SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2020
Project Title:	Stochastic Coastal/Regional Uncertainty Modelling 2: consistency, reliability, probabilistic forecasting, and contribution to CMEMS ensemble data assimilation
Computer Project Account:	SPGRVER2
Principal Investigator(s):	Vassilios D. Vervatis (1), Pierre De Mey-Frémaux (2)
Affiliation:	(1) National Kapodistrian University of Athens (UoA) .. (2) Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Sarantis Sofianos (1), Nadia Ayoub (2), Bénédicte Lemieux-Dudon (2)
Start date of the project:	29 May, 2018
Expected end date:	31 December, 2020

Computer resources allocated/used for the current year and the previous one
(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	18MSBU	~18MSBU	2MSBU	~1.3MSBU
Data storage capacity	(Gbytes)	24TB	~24TB	3TB	~2TB

Summary of project objectives (10 lines max)

The ECMWF-SP resources are used in a joint project named SCRUM2, within the CMEMS Service Evolution open tender under Lot 5: "cross-cutting developments on observation, assimilation and product quality improvements". The work proposed here builds upon, and expands, the previous ECMWF-SP project SCRUM. The work is based on stochastic modelling of ocean physics and biogeochemistry, in the context of coastal/regional Ensemble Data Assimilation (EDA) forecasting systems, and includes methods suitable to assess the reliability of Ensembles in probabilistic assimilation systems

Summary of problems encountered (10 lines max)

We did not encounter any problems, thanks to the additional resources obtained the previous year. We are thankful for the fact that our request for additional resources was accepted by the ECMWF technical and scientific advisory committee.

Summary of plans for the continuation of the project (10 lines max)

This is the third and last year of the current ECMWF-SP and also the last year of the CMEMS SE SCRUM2 joint project. An Horizon 2020 project named DIOGENE (Data assimilation to Improve Ocean forecasting by GEnerating Novel and Efficient algorithms), has been recently submitted and is currently under evaluation. This Horizon 2020 project would be (upon acceptance) a continuation of the previous CMEMS Service Evolution projects and for this, we are going to submit a request (provided in a separate form) for a new Special Project for the years 2021-2023.

List of publications/reports from the project with complete references

De Mey-Frémaux P, Ayoub N, Barth A, Brewin R, Charria G, Campuzano F, Ciavatta S, Cirano M, Edwards CA, Federico I, Gao S, Garcia Hermosa I, Garcia Sotillo M, Hewitt H, Hole LR, Holt J, King R, Kourafalou V, Lu Y, Mourre B, Pascual A, Staneva J, Stanev EV, Wang H and Zhu X (2019) Model-Observations Synergy in the Coastal Ocean. Front. Mar. Sci. 6:436. doi: 10.3389/fmars.2019.00436.

Vervatis, D. V., P. De Mey-Frémaux, et al., Assessment of a physical-biogeochemical stochastic model in a high-resolution Bay of Biscay configuration. Part 1: generation and first analysis, Ocean Modelling, under review.

Vervatis, D. V., P. De Mey-Frémaux, et al., Assessment of a physical-biogeochemical stochastic model in a high-resolution Bay of Biscay configuration. Part 2: empirical consistency, Ocean Modelling, under review.

J. Karagiorgos, V. Vervatis, B. Lemieux-Dudon, P. De Mey-Frémaux, N. Ayoub, S. Sofianos (2020), On the Use of Atmospheric Ensembles to Generate Physical-biogeochemical Ocean Model Uncertainties, AGU, Ocean Sciences Meeting 2020.

De Mey-Frémaux, P., 2020: Array-space analysis tools in SCRUM and SCRUM2 (Version 2). Zenodo. http://doi.org/10.5281/zenodo.3688509

Vervatis, V.D., Pierre De Mey-Frémaux, Nadia Ayoub, Sarantis Sofianos, Charles-Emmanuel Testut, Marios Kailas, John Karagiorgos, Malek Ghantous, 2019: Physical-biogeochemical regional ocean model uncertainties stemming from stochastic parameterizations (Version 1). Zenodo. http://doi.org/10.5281/zenodo.2556530

Lemieux-Dudon, B., 2020: Probabilistic Scores for the verification of ensemble prediction systems - Brier, DRPS, CRPS underlying mathematics. Zenodo. http://doi.org/10.5281/zenodo.3743087

Summary of results

If submitted **during the first project year**, please summarise the results achieved during the period from the project start to June of the current year. A few paragraphs might be sufficient. If submitted **during the second project year**, this summary should be more detailed and cover the period from the project start. The length, at most 8 pages, should reflect the complexity of the project. Alternatively, it could be replaced by a short summary plus an existing scientific report on the project attached to this document. If submitted **during the third project year**, please summarise the results achieved during the period from July of the previous year to June of the current year. A few paragraphs might be sufficient.

