SPECIAL PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

Project Title:	Stochastic Coastal/Regional Uncertainty Modelling: sensitivity, consistency and potential contribution to CMEMS ensemble data assimilation
Computer Project Account:	spgrverv
Start Year - End Year :	2016 - 2018
Principal Investigator(s)	Vassilios D. Vervatis (1), Pierre De Mey-Frémaux (2)
Affiliation/Address:	(1) National Kapodistrian University of Athens (UoA) .. (2) Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)
Other Researchers (Name/Affiliation):	Sarantis Sofianos (1), Nadia Ayoub (2), Charles-Emmanuel Testut (Mercator Ocean, Ramonville St. Agne, France)

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 within the CMEMS Service Evolution open tender under Lot 3: links with coastal environment. The project aimed at strengthening CMEMS in the areas of coastal/regional ocean uncertainty modelling, ensemble consistency verification and ensemble data assimilation. The work was 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 you encountered any problems of a more technical nature, please describe them here.)

The problems encountered during the project were mainly linked to the ECMWF necessary updates in the default versions of the software packages, libraries and compiler environments. All issues were solved promptly after contacting the ECMWF admins.

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 a joint CMEMS research project. 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.)

Executive Summary

Ocean model uncertainties arise on a range of spatiotemporal scales from the formulation of forecast models themselves. In the past two decades, successful applications of advanced ocean data assimilation methods have led to an increased interest in their use. However, ocean Ensemble forecasting and data assimilation still encounter great challenges.

The project aimed at strengthening CMEMS in the areas of coastal/regional ocean uncertainty quantification, 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. One important paradigm of this work is to adopt a balanced approach between building Ensembles and testing their relevance.

We use stochastic modelling to generate Ensembles describing uncertainties in open ocean and on the shelf, mainly focusing on upper ocean processes. Our stochastic implementation is based on autoregressive processes in the context of the stochastic parameterization of perturbations of the atmosphere-ocean state in the ocean model and of the biogeochemical state in the ecosystem model. The method is complemented to account for spatial correlations and anamorphosis transformation of anisotropic uncertainty patterns, which is of vital importance in high-resolution coastal/regional configurations. Wind uncertainties are found to dominate all other atmosphere-ocean sources of model errors. The Ensemble spread (i.e. 1σ), after one-month spin up period focusing on upper ocean properties, is approximately 0.01 m for SSH and 0.15 °C for SST, though these values vary depending on season and cross shelf regions. In the context of a seasonal Ensemble, the aforementioned values reach close to 0.05 m for SSH and 0.5 °C for SST. Ecosystem model uncertainties emerging from perturbations in physics appear to be moderately larger than those perturbing the concentration of the biogeochemical compartments, resulting in total chlorophyll spread at about 0.01 mg.m^{-3}.

We introduce several methodologies aimed at checking the consistency of the above Ensembles with respect to TAC data and arrays. An Ensemble-based Consistency Analysis Toolbox was developed and used to document the statistical consistency of our Ensembles vs. CMEMS data. It will also be delivered to CMEMS groups so that it can be used beyond this project. All Ensembles were found to be underdispersive. Regarding Chl, we identified a statistical biogeochemical spin-up time of the order of 3 months. As a general rule, Ensembles were found to be most consistent in terms of SST, moderately consistent in terms of SLA, and often inconsistent in terms of Chl. We also found that in general Ens-3 (with perturbations in both physics and bgc) was the most consistent. For all variables, pattern consistency was found to be fairly good at the larger scales; it degraded at smaller scales for gridded SLA and for Chl. Some error processes seemed to be missing from the range of perturbations – in particular, some high-frequency error processes are currently unaccounted for on the shelves. We could not clearly attribute the missing processes to any particular error process with the tools at hand.

The consistency analysis also addressed the question of which routinely-available surface datasets were fit for validating Ensembles. In particular, the use of high-resolution SST products appears necessary when it comes to validating eddy-resolving Ensembles (such as 1/36° here). This is probably also the case for Chl, but is masked at this time by other deficiencies.

