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	2017
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)	2 + 7 = 9 MSBU	~ 9.5 MSBU	4 MSBU	~ 3.4 MSBU
Data storage capacity	(Gbytes)	4 + 11.5 = 15.5 TB	~ 12 TB	8 TB	~ 6 TB

Summary of project objectives
(10 lines max)

The ECMWF-SP resources are used in a joint project within the CMEMS Service Evolution open tender under Lot 3: links with coastal environment. The project aims at strengthening CMEMS in the areas of coastal/regional ocean uncertainty modelling, ensemble consistency verification and ensemble data assimilation. The work is based on stochastic modelling of ocean physics and biogeochemistry in the Bay of Biscay, on an identical sub-grid configuration of the IBI-MFC system in its latest CMEMS operational version.

Summary of problems encountered (if any)
(20 lines max)

The problems encountered during the ensemble runs of the previous year were mainly linked to the memory consumption by XIOS and netcdf libraries. The issue was reported to NEMO and XIOS developers in order to investigate the source of the leak and suggest possible solutions. In addition, we have contacted the ECMWF administrators to help us solve the problem, which was triggering an Out Of Memory (OOM) event cancelling the job. The proposed workaround was to change the resources geometry of the batch jobs and continue with the scheduled ensemble experiments. This solution has solved the problem only partially, since now the OOM killer is not triggered, but another problem occurred currently under investigation by ECMWF administrators. The problem is that in some cases a batch job might not exit properly even though the simulation is successful, thus wasting unnecessary resources till is killed by the wall-time limit. This behaviour appears to be random. In order to solve this problem, in collaboration with ECMWF administrators, we are currently performing short test runs. Finally, because of this problem, we are submitting an “additional resources” report to continue with the rest of the experiments scheduled this year.

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

In this section, we give a brief overview of SCRUM based on the mid-term report of the joint CMEMS project.

The scientific objectives of this project cover a broad spectrum of CMEMS interdisciplinary components, focusing on the generation of ensembles and their consistency against observations. In summary, we outline the following key points:

- Estimate coastal/regional uncertainties in a high-resolution Bay of Biscay configuration, by generating ensembles using Auto-Regressive (AR) processes implementing a Stochastic Parameterization of Perturbed Tendencies (SPPT) scheme.
- Complementing the above method to calculate explicitly anisotropic/variable stochastic patterns in high-resolution models, by solving an elliptic Gaussian equation and anamorphosis functions, in place of spatial filtering techniques (e.g. Laplacian operator).
- Focus on those variables of particular interest for CMEMS applications, such as surface/upper ocean variables on the shelves, shelf break and in the deep ocean, at time scales from inertial frequency to the advective time scale (~week).
- Estimate the relative sensitivity to physical and biogeochemical perturbations in interfaces between (i) the open ocean and the shelf, (ii) the ocean and the atmosphere.
- Analyze the multivariate connections between physical and biogeochemical variables to guide future ensemble modelling strategies in coastal/regional environments, in support of assimilation methods such as the EnKF and sub-optimal variants.
- Assess ensembles against observations within the CMEMS infrastructure. Observations are downloaded from different TACs (e.g. SST, altimetry data, ocean colour and in situ data), in order to perform data space innovation statistics.
- Showcase the use of those ensemble-modelled uncertainties in a pilot Data Assimilation (DA) impact exercise.

Methodology for Ensemble-based sensitivity studies

The ocean model used in this project is the NEMOv3.6 community model (Madec, 2008) and the ecosystem model is based on PISCES-v2 (Aumont et al., 2015), as part of the NEMO modelling platform. The configuration includes the Bay of Biscay abyssal plain with typical depths of ~4500 m, the shelf break and a variety of wide and narrow expanses of continental shelf, with an increasing width towards the north opening to the English Channel and to the Celtic Sea (about 300 km at 47°N).

A generic tool to generate probabilistic output from deterministic ocean models has been implemented within NEMO, including several kinds of stochastic parameterizations based on AR processes in the context of an SPPT scheme (Brankart et al., 2015). For this project, we have complemented the method for high-resolution configurations, by solving an elliptic Gaussian equation in place of Laplacian filtering operators to explicitly introduce spatial correlations and anamorphosis transformation of anisotropic uncertainty patterns (not yet in the NEMO svn trunk; Vervatis et al., in prep.).

