SPECIAL PROJECT PROGRESS REPORT

Progress Reports should be 2 to 10 pages in length, depending on importance of the project. All the following mandatory information needs to be provided.

Reporting year	2016 ……………………………………………………..……
Project Title:	Stochastic Coastal/Regional Uncertainty Modelling: sensitivity, consistency and potential contribution to CMEMS ensemble data assimilation ……………………
Computer Project Account:	SPGRVERV…………………………………………..……
Principal Investigator(s):	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) ……………………
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Sarantis Sofianos (1), Nadia Ayoub (2), Charles-Emmanuel Testut (Mercator Ocean, Ramonville St. Agne, France) …………………………………………
Start date of the project:	24 March, 2016………………..……………………..……
Expected end date:	31 December, 2018……………………………………..……

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)	n/a	n/a	2.000.000	~45.000
Data storage capacity	(Gbytes)	n/a	n/a	4.000	~150

Summary of project objectives

(10 lines max)

This study focuses on topics of ocean stochastic modelling in the context of coastal/regional Ensemble Data Assimilation (EDA). The SP objectives are twofold. Initially, we wish to carry out sensitivity experiments using stochastic modelling to guide our choices in EDA schemes via the correct description of uncertainties. 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 ECMWF-SP resources are used in a joint project within the CMEMS Service Evolution open tender under Lot 3: links with coastal environment.

Summary of problems encountered (if any)

(20 lines max)

Below we give a short list with problems encountered and solutions given:

1) The NEMO ocean platform was installed successfully in CCA and CCB clusters. An incompatibility was observed when XIOS library compiled with Cray compilers. ECMWF administrators have helped us to switch into Intel environment and successfully compile both XIOS and NEMO.

2) The project initiation coincided with the second phase of the Cray HPCF. During this period CCA and CCB nodes have been upgraded with new Intel processor technology. As a consequence, one cluster at a time was unavailable for a week.

3) The PI of the project participated in the webinar “HPCF Cray Phase 2”, organized by ECMWF on 12 of April, 2016. The webinar presented best practices on how to use the new architecture of the clusters. With the help of ECMWF administrators we were able to change the resources geometry of our jobs and fill properly the nodes.

4) The requested resources for the specific SP were based on an estimate from a previous study, implementing a different ensemble approach and no ecosystem model (Vervatis et al., 2016). For this, we are currently in a test phase reassessing the project’s resources. First results show that we have underestimated the resources needed to complete the project, as well as the disk space to store the ensemble data. A report is submitted, attached to this progress report, requesting additional resources.

Summary of results of the current year (from July of previous year to June of current year)

This section should comprise 1 to 8 pages and can be replaced by a short summary plus an existing scientific report on the project

The project started in March, 2016 and during this period we have implemented an ocean-biogeochemical coastal configuration for the Bay of Biscay, based on the NEMO platform with ensemble capabilities. The system description is presented in Table 1. Scalability tests indicated that the optimal domain decomposition, excluding land points, is 96 CPU cores (Fig. 1). In addition, we have performed short runs with debugging and optimization options, evaluating the restartability and reproducibility of the configuration on CCB cluster. Test simulations show that we can reproduce a free run performed in previous studies (Vervatis et al., 2016). First results from a one year free run, focusing on CHL concentrations, are discussed in the following paragraph. Currently, we are in a test phase optimizing the PBS options of our scripts (e.g. nodes, memory, walltime, wrappers etc) in order to perform ensemble sensitivity experiments later in the project.

Table 1. System description

Geographical coverage	Bay of Biscay (11-0.26°W and 41-51.5°N)
Grid and resolution	1/36° curvilinear Arakawa C-grid (global ORCA tripolar grid) & 50 geopotential levels with partial steps (388x538x50)
Code version	NEMOv3.6-PISCESv2
Initial/boundary conditions	PSY2V4R2-BIOMER4V1R1 (July 06, 2011 to June 30, 2012)
Atmospheric forcing	0.25° ECMWF-3hr (CORE Bulk parameterization)
Tidal forcing	11 tidal harmonics: M2, S2, K2, N2, K1, O1, P1, Q1, M4, Mf, Mm
Rivers	Daily fluxes of three major rivers: Loire, Gironde and Adour

