LATE REQUEST FOR A SPECIAL PROJECT 2016–2018

MEMBER STATE: Greece, France ……………………………………………………......

Principal Investigators: Vassilios D. Vervatis (1), Pierre De Mey (2)………………….………

Affiliation:
(1) National Kapodistrian University of Athens (UoA)..…………
(2) Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)..……………………………………………

Address:
(1) Ocean Physics And Modelling group (OPAM), Faculty of Physics, Division of Environmental Physics and Meteorology, Campus Building PHYS-V, 15784 Athens, Greece…………………
(2) LEGOS/CNES 18, av. Edouard Belin, 31401 Toulouse cédex 9, France ……………………………………………………………

E-mail:
(1) vervatis@oc.phys.uoa.gr ………………………………………...
(2) pierre.de-mey@legos.obs-mip.fr ………………………………..

Other researchers: Sarantis Sofianos (1), Nadia Ayoub (2), …………….……………
Charles-Emmanuel Testut (Mercator Ocean, Ramonville St. Agne, France) ……………………………………………………………

Project Title: Stochastic Coastal/Regional Uncertainty Modelling: sensitivity, consistency and potential contribution to CMEMS ensemble data assimilation ………………………………………………………

Would you accept support for 1 year only, if necessary?	YES ☒	NO ☐

Computer resources required for 2016-2018: (The project duration is limited to a maximum of 3 years, agreed at the beginning of the project.)	**2016**	**2017**	**2018**
High Performance Computing Facility (units)	2.000.000	4.000.000	500.000
Data storage capacity (total archive volume) (gigabytes)	4.000	8.000	1.000

An electronic copy of this form ***must be sent*** *via e-mail to:* *special_projects@ecmwf.int*

Electronic copy of the form sent on (please specify date):
…………………………………………

Continue overleaf

Principal Investigators: Vassilios D. Vervatis and Pierre De Mey …………….…………

Project Title: Stochastic Coastal/Regional Uncertainty Modelling: sensitivity, consistency and potential contribution to CMEMS ensemble data assimilation …………………………………………...…………

Extended abstract

It is expected that Special Projects requesting large amounts of computing resources (500,000 SBU or more) should provide a more detailed abstract/project description (3-5 pages) including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used. The Scientific Advisory Committee and the Technical Advisory Committee review the scientific and technical aspects of each Special Project application. The review process takes into account the resources available, the quality of the scientific and technical proposals, the use of ECMWF software and data infrastructure, and their relevance to ECMWF's objectives. - Descriptions of all accepted projects will be published on the ECMWF website.

1. Introduction

Ocean forecasting systems fuse information from dynamical models and observational networks through data assimilation in an attempt to provide the best ocean state estimate and forecast. In recent years considerable effort has been dedicated to reduce forecast errors based on model improvement, observational array design and evolution of data assimilation methods, within the context of ever increasing computing capacities. Regional/coastal ocean estimates are subject to errors by cause of dynamical simplifications, horizontal and vertical grid resolution, numerical artifacts, improper parameterization of physical processes and sensitivity to imperfect boundary conditions. These errors are generally poorly known and ad hoc stochastic methods can be used along with other techniques to assess predictability limits.

Stochastic modelling offers the possibility to estimate the ocean response to given perturbations on state variables and constitutes one of the pillars of Ensemble filters such as the Ensemble Kalman Filter (EnKF) (Evensen, 2003). Dynamical ensemble integration provides estimates and forecasts of flow-dependent background error covariances. In practice, due to limited resources a choice often has to be made between increasing the model resolution and adopting advanced error-predicting schemes, one at the expense of the other. Since expensive, high-resolution ocean models are often needed to represent realistically the mesoscale, sub-mesoscale and coastal dynamics, sub-optimal schemes based on the KF have been proposed to tackle the problem. The computational burden of the EnKF for short-term forecasting and reanalysis purposes is such, that only TOPAZ4 (Sakov et al., 2012) adopts the method in a large-scale data assimilation system, among the 7 Monitoring and Forecasting Centers (MFCs) in the MyOcean EU project (http://www.myocean.eu.org) and its successor the Copernicus Marine Environmental Monitoring Services (CMEMS; http://marine.copernicus.eu).

In the context of this proposal we aim to simulate ensembles in a regional/coastal configuration and to enrich covariances adopting ensemble-based error estimates. This probabilistic representation of errors can in turn be used as a proxy tracer in space and time to improve observational networks and data assimilation techniques. The following step will deal with the integration of ensemble covariances into a data assimilative system, and consequences thereof. Sensitivity experiments will be conducted to evaluate the relative performance of an ocean-biogeochemical system with different ensemble approaches and observational networks.