Development of the SCRUM Consistency Analysis Toolbox

An Ensemble-based Consistency Analysis Toolbox (nicknamed scrumcat) is being developed in SCRUM. The resulting toolbox is being used to document the statistical consistency of Ensembles vs. CMEMS data and it will be delivered as a FORTRAN code so that it can be used beyond this project. The produced Ensembles undergo verification with respect to CMEMS observations, and in particular with respect to model-data misfits (hereafter called "innovation" for simplicity). First, it is understood that the innovation spread is the result of prior uncertainties of both the model and observations; therefore that spread should be statistically consistent with the sum of prior model uncertainty estimates and observational uncertainty estimates (measurement and representativity errors). In order to perform consistency analysis of the Ensembles, we are choosing several observational networks within the

CMEMS TACs infrastructure (http://marine.copernicus.eu). Observational error estimates have also been sought via the TAC infrastructure, when available, and also through discussions within the community.

The scrumcat toolbox is built upon the Sequoia Data Assimilation Platform (SDAP; https://sourceforge.net/projects/sequoia-dap). During the last part of 2016, scrumcat has been developed as a SDAP application. It uses several generic SDAP v1.5 elements, such as:

- SDAP Makefile and configuration tool
- SDAP versioning with svn
- Model grid services, including e.g. interpolation and regional masks
- Data services, including e.g. observation operators and decimation
- Off-Line Analysis services (OLA)
- Time sequencer.

The scrumcat toolbox adds specific elements such as:

- An interface with NEMO 3.6: grid, bathymetry, variables, observation operators
- NetCDF I/O: CMEMS data input, Ensemble input, grid input, regional masks input
- An observational error covariance model.
- The scrumcat consistency analysis tools, namely:
    - Data-space consistency analysis tool
    - Rank histograms consistency analysis tool
    - Array-space consistency analysis tool.

The tools calculate first- and second-order statistics pertaining to the distribution of observational samples, Ensemble samples, and innovation (observation minus Ensemble – OmE) samples. In addition, consistency analysis performed by means of rank histograms (Talagrand diagrams) to see how data project onto Ensemble-derived classes. This can be used to infer biases as well as differing spreads between data and Ensemble – e.g. a too small Ensemble spread will translate into most data falling into "outer" classes of the RH. The ArM ("Array Modes") toolbox propose criteria to characterize (1) array performance at detecting forecast errors, (2) consistency between forecast Ensemble statistics and innovation statistics. The tools work in space/time and across variables. The observational errors can be correlated.

Preliminary results

In **Fig. 1**, we present daily averaged maps of model errors for all medium range experiments after one month spin-up (i.e. 30 April, 2012). Error regimes are reshaped by the wind forcing, supported by the fact that wind uncertainties are derived by short-term correlated stochastic patterns with large uncertainty amplitude unlike other forcing fields. Wind uncertainties are expected to have a large impact in terms of Sverdrup dynamics influencing both geostrophic and Ekman components. Imposing the same perturbation field for both u and v wind velocities (i.e. not changing the wind azimuth) result in similar uncertainties for the vorticity and Ekman pumping, which is a wanted condition for robust stochastic patterns further enhancing model errors. The rest of the perturbed variables locally augment the ensemble spread in filament-like patterns in the periphery of eddies, near river plumes and in the shelf slope due to energy trapping. Air temperature uncertainties have a significant impact on SST and CHL compared to other experiments excluding wind perturbations. Uncertainties in the wind drag coefficient have moderate impact on the wind stress and consequently on the spread of surface ocean-biogeochemical variables against those of wind forcing (an order of magnitude smaller, i.e. $O(10^{-1})$). The Inverse Barometer (IB) response is mostly isostatic and of limited impact on SSH on the shelf. The latter is justified by the fact that IB perturbations are equal or larger than the external Rossby radius and/or the IB pumping on the abyssal plain (Tai, 1993). Uncertainties in the bottom drag coefficient amplify error regimes along the shelf break and on the shelves, and especially on the macrotidal area of the English Channel dynamically controlled by strong tidal currents.

