience with the Special Project framework

(Please let us know about your experience with administrative aspects like the application procedure, progress reporting etc.)

The use of the ECMWF HPC Facilities was very important, in order to complete successfully the joint CMEMS research project mentioned above. All administrative issues were easy to follow.

## Summary of results

(This section should comprise up to 10 pages, reflecting the complexity and duration of the project, and can be replaced by a short summary plus an existing scientific report on the project.)

In this section, we give a brief summary of the research activities carried out in the context of a joint CMEMS SE project. For a more detailed description the reader is referred to the previous SP reports, and to the Sourceforge and Zenodo documents and links provided at the end of this section.

The project aimed at strengthening CMEMS in the areas of regional and coastal ocean uncertainty modelling, empirical Ensemble consistency verification, Ensemble probabilistic forecasting, and Ensemble data assimilation. Specifically, the work was based on stochastic modelling of ocean physics and biogeochemistry, in the context of Ensemble forecasting systems, and included methods to assess the suitability of Ensembles for probabilistic forecasting.

The notion of Ensemble empirical consistency was meant as the consistency of the model pdf (as approximated by an Ensemble) with the pdf of observations. An underlying notion is the joint probability on the right-hand-side of the equal sign in the Bayes theorem: if model and observation pdfs are disjoint, the joint probability is zero, and any solution found by convex DA will have in effect zero probability to occur. In order to explore ensemble empirical consistency, we defined three approaches (basic statistics, rank histograms and array-space analysis) and several consistency criteria, all of them gathered in a toolbox interfaced with NEMO.

On the other hand, the implementation of probabilistic scores is a fairly new area in ocean sciences. Published algorithms (brier score and Hersbach CRPS) were implemented within the aforementioned toolbox and the code was interfaced with a Python driver. The scores were tested for several applications with Quasi-Reliable testbed (e.g. using synthetic observations from a member of the ensemble excluded later from the analysis), as well as with real observations from the CMEMS catalogue. We illustrated the application of those probabilistic scores to an upwelling event off the Galician coasts.

Overall, the following activities were carried out in this SP:

- We produced short- to medium-range Ensembles over repeated periods to mirror the MFC operational practices. This permitted to take into account the Ensemble spin-up period (up to 10 days) and give access to performance categorization based on forecast lead time (up to 20 days).
- Ensembles were generated making use of the European Center for Medium-Range Weather Forecasts Ensemble Prediction System (ECMWF-EPS; 50 members). The goal using multivariate atmospheric Ensembles was to obtain a controlled (without DA) ocean Ensemble spread and focus on variables and timescales of particular interest to CMEMS.
- We revisited the stochastic approaches investigated in the previous SP project to optimize further the Ensemble strategies for coastal and regional configurations.
- We examined the new Ensembles focusing on empirical consistency methods and probability scores (and associated statistical properties). Data from Thematic Assembly Centers (TAC) and arrays from the CMEMS catalogue were used in this task, including Sentinel 3A data.
- New empirical consistency approaches took into account correlated observational errors, aiming at updating the R matrix.
- We developed and test probabilistic scores based on Ensemble forecasts. This required defining the occurrence of "events" based on "stake variables" and examined how the scores perform depending on forecast lead time.

The codes and compendia are available upon request. The reader is also referred to the following Sourceforge and Zenodo documents and links:

- Toolbox in SDAP Sourceforge
- NEMO code for stochastic implementation https://doi.org/10.5281/zenodo.2556529
- Consistency formalism https://doi.org/10.5281/zenodo.3688509
- Probabilistic formalism http://doi.org/10.5281/zenodo.3743087

## List of publications/reports from the project with complete references

The most recent publications this year (acknowledging the ECMWF SP resources) are:

Vervatis, D. V., P. De Mey-Frémaux, N. Ayoub, J. Karagiorgos, M. Ghantous, M. Kailas, C.-E. Testut and S. Sofianos (2021), Assessment of a regional physical-biogeochemical stochastic ocean model. Part 1: Ensemble generation, Ocean Modell., 160, 101781, https://doi.org/10.1016/j.ocemod.2021.101781

Vervatis, D. V., P. De Mey-Frémaux, N. Ayoub, J. Karagiorgos, S. Ciavatta, R.J.W. Brewin and S. Sofianos (2021), Assessment of a regional physical-biogeochemical stochastic ocean model. Part 2: Empirical consistency, Ocean Modell., 160, 101770, https://doi.org/10.1016/j.ocemod.2021.101770

## Future plans

(Please let us know of any imminent plans regarding a continuation of this research activity, in particular if they are linked to another/new Special Project.)

We plan to continue our activities on stochastic ocean modelling in coastal and regional domains. We are currently running a continuation of this work with a new SP for the year 2021 (i.e. SDAP/NEMO System https://www.ecmwf.int/en/research/special-projects/spgrver2-2021).