The Special Project (SP) resources will be used in a joint proposal, to be submitted by OPAM/UoA (http://www.oc.phys.uoa.gr) and LEGOS/CNRS (http://www.legos.obs-mip.fr), within the CMEMS Service Evolution open tender http://ted.europa.eu/udl?uri=TED:NOTICE: 3761092015:TEXT:EN:HTML&tabId=1&tabLang=en. If the CMEMS Service Evolution proposal is accepted for funding it will have duration of two years beginning winter 2016. Both institutions have been actively collaborating in the past for operational and other research activities, within several EU funded projects. This collaboration was established during the first stages of operational oceanography in Europe, within the FP6 ECOOP (European COastal sea OPerational observing and forecasting system) project and the FP7 MyOcean GMES program.

2. Scientific Objectives

The scientific objectives of the SP follow the major R&D priorities outlined in the CMEMS Service Evolution open tender. Our research will focus on the required developments for stochastic modelling of ocean physics and biogeochemistry, in the context of coastal/regional Ensemble Data Assimilation (EDA) forecasting systems.

More in details the SP goals are twofold. Initially, we wish to carry out sensitivity studies using stochastic modelling to guide our choices in EDA schemes via the correct description of uncertainties; the main objectives are:

- Representing regional/coastal uncertainties in a high-resolution Bay of Biscay configuration, using AR (Auto-Regressive) processes implementing an SPPT (Stochastic Parameterization of Perturbed Tendencies) scheme.
- Focusing mainly on surface/upper-ocean variables in the shelves, shelf break and in open ocean at time scales from inertial frequency to the advective time scale (~week).
- Estimating the relative contribution of physical and biogeochemical perturbations in upwelling regions and interfaces between (i) the open ocean and the shelf, (ii) the ocean and the atmosphere.
- Evaluating the multivariate nature of physical and biogeochemical variables to guide future ensemble modelling approaches for assimilation methods based on the Ensemble Kalman Filter (EnKF; Evensen, 2003) and its sub-optimal variants.

In addition, we aim at investigating the “consistency of ensembles” as follows:

- Assessment of ensembles against observations within the CMEMS infrastructure. The available networks will be inquired from the different Thematic Assembly Centers (TACs; e.g. SST, SLA, ocean colour, in situ data), in order to perform data-space innovation statistics and rank histograms.

- Consistency analysis of ensembles using a set of tools for data assimilation, developed within the SANGOMA (Stochastic Assimilation for the Next Generation Ocean Model Applications; http://www.data-assimilation.net) FP7 project. The principle is to compare ensemble to innovation 2nd order statistics, same as above, but in array-space (ArM toolbox for “stochastic array design and consistency analysis”; De Mey, pers. comm., 2015).

- Feedback from the consistency analysis will be used to re-assess the stochastic protocol generating model/observation uncertainty for sensitivity studies.

3. Methodology and Experimental Design

In this study, we aim to perform ensemble simulations with the NEMO community model (Nucleus for European Modelling of the Ocean; Madec, 2008; http://www.nemo-ocean.eu). A regional configuration has been set up targeting from coastal to open ocean applications. The model domain covers the Bay of Biscay and the western part of the English Channel, using a $1/36^{o}$ curvilinear Arakawa C-grid (BISCAY36). For a complete description of the numerical set-up, identical to the IBI-MFC within CMEMS (http://marine.copernicus.eu), the reader is referred to Maraldi et al. (2013). The NEMO ocean engine OPA, in its latest version 3.6, is coupled with the passive tracer package TOP2 including the biogeochemical model PISCES-v2 (Pelagic Interactions Scheme for

Carbon and Ecosystem Studies volume 2; Aumont et al., 2015). The meteorological fields are provided by the European Center for Medium-Range Weather Forecasts (ECMWF).

A generic implementation of transforming deterministic ocean models into probabilistic, has been recently implemented within NEMO, including several kinds of stochastic parameterizations to simulate unresolved processes, scales and diversity (Brankart et al., 2015). In this work, we will generate stochastic patterns to investigate uncertainties in the ocean circulation and ecosystem models. For this reason, we design sensitivity experiments to assess error regimes based on stochastic modelling of the atmospheric forcing, simplifications in the air-sea interaction and uncertainties in intrinsic 3D ocean-biogeochemical variables. Our stochastic implementation is based on AR processes in the context of an SPPT scheme for the ocean state and the unresolved biodiversity for the ecosystem model. Brankart et al., (2015) have carried out stochastic simulations with low-resolution global/regional configurations (i.e. ORCA2/NATL025), in which spatial correlations are not explicitly introduced in the ocean and the ecosystem models. In high-resolution regional/coastal configurations this is of vital importance and we have recently developed a generic way to introduce spatial correlations. This is done by solving an elliptic Gaussian equation (Vervatis, pers. comm., 2015) integrated within the NEMO stochastic modules. For the ecosystem model we may wish to develop an anamorphosis transformation of the Gaussian distribution to carry out lognormal perturbations.