Biogeochemical uncertainties arise from inborn ecosystem model errors and from errors in the ocean state variables. All sensitivity experiments perturbing physics (**Fig. 1**) leave an imprint on CHL uncertainties, which is on several occasions significant against SSH or SST model errors. When perturbing only the ecosystem model, we implement an identical stochastic pattern across all variables and vertical levels in order to obtain robust error regimes. In all cases, ocean model errors are found to have larger impact on CHL spread than those of the ecosystem perturbations (**Fig. 2**). An uneven ensemble spread is apparent comparing "nano" and "diatoms" compartments of CHL, with larger uncertainties observed in nano-CHL. The latter is most likely explained from the additional requirements in nutrient supply of Si for the primary production of diatoms.

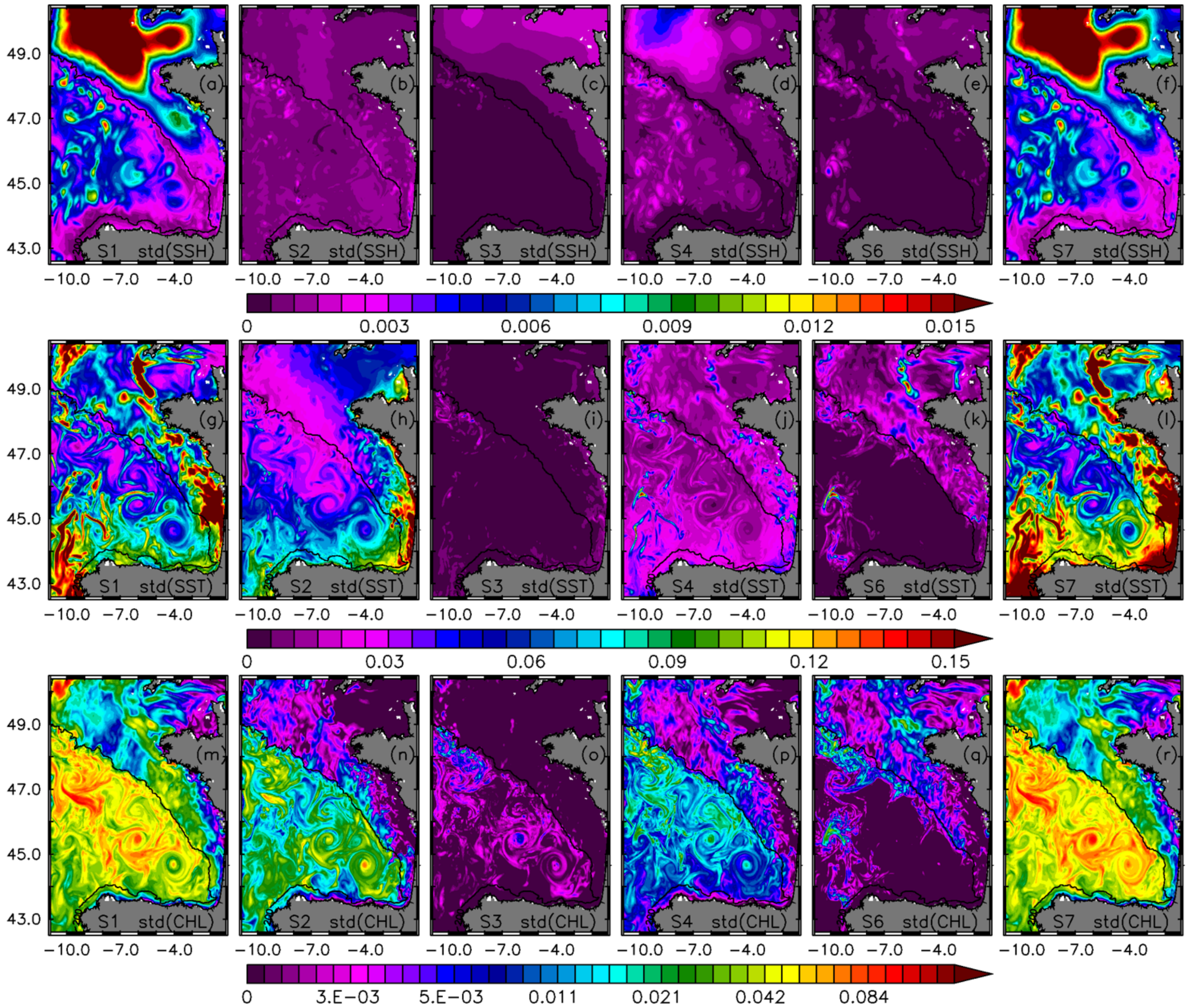

Fig. 1. Ensemble spread (i.e. 1σ) daily averaged maps on April 30, 2012 after one month spin-up: (a-f) SSH (m), (g-l) SST (°C) and (m-r) CHL surface concentration (mg/m^3). The colorbars are from top to bottom respectively; CHL concentration on logarithmic scale. Sensitivity experiments from left to right: wind, air temperature, SLP, wind drag coefficient, bottom drag coefficient, all perturbations.