Here, we only present results for the most demanding R&D activities in terms of computational resources, carried out the previous year in the context of this Special Project.

More in details, three approaches were envisaged to generate ocean physics and biogeochemical uncertainties: (1) the one being a NEMO stochastic implementation, (2) the other incorporating an atmospheric Ensemble using the ECMWF-EPS system, and (3) merging the two previous approaches. For the stochastic approach we used the SPPT and SPP schemes based on AR1 processes (**Tables 1 and 2**), with an increased uncertainty compared with the previous ECMWF SP (https://www.ecmwf.int/en/research/special-projects/spgrverv-2016) for the wind from 0.3 to 0.4 st.dev., and for the biogeochemical SMS(C) terms from 0.6 to 0.8 st.dev., in order to deliberately inflate the Ensemble spread. For the same reason, we used simultaneously the SPUF stochastic method sampling gradients from the T/S state vector performing random walks. We performed only one random walk and only for physics (i.e. EOS ρ(T,S,p) perturbations), as a proof of concept for computational efficiency in the context of an operational system (although a larger number of random walks and/or SPUF biogeochemical perturbations were possible, but much costlier in terms of resources). A total number of 50 members were performed in each Ensemble experiment.

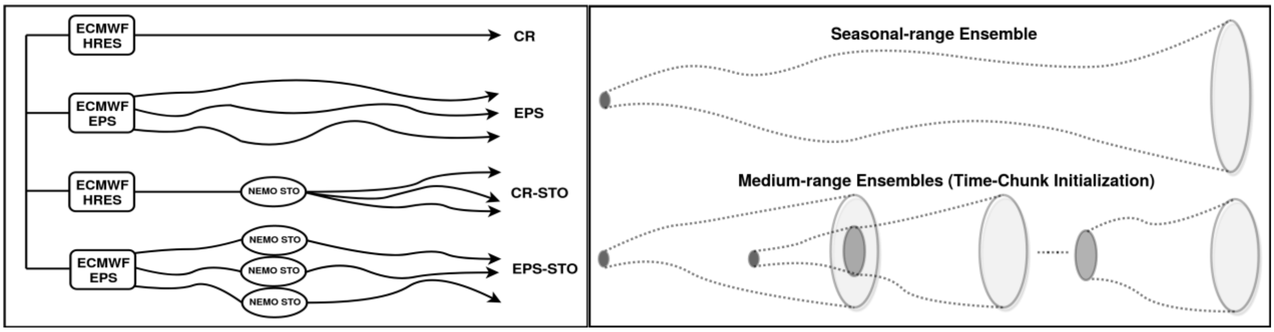
Ensemble exp. (50 members)	**Contents**
Ens4 EPS-Seasonal	**Seasonal-range** Ensemble incorporating the atmospheric **ECMWF-EPS** system
Ens5 CR-STO	Stochastic **time-chunked** Ensemble based on SPPT, SPP and SPUF methods incorporating the **ECMWF-CTRL** system
Ens6 EPS-STO	Combining the Ens4 and Ens5 Ensembles, incorporating the atmospheric **ECMWF-EPS** system, in the form of a **time-chunked** Ensemble
Ens7 EPS	Short- to medium-range Ensemble incorporating the atmospheric **ECMWF-EPS** system, in the form of a **time-chunked** Ensemble

Table 1: Ensemble generation in this project (the numbering of Ensemble experiments continuous from the previous ECMWF SP https://www.ecmwf.int/en/research/special-projects/spgrverv-2016).

Physics	**Atmospheric forcing**	**SPPT**: Wind, Tair, SLP
	Parameters	**SPP**: cd, ce, ch, cb
	Ocean state	**SPUF**: EOS ρ(T,S,p)
Biogeochemistry	**SPPT**: sources-minus-sinks of 24 prognostic variables SMS(C)	

Table 2: Stochastic methods and perturbed variables implemented in CR-STO (i.e. Ens5) and EPS-STO (i.e. Ens6) Ensemble experiments.