The expected duration of the SP is three years and the experimental protocol, with a tentative timeline, is summarized in Table 1. In Year-1 of the project, we wish to perform medium-range to seasonal ensemble simulations investigating error regimes in different ocean-biogeochemical variables. Initially, we will carry out perturbations only for the ocean physics focusing mainly on surface/upper-ocean properties (Ens-1). To introduce ocean uncertainties due to erroneous atmospheric forcing we propose to perturb the wind, the air temperature (Tair) and the Sea Level Pressure (SLP) fields. Uncertainties due to simplifications in parameterized variables will be expressed by perturbing air-sea interaction components, such as the drag and turbulent heat fluxes coefficients, as well as the absorption coefficient of the Photosynthetically Active Radiation (k_{PAR}) influencing the penetrative solar radiation. Uncertainties due to tidal mixing in the shelf may be imposed by perturbing the bottom drag coefficient. Sensitivity experiments (computationally inexpensive) over a short period perturbing only one variable and taking the difference with the unperturbed run, may assess the relative contribution of each stochastic field (e.g. Tair perturbations impact mainly on SST, SLP perturbations impact mainly on SSH). In addition, comparisons should be made using as reference recent stochastic modelling studies in the region (Quattrocchi et al., 2014; Kourafalou et al., 2015; Vervatis et al., 2015).

During the same Year-1, a second ensemble is designed to investigate error regimes based solely on uncertainties in the ecosystem model (Ens-2). For this task, we will perturb the tracer concentrations focusing on variables with available observations and mainly on chlorophyll. In this way, we aim at increasing patchiness on the ensemble spread, where we intuitively argue that the unresolved biodiversity should look like. In Year-2, we wish to generate a third ensemble by perturbing physics and biology simultaneously, in order to augment the ensemble spread of the coupled system (Ens-3). The target of this step is to enrich ensemble covariances and increase our directions for ocean-biogeochemistry EDA. In parallel, during each ensemble run, several observational networks will be envisaged within the CMEMS TACs products (http://marine.copernicus.eu), in order to perform consistency analysis of the ensembles in data/array-space. A feedback from the consistency analysis may be used to optimize our choices for EDA experiments to be carried out in Year-2 of the project. For this task, we will couple the ocean-biogeochemical system with the operational assimilation platform SAM2 (Système d'Assimilation Mercator, in its 2nd release) integrated in Mercator Ocean forecasting systems (http://www.mercator-ocean.fr). The EDA experiments will be performed using a kernel based on the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007), which is currently under development (Testut, pers. comm., 2015).

4. Technical Requirements

In this section, the SP estimates for computational resources and storage capacity are given approximately, based on recent experiments performed by Vervatis et al. (2015) on ECMWF/CEP machine and the local cluster NAVITI at Mercator Ocean. Thus, the SP estimates should be revisited after performing sensitivity experiments with the proposed coupled configuration of NEMO3.6 and PISCES-v2 at ECMWF HPCF (High Performing Computing Facility).

Vervatis et al. (2015) performed a free run and a large ensemble of 102 members requiring at about 10.000 cores at the ECMWF/CEP HPCF. The experiments were carried out in the IBM c2a cluster comprised of super-nodes of POWER7 processors. The model is optimized for vector computers and parallelised by domain decomposition with MPI software. The optimal domain decomposition for the BISCAY36, in order to exclude land grid-points requires 96 processors. The version used in the experiments was NEMO2.3, including tides and using a relative expensive time-step; the stochastic patterns where calculated offline performing an EOF analysis of the wind components; the ensemble runs covered a period of seven months (the first month was used for the ensemble spin-up). The ensemble costed out approximately 2.000.000 SBU and the model daily output was at about 3 TB (including monthly restarts). For the ensemble consistency analysis against observations is feasible to download CMEMS products at daily frequency to match the model outputs, so as not having to oversample in time and calculate online innovations.