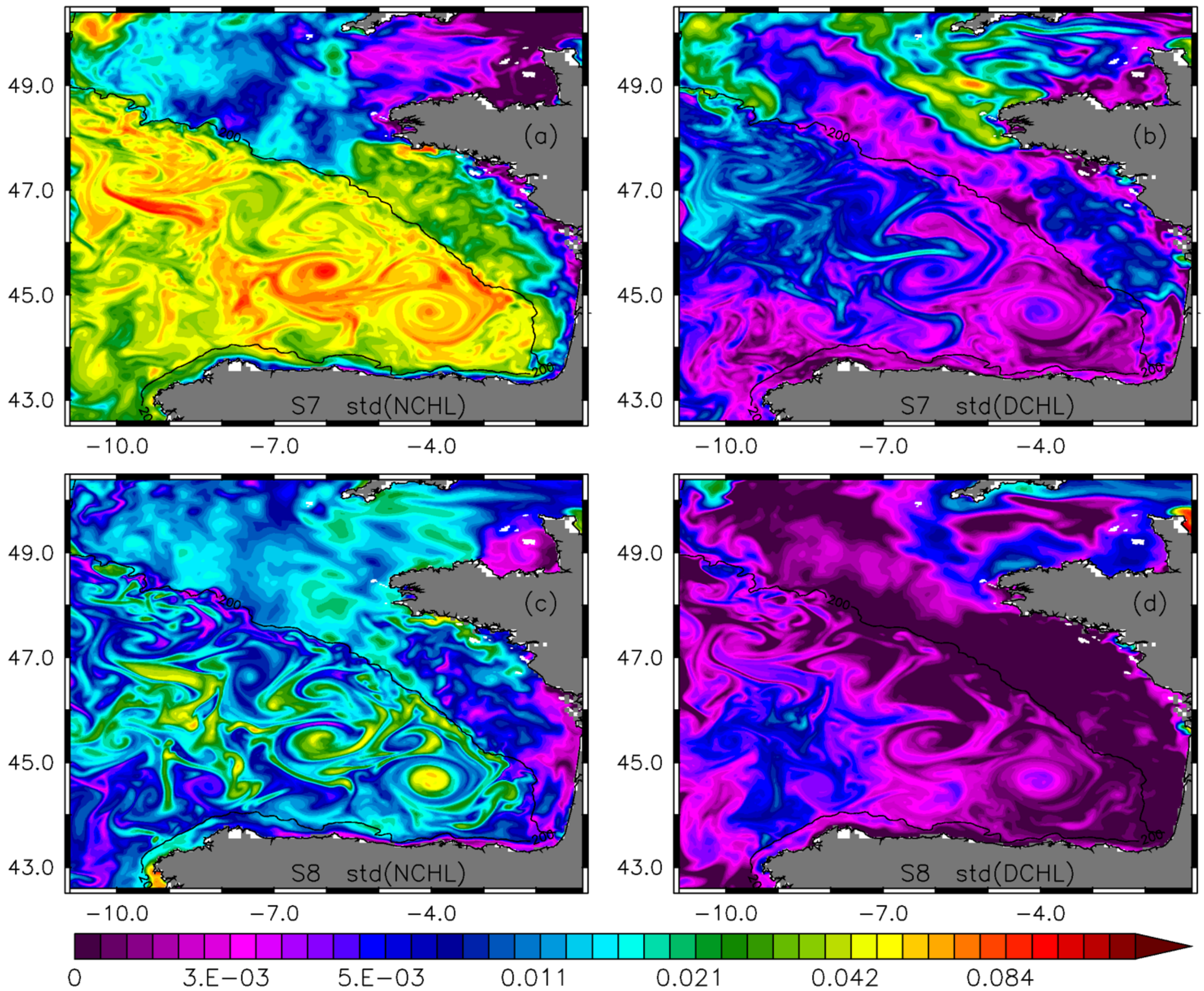

Fig. 2. Ensemble spread (i.e. 1σ) daily averaged maps on April 30, 2012 after one month spin-up: (a) nano and (b) diatoms CHL surface concentration (mg/m^3) perturbing only physics, (c-d) same as in (a-b) perturbing only Sources Minus Sinks (SMS) of biogeochemical state variables.

Fig. 3 shows an example of results with the data-space consistency analysis tool of the scrumcat toolbox. An ensemble of OSTIA (L4) foundation SST data, assuming an observational error st.dev. of 1.0°C, is compared with the ensemble perturbing only physics. In this example, the proxy for the foundation SST has been chosen to be the temperature at the level closest to 10-meter depth. The calculation is carried out for the whole months of January and June 2012. The left panels directly compare the ensemble spreads (first- and second-order Ensemble statistical moments, averaged over observations). Both ensembles appear to be compatible with each other since vicinities overlap. Assimilating those observations with those error estimates with e.g. an Ensemble or Bayesian filter would probably be well-posed, at least in data space, in the sense that the joint probability associated to both sources of information appears to be always nonzero. The right panel shows innovation statistic metrics. The model error estimate is always contained within the innovation spread, which is good. The mean bias between the center of both distributions stays within the [-0.5 0.5] °C interval. It can also be seen that the model spread slowly grows in time, since the perturbation mechanisms at work remain active throughout the period.