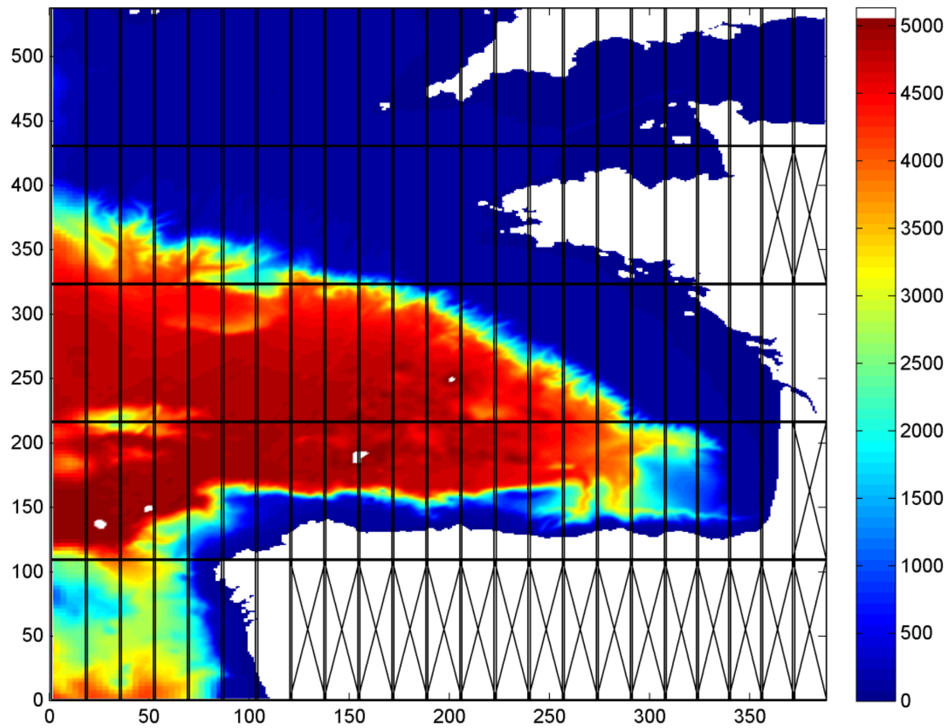

Fig. 1 Bay of Biscay configuration (i.e. BISCAY36 388x538 grid points) bathymetry and optimal domain decomposition of 96 CPU cores, excluding 19 land processors (i.e. 23×5 – 19 = 96).

Chlorophyll is a common parameter used in the validation of biogeochemical models due to wide availability of data. In Fig. 2, the satellite ocean colour dataset is used to quantify the model's Chlorophyll misfit against the observational network and its associated errors. Remotely sensed chlorophyll fields at daily frequency and 4km resolution were downloaded from the CMEMS site, with product identifier OCEANCOLOUR_GLO_CHL_L3_NRT_OBSERVATIONS_009_032. The spatial coverage of the data in this coastal region and in particular during cloudy periods drops in many cases below 20% (Fig. 2a). Enhanced levels of chlorophyll abundance are observed in the English Channel and on the southern part of the Irish shelf, during the whole period except in early-winter due to low temperatures and light availability in the region. In spring bloom elevated levels of chlorophyll are manifested, thought the model's variability is notably smaller compared with observations. Overall, the model performs reasonably well in the shelves, while underestimates chlorophyll concentrations offshore. Model's chlorophyll underprediction bias is distinct against the sparse satellite observations, though in general is smaller than the observational error (Fig. 2b).