In **Schematic 1**, we illustrate the Control Run (CR), the different Ensemble approaches and initialization techniques (**Table 1**). In more details, we performed a seasonal-range Ensemble (i.e. Ens4) and short- to medium-range Ensembles over repeated periods to mirror the MFCs operational practices (i.e. Ens5-7). The latter approach permits to take into account the Ensemble spin-up period and give access to aging errors within a given forecast Lead Time (LT). Hereafter, we will refer to this approach as time-chunked Ensembles, meaning successive segments of time over which Ensemble calculations are conducted.



Schematic 1: Ensemble approaches (cf. **Table 1**) and initialization techniques.

In **Fig 1-2**, we show results for CHL and SST model uncertainties for the period under investigation between Dec.2016 and Jun.2017. CHL model errors appear to be in phase between open-ocean and coastal regions, pertaining to the fact that primary and secondary blooms occur with a lag at the beginning and end of spring (**Fig 1**). Another remark is that the most effective Ensemble approach in terms of biogeochemical uncertainties is the stochastic approach. The ECMWF-EPS model errors have a moderate impact in terms of biogeochemical uncertainties compared with stochastic modelling. Model errors with increased forecast LT (e.g. seasonal-range Ensemble), show moderate increase in biogeochemical model errors mostly open-ocean.

In **Fig 2**, we present the SST for the CR and the SST model uncertainties for the EPS and EPS-STO time-chunked Ensembles on two different periods and forecast LTs. It is verified that the EPS-STO has a greater impact in terms of model uncertainties in physics compared with the EPS locally augmenting the model spread. Both methods are more efficient on generating SST model errors with a few days forecast LT during spring, mainly because of the thermocline shoaling.

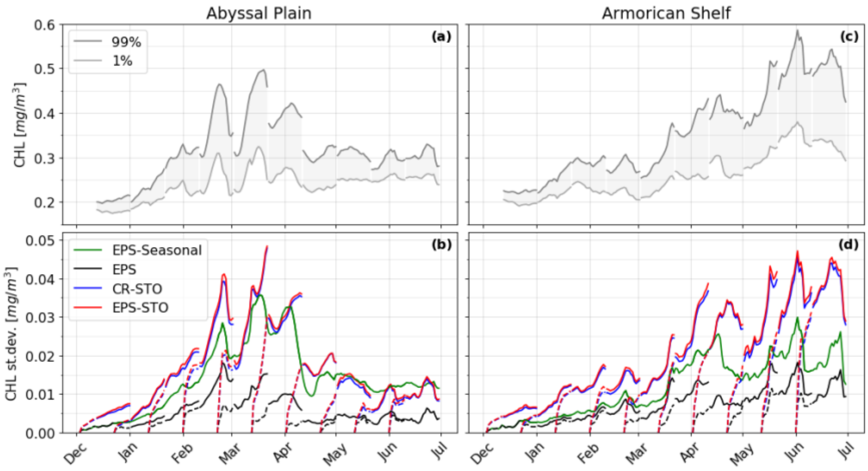


Fig 1: Model Ensembles in the form of time-chunks and seasonal-range: (a-c) CHL quantiles and (b-d) model spread in mg/m3 for the Abyssal plain and the Armorican shelf.