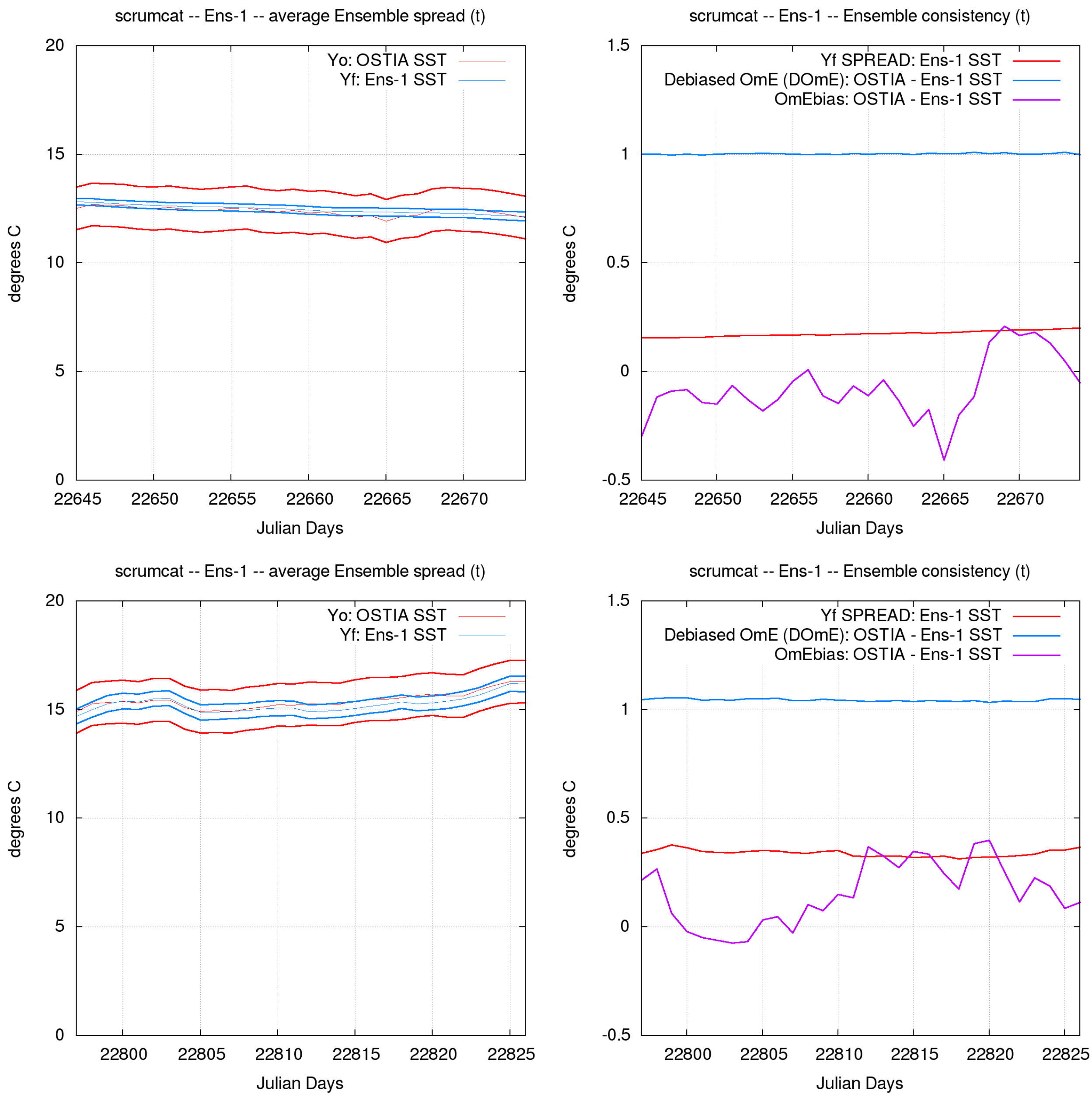

Fig. 3. Comparison of 40-member ensemble SST with OSTIA (L4) foundation SST pseudo-ensemble in (top) January 2012 and (bottom) June 2012 with the data-space consistency analysis tool: (left) Yo and Yf ensemble spreads as shown in legend (assuming observational error stdev=1.0°C); (right) ensemble consistency diagnostics as shown in legend.

<u>*Refernces*</u>

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.

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

Tai, C.K., 1993. On the Quasigeostrophic oceanic response to atmospheric pressure forcing: the Inverted Barometer pumping. NOAA Technical Memorandum NOS OES 005.

List of publications/reports from the project with complete references

Results using computational resources from this ECMWF-SP has been presented in the following workshop/conferences:

Vervatis, V., P. De Mey, N. Ayoub, M. Kailas and S. Sofianos (2017), Stochastic Coastal/Regional Uncertainty Modelling: a Copernicus marine research project in the framework of Service Evolution, Geophys. Res. Abstr., 19, EGU 2017-10054