Recently, we have carried out sensitivity experiments with the proposed configuration (NEMO3.6, PISCES-v2 and stochastic modules) on a local cluster comprised of 16-nodes, 8-cores per node and 48 GB RAM, using Intel Xeon 3.20 GHz 2-CPU processors. The new version of NEMO3.6 is approximately twice fast compared with the old version NEMO2.3 for the specific BISCAY36 configuration. However, coupling the ocean model OPA with the passive tracer package TOP2 and the biogeochemical model PISCES-v2 is three times more expensive compared with a non coupled system and therefore, we should expect an increase of the cost at about 50%. Additional consideration should be given to the increase of computational time, at about 15%, calculating online AR stochastic patterns.

The output from the ecosystem model is significantly larger than the ocean model. There are 24 prognostic tracers and a careful selection of variables should be made in order to constrain the biogeochemical output. Our intention is to focus only on chlorophyll and the main nutrients, where we can have data (real time and climatology) to validate our simulations. On top of these calculations we should add storage requirements for the stochastic and biogeochemical restarts provided from the new stochastic modules and PISCES-v2. Nevertheless, a reasonable estimate is to double the output requirements compared with the non coupled system.

In Year-1, the cost is associated with the aforementioned free ensembles Ens-1-2: physics and biogeochemistry separately. The number of members and the period of the ensembles would be finalized after performing sensitivity runs on ECMWF HPCF, to assess computational and storage limits, outlined in the specific boxes of the SP form. A tentative experimental design for the BISCAY36 configuration would be at least ~40-50 members per ensemble (convergence statistics after Vervatis et al., 2015) for a period of ~6-7 months or 80-100 members per ensemble for 2-3 months.

In Year-2, the cost is associated with the third free ensemble Ens-3: physics and biogeochemistry simultaneously, as well as, the coupling of the ocean-biogeochemical model and the assimilation platform SAM2. The cost performing an LETKF experiment is twice the cost of a free ensemble, since SAM2 performs a sequential Increment Analysis Update (IAU) at each assimilation cycle, where the model is rewind applying a 4D model update. In order to constrain Year-2 requirements for computing time, we will make decisions during Year-1 and beginning of Year-2, based on the analysis of our free ensembles, so as to perform selective ocean-biogeochemistry EDA experiments.

In Year-3 a relative small amount of CPU time is requested to guide our choices for future EDA simulations in regional/coastal forecasting systems.

Finally, since this is a joint SP it will be reasonable for the collaborating scientific groups to have common permission/access on the ensemble data stored at ECMWF premises.

References

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., Gehlen, M., 2015. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. Geosci. Model Dev. 8, 2465-2513.

Brankart, J.-M., Candille, G., Garnier, F., Calone, C., Melet, A., Bouttier, P.-A., Brasseur, P., Verron, J., 2015. A generic approach to explicit simulation of uncertainty in the NEMO ocean model. Geosci. Model Dev. 8, 1285-1297.

Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343-367.

Hunt, B.R., Kostelich, E.J., Szunyogh, I., 2007. Efficient Data Assimilation for Spatiotemporal Chaos: A Local Ensemble Transform Kalman Filter. Phys. D, 230, 112-126.

Kourafalou, V., De Mey, P., Staneva, J., Ayoub, N., Barth, A., Chao, Y., Cirano, M., Fiechter, J., Herzfeld, M., Kurapov, A., Moore, A.M., Oddo, P., Pullen, J., Van Der Westhuysen, A., Weisberg, R., 2015. Coastal Ocean Forecasting: science foundation and user benefits. J. Oper. Oceanogr., doi:10.1080/1755876X.2015.1022348.

Madec, G., 2008. NEMO ocean engine. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27 ISSN No 1288-1619.

Maraldi, C., Chanut, J., Levier, B., Reffray, G., Ayoub, N., De Mey, P., Reffray, G., Lyard, F., Cailleau, S., Drevillon, M., Fanjul, E.A., Sotillo, M.G., Marsaleix, P., and the Mercator Team, 2013. NEMO on the shelf: assessment of the Iberia-Biscay-Ireland configuration. Ocean Sci. 9, 745-771.

Quattrocchi, G., De Mey, P., Ayoub, N., Vervatis, V.D., Testut C.E., Reffray, G., Chanut, J., Drillet, Y., 2014. Characterisation of errors of a regional model of the Bay of Biscay in response to wind uncertainties: a first step toward a data assimilation system suitable for coastal sea domains. J. Oper. Oceanogr. 7 (2), 25-34.

Sakov, P., Counillon, F., Bertino, L., Lisaeter, K.A., Oke, P.R., Korablev, A., 2012. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Sci. 8, 633-656.