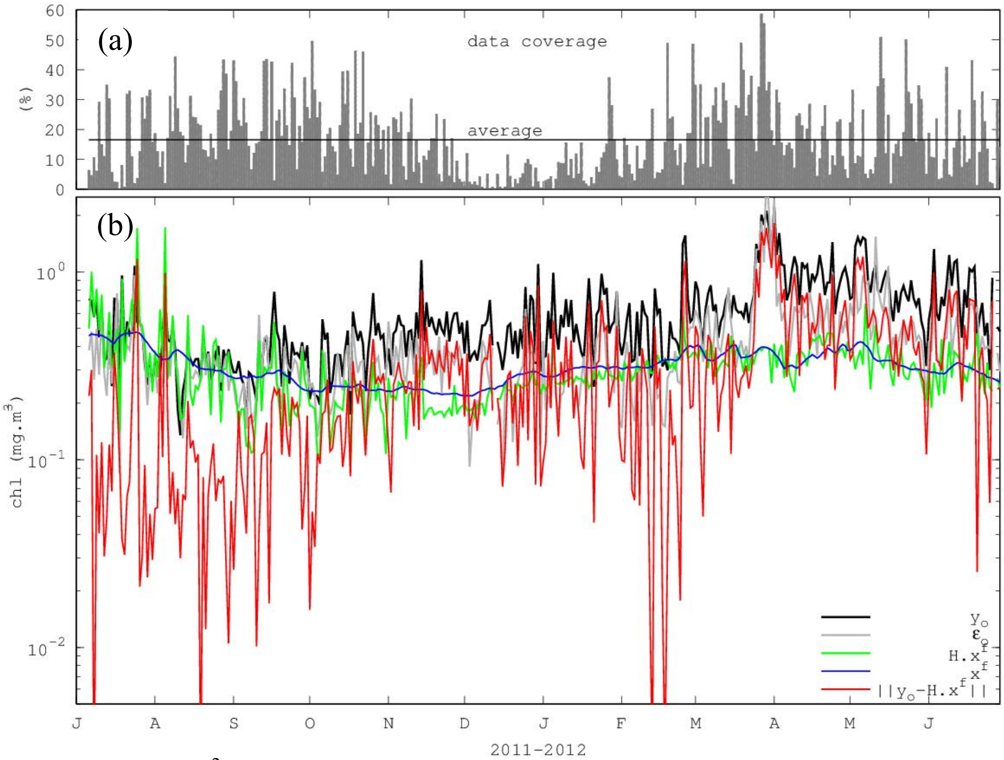

Fig. 2 Timeseries in mg/m^3 during one year free run: (a) satellite ocean colour observations Chl-a spatial coverage, (b) Chlorophyll of the model state x^f (blue), observations y_o (black), error of observations ε_o (gray), model state in observation space $H.x^f$ (green), misfit $||y_o - H.x^f||$ (red).

In this work, we implement a NEMO generic tool to perform ensembles. Currently, short test runs of stochastic simulations are carried out, with temporal behaviour modelled using autoregressive (AR) processes, in the context of a Stochastic Parametrization of Perturbed Tendencies (SPPT) scheme. This scheme is implemented today in meteorological weather forecasting systems at ECMWF (Palmer et al., 2009) and it has been recently introduced in the NEMO ocean model by Brankart et al. (2015). The AR scheme is complemented incorporating suitable techniques to open degrees of freedom of errors in high-resolution coastal configurations. For this, we have developed a method for variable/anisotropic correlation length fluctuations by solving an elliptic Gaussian equation (Vervatis, pers. comm., 2015).

We briefly present the variables proposed for the sensitivity experiments (to be performed till the end of the year 2016) and their statistical properties derived from a one year control run. The experiments are designed in order to investigate coastal uncertainties due to (a) erroneous atmospheric forcing i.e. wind, air temperature and sea level pressure (b) boundary parameterizations of unresolved processes in the ocean i.e. air-sea interaction (Cd, Ce, Ch, K_{PAR}) and bottom drag (Cd_b) coefficients and (c) parameterizations in various compartments of the ecosystem model. The medium range experiments, during April 2012 spring bloom, will guide our choices to optimize the perturbation approach for a seasonal range ensemble. For the latter experiment, we will impose fluctuations in all variables simultaneously for the atmospheric-oceanic system (Ens1), as well as the coupled ecosystem model (Ens2; to be examined later in the SP).