Stochastic Coastal/Regional Uncertainty Modelling: a Copernicus marine research project in the framework of Service Evolution. Forced and Chaotic ocean variability: toward probabilistic oceanography, a workshop by invitation, Grenoble, FR, April 20-21, 2017 (https://chaoticocean17.sciencesconf.org)

Vervatis, V., P. De Mey, M. Kailas, N. Ayoub and S. Sofianos (2017), Stochastic Coastal/Regional Uncertainty Modelling: insights from ensemble sensitivity/consistency experiments, 5th GODAE OceanView COSS-TT-ICM5, Cape Town, South Africa, 3-7 April 2017

Vervatis, V., P. De Mey, S. Sofianos, N. Ayoub, M. Kailas and Mercator Ocean Team (2016-2017), Stochastic Coastal/Regional Uncertainty Modelling (SCRUM): Sensitivity, Consistency and potential contribution to CMEMS Ensemble Data Assimilation:

- CMEMS Service Evolution Mid-Term Meeting, Mercator Ocean, Toulouse, FR, January 23-27, 2017
- CMEMS Service Evolution Coordination Meeting, Bergen, Norway, December 1-2, 2016
- CMEMS Service Evolution Kick of Meeting, Mercator Ocean, Toulouse, FR, March 9, 2016

Summary of plans for the continuation of the project

(10 lines max)

This is the second year of the project and our plans are to continue, since the resources are used in a joint CMEMS project for the same period.