Vervatis, V., Testut, C.E., De Mey, P., Ayoub, N., Chanut, J., Quattrocchi, G., 2015. Data assimilative twin-experiment in a high-resolution Bay of Biscay configuration: 4D EnOI based on stochastic modelling of the wind forcing. Ocean Modell., in revision.

Table 1. Special Project (SP) timeline, resources and storage requirements.

SP pre-operating period: User registration and inexpensive experiments to (re)assess computational resources and storage		
Year	Requirements Quarterly	OPAM/UoA, LEGOS/CNRS and “other researchers”

1	1 Resources: inexpensive experiments Storage: ~*O* a few GB	(a) NEMO3.6/PISCES-v2 coupling. Free runs to optimize domain decomposition and performance for the regional/coastal configuration. (b) Integration of stochastic modules and code development to include several stochastic variables. (c) Assessment of Ens-1 stochastic protocol via sensitivity experiments.
	2 Resources: 1Million SBU Storage: 2TB	(a) Inquire/download CMEMS observational products (Obs-1), relevant with Ens-1 (e.g. daily SST, SLA, in-situ data). (b) Build scripts to process Obs-1. Sensitivity experiments for the consistency analysis tools incorporating Obs-1. (c) Perform Ens-1 (physics).
	3 Resources: inexpensive experiments Storage: ~*O* a few GB	(a) Archive/disseminate Ens-1. Innovation/convergence statistics and rank histograms in observation-space Obs-1 (e.g. daily SST, SLA, in-situ data). (b) Code development of the ArM toolbox (observational network: physics). Consistency analysis of Ens-1/Obs-1 with the ArM toolbox in array-space. (c) Assessment of Ens-2 stochastic protocol via sensitivity experiments.
	4 Resources: 1Million SBU Storage: 2TB	(a) Inquire/download CMEMS observational products (Obs-2), relevant with Ens-2 (e.g. daily Globcolour CHL). (b) Build scripts to process Obs-2. Sensitivity experiments for the consistency analysis tools incorporating Obs-2. (c) Perform Ens-2 (BGC).
2	1 Resources: inexpensive experiments Storage: ~*O* a few GB	(a) Archive/disseminate Ens-2. Innovation/convergence statistics and rank histograms in observation-space Obs-2 (e.g. daily Globcolour CHL). (b) Code development of the ArM toolbox (observational network: biogeochemistry). Consistency analysis of Ens-2/Obs-2 with the ArM toolbox in array-space. (c) (Re)assessment/optimize via Ens-1-2 experiments the stochastic protocol of Ens-3. Perform Ens-3 sensitivity experiments.
	2 Resources: 1Million SBU Storage: 2TB	(a) Reorganize/merge directories for Obs-1-2 daily datasets SST, SLA, in-situ and Globcolour CHL into Obs-3 and build scripts to (re)process altogether. Sensitivity experiments for the consistency analysis tools incorporating Obs-3. (b) Perform Ens-3 (physics and BGC). (c) Archive/disseminate Ens-3. Innovation/convergence statistics and rank histograms in observation-space Obs-3 (e.g. daily datasets SST,

		SLA, in-situ and Globcolour CHL). (d) Consistency analysis of Ensemble-3/Obs-3 with the ArM toolbox in array-space. Synthesis of all consistency analysis experiments.
	3 Resources: 2Million SBU Storage: 4TB	(a) Explore the multivariate nature of physical-biogeochemical covariances with respect to free ensembles Ens-1-2-3. (b) NEMO3.6/PISCES-v2/SAM2 coupling and code development. Prepare observational networks for EDA/LETKF experiments. (c) (Re)assess stochastic protocols with respect to free ensembles Ens-1-2-3. Perform a few cycles/members sensitivity EDA/LETKF experiments. (d) Perform specific/selective medium range to seasonal EDA/LETKF experiments with a large number of members (≥ 40 members).
	4 Resources: 1Million SBU Storage: 2TB	(a) Archive/disseminate EDA/LETKF outputs. Innovation/convergence statistics and rank histograms in observation-space Obs-3 (e.g. daily datasets SST, SLA, in-situ and Globcolour CHL). (b) Consistency analysis of EDA/LETKF/Obs-3 with the ArM toolbox in array-space. (c) Overall analysis of EDA/LETKF results. As needed, additional specific/selective EDA/LETKF simulations.
3	1-4 Resources: 0.5 Million SBU Storage: 1TB	(a) Design/preparation of future EDA experiments for regional/coastal forecasting systems within CMEMS.