In Table 2, we summarize our options for the sensitivity experiments of all modelling systems. We analyze output diagnostics from a one year control run (i.e. deterministic free run) to calculate intraseasonal statistics (Fig. 3). We deduce pdfs of the variables of interest: all of them are normal except for the air temperature that shows a bimodal distribution. The wind noise to signal ratio is set to typical values reported in the literature, whereas air temperature and sea level pressure are considered less uncertain. In this study, we focus on synoptic timescales and we set the AR correlation period to a few days for all atmospheric variables, even for air temperature assuming uncertainties of synoptic phenomena i.e. storms. The energy cascade from large to small scales through the interaction of eddies, provides an estimate for the circulation spatial patterns in the Bay of Biscay (Fig. 3). The spatial atmospheric fluctuations are set approximately two times to half an order larger compared to the aforementioned ocean scales. The correlation timescale for the K_{PAR} coefficient is constrained for intramonthly AR processes, in order to enhance uncertainties during phytoplankton blooms. The bottom drag coefficient in many cases is approximated as a permanent feature in ocean models (e.g. fixed values in the Abyssal plain) and therefore, we impose large correlation timescales up to a month. In addition, considering limited knowledge for the dominant scales in the bottom layer, we introduce white noise and then a Laplacian filter (as in Brankart et al., 2015 for the ice strength parameter) to assess spatial correlations for the bottom drag coefficient. Finally, the ecosystem model variables are considered highly uncertain, focusing mainly on chlorophyll concentrations and in agreement with previous studies (Simon and Bertino, 2009; Brankart et al., 2015). All other stochastic parameters are closely related to the ocean modelling system and in particular to K_{PAR} applying lognormal perturbations.

Table 2. Scheduled ensemble sensitivity experiments and their stochastic parameters of first order AR processes for all modelling systems. An elliptic 2D Gaussian equation is applied to introduce spatial correlations with variable/anisotropic length ($1\text{-}\sigma_{x,y}$ in degrees) and uncertainty amplitude ($1\text{-}\sigma$ no units) of the 2D Gaussian distribution.

experiment	variable	uncertainty ($1\text{-}\sigma$)	correlation timescales	spatial scales ($1\text{-}\sigma_{x,y}$)	distribution
			Atmospheric model		
S1	Wind	0.3	3 days	1^o	Gaussian
S2	Tair	0.1	15 days	2^o	Gaussian
S3	SLP	0.01	5 days	3^o	Gaussian
			Ocean model		
S4	Cd, Ch, Ce	0.1	3 days	0.5^o	Gaussian
S5	K_{PAR}	0.2*	15 days	0.5^o	lognormal
S6	Cd_b	0.2	30 days	0.2^o	Laplace flt
S7	S1-6	medium range ensemble (April 2012 spring bloom; 20 mem)			
Ens1	S1-6	seasonal range ensemble (01 Dec, 2011 – 30 June, 2012; 40 mem)			
			Ecosystem model		
S8	all BGC	0.6*	10 days	0.5^o	lognormal
		medium range ensemble (April 2012 spring bloom; 20 mem)			
Ens2	S8	seasonal range ensemble (01 Dec, 2011 – 30 June, 2012; 40 mem)			

abbreviations: flt-filter; mem-members

*A lognormal anamorphosis function is applied for K_{PAR} ($\sigma_{norm} = 0.2$; $\sigma_{logn} = e^{0.2} \approx 1.2$) and BGC concentrations ($\sigma_{norm} = 0.6$; $\sigma_{logn} = e^{0.6} \approx 1.8$)