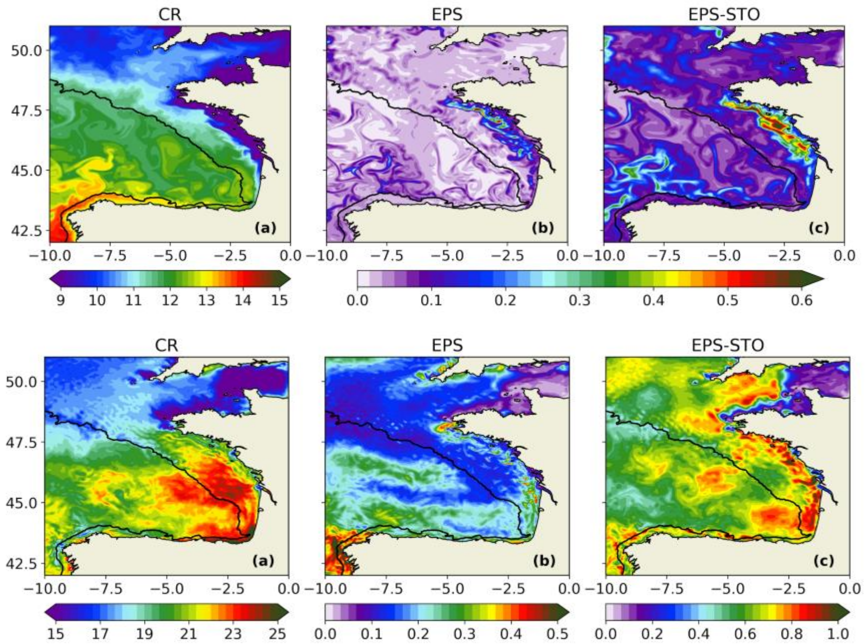


Fig 2: (a) CR and (b-c) SST model uncertainties in oC for the EPS and EPS-STO time-chunked Ensembles on two different periods and forecast LTs; upper panels: 20170208 (LT: 28-day), lower panels: 20170619 (LT:19-day).

The main source of model errors in ocean forecasting systems is the wind forcing, with the other perturbed variables having a moderate impact on model uncertainties in physics. Here, we investigate the model spread of the wind forcing for the Ensembles generated in this project, to explain the efficiency of stochastic modelling to generate larger physical-biogeochemical model uncertainties, in comparison with the ECMWF-EPS approach. In **Fig 3**, we present the Ensemble mean and spread of the wind forcing for the EPS, CR-STO and EPS-STO Ensembles on 20170501. An important attribute for all three Ensembles is that their means are similar to each other (**Fig 3a-c**). Moreover, we observe that the CR-STO has smaller scales in **Fig 3b** than the other means, because it uses the ECMWF-HRES at 9km resolution, whilst the ECMWF-EPS has a coarser resolution at 18km and smoother wind patterns. Overall, the EPS can be considered as a complementary approach to the CR-STO, locally augmenting the wind spread in areas with spatial gradients, resulting in the EPS-STO simulation (**Fig 3d-f**). However, we argue that for basin scale and global domains the ECMWF-EPS impact on generating ocean model errors may be comparable to the stochastic approaches (**Fig 4**).

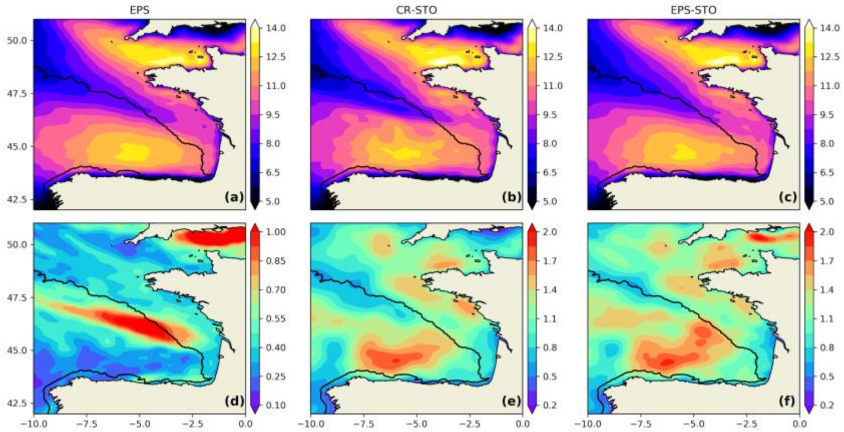


Fig 3: (a-c) Ensemble mean and (d-f) spread for the wind modulus in m/s for the EPS, CR-STO and EPS-STO Ensembles on 20170501. Note the different colorbars between the EPS and the two stochastic approaches CR-STO and EPS-STO.

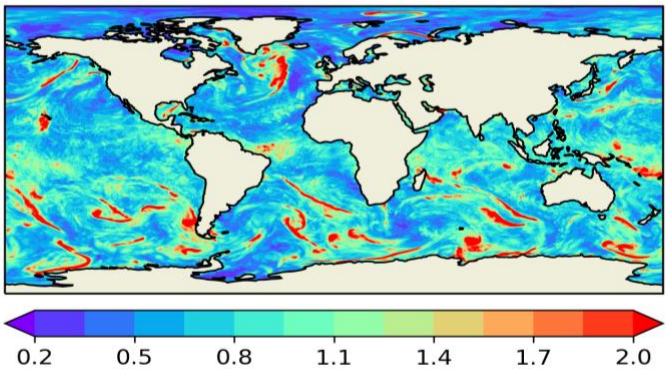


Fig 4: Same as **Fig 3d** for the ECMWF atmospheric global model (land wind values are masked). Note that here the EPS colorbar is the same with those used for the stochastic approaches CR-STO and EPS-STO in **Fig 3e-f**.