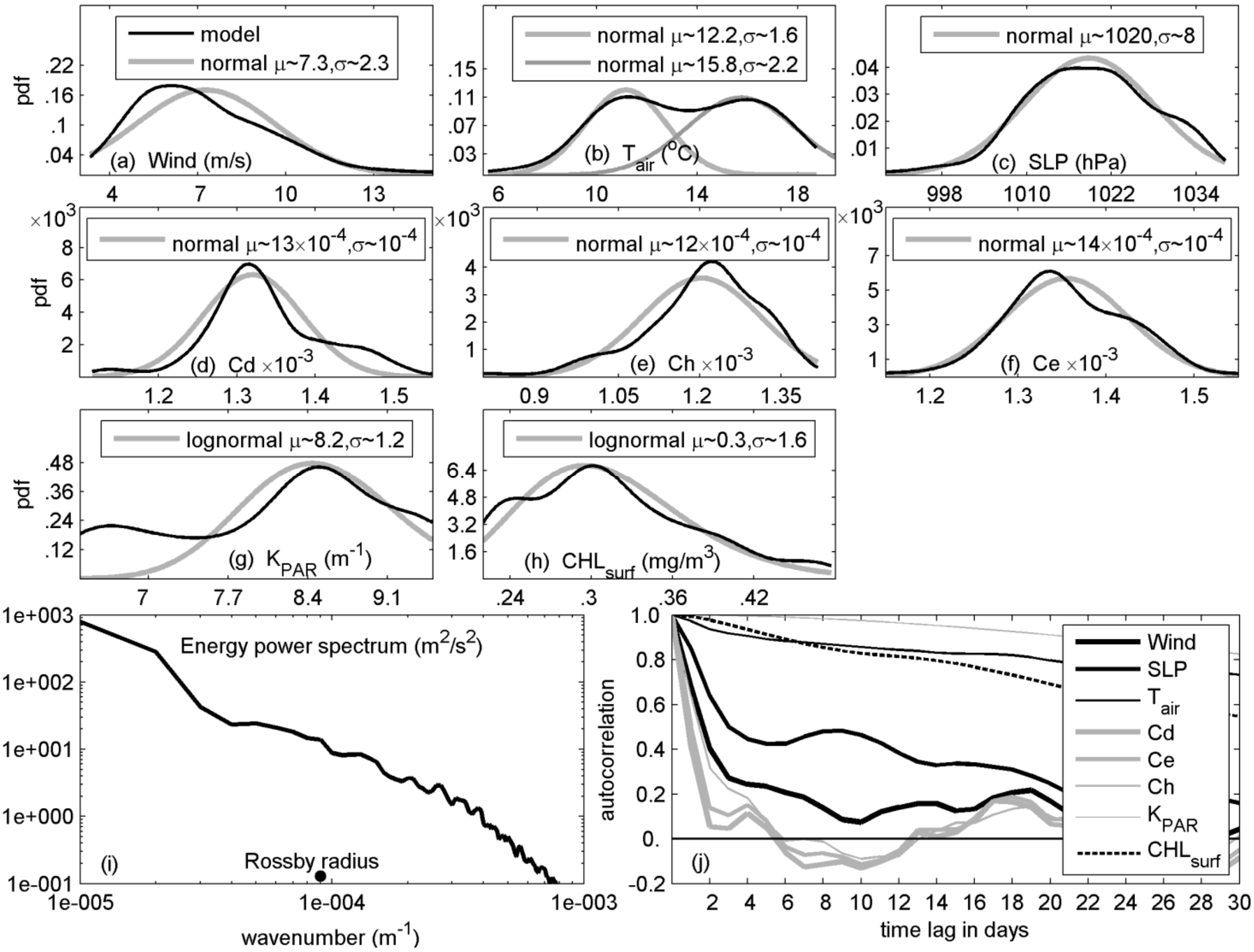

Fig. 3 (a-h) Spatiotemporal pdfs and statistical properties of the control run variables (black lines) and their fitting distributions (grey lines), (i) energy power spectrum (m^2/s^2) of ocean surface circulation per wavenumber (m^{-1}); superimposed the Abyssal plain first Rossby mode, (j) time lagged autocorrelations (lines: atmosphere-black; ocean-grey; chlorophyll-dashed). n/a: statistical properties for the bottom drag coefficient.

References

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.

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M., Weisheimer A., 2009. Stochastic parametrization and model uncertainty, ECMWF Tech. Memo. 598, 42 pp.

Simon, E., Bertino, L., 2009. Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. Ocean Sci. 5, 495-510.

Vervatis, V., C.E. Testut, P. De Mey, N. Ayoub, J. Chanut, and G. Quattrocchi, 2016: Data assimilative twin-experiment in a high-resolution Bay of Biscay configuration: 4D EnOI based on stochastic modelling of the wind forcing. Ocean Modell, 100, pp. 1-19.

List of publications/reports from the project with complete references

n/a ……………………………..…………………...………………………………………………………..

Summary of plans for the continuation of the project

(10 lines max)

This is the first year of the project. Currently, test runs are carried out to reassess the computational resources of the ensemble sensitivity experiments. Our plans are to continue the ECMWF-SP till 2018, since the resources are used in a joint CMEMS project for the same period.