We calculated multivariate representers and EnKF analyses from our Ensembles. One objective was to access the impact of observations onto unobserved variables, such as other data types or subsurface variables. The correlation structures revealed general differences between the abyssal and coastal areas, as well as between variables. On the shelves (English Channel and South Armorican shelf), the filament-shaped structures for SST, SSS and Chl were likely linked to near-shore features such as river discharges (Loire river plume), mid-shelf thermal fronts and tidal fronts. The effect of spring bloom could be seen on the shelf, confirmed by the negative correlation between SST (e.g., heating up) and Chl (plankton depletion following a bloom). In the abyssal plain, the scales were often characteristic of the underlying quasigeostrophic mesoscale features, Chl included due to the mesoscale vertical velocity field. Most of the Chl correction arise from uncertainties in physics, but bgc model errors tend to enhance the effect of the physics. Logically, assimilating Chl seems to have a very measurable impact on physical variables.

Scientific Objectives

As in the open ocean, the most straightforward approach to coastal ocean forecasting and applications is based on two steps: deterministic, realistic numerical modelling and validation with respect to observations; and optional data assimilation, which in turn enables forecasting. However probabilistic approaches could provide an interesting alternative in the coastal ocean. There may not be enough observations in many coastal regions of the world to reliably estimate uncertainties of numerical models, to assimilate in those models, and to enable deterministic forecasts. Therefore, it can be expected that probabilistic approaches will widely be used in the future, in complement to the “deterministic” approach, for quantifying uncertainties in coastal products, and for providing probabilistic forecasts. In addition, for authorities and service companies involved in applications such as e.g. coastal flooding, fish stock management or surface drift predictions, probabilistic products have the potential to facilitate crisis-time decision-making as well as the longer term policies aimed at the mitigation of risks, with respect to using deterministic, best-estimate products alone. It is time to learn how to draw benefit from Ensembles, which are just sampling-based probabilistic products by nature.

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. One important paradigm of this work is to adopt a balanced approach between building Ensembles and testing their relevance. In this context, we outline the following key points:

- Quantify 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 introducing a Gaussian 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 (Evensen, 2003) 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.
- Demonstrate the consistency analysis of Ensembles in the space of “array modes”, using a set of tools developed within the FP7 project SANGOMA (http://www.data-assimilation.net).
- Feedback from the consistency analysis re-assessing the stochastic protocol and generating model uncertainties for sensitivity studies; this task could have a direct impact in the IBI-MFC system.
- Showcase the use of those Ensemble-modelled uncertainties in a Data Assimilation Impact (DAI) exercise, via representers and EnKF analyses.
- Contribute to the CMEMS team guidance in support of upcoming decisions regarding the evolution of coastal/regional DA schemes.

Main numerical codes used in the project

In this work, we used a regional configuration of the NEMO community ocean model (Nucleus for European Modelling of the Ocean; Madec, 2008; http://www.nemo-ocean.eu), BISCAY36. The model domain covers the Bay of Biscay and the western part of the English Channel, using a 1/36° curvilinear Arakawa C-grid. The NEMO ocean engine OPA was 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 were provided by ECMWF; the initial state and the components of the ocean-biogeochemical open boundary conditions will be inquired through the daily-weekly archives of the CMEMS infrastructure.

We used NEMO with the Sequoia Data Assimilation Platform (SDAP, https://sourceforge.net/projects/sequoia-dap), implementing localized, 4D Ensemble Kalman Filtering, developed at LEGOS. All statistical diagnostic tools described here were implemented within SDAP and interfaced with NEMO.

Ensemble production

We investigate whether Ensemble-based cross-covariances are enriched (with the possibility to increase DA performance), by perturbing both components of the coupled ocean-biogeochemical model. In this regard, an increasing complexity of experiments to augment chlorophyll spread is designed as follows: initially perturbing only physics (Ens-1), then only biogeochemistry (Ens-2) and finally both models simultaneously (Ens-3). In the coupled Ens-3 simulation, the evolution of the biogeochemical tracers is described by the advection-diffusion equation:

$$\partial_t C = \overbrace{\overbrace{-\underbrace{\nabla(u \cdot C)}_{\text{advection}} - \underbrace{K_h\, \nabla_h^2 C}_{\substack{\text{horizontal}\\ \text{diffusion}}} + \underbrace{\partial_z(K_z\ \partial_z C)}_{\substack{\text{vertical}\\ \text{diffusion}}}}^{\text{Ens1}} + \overbrace{\underbrace{SMS(C)}_{\text{biology}}}^{\text{Ens2}}}^{\text{Ens3}} \tag{1}$$

where on the right hand of Eq. (1) the first term represents the advective transport of tracers along isopycnals, the second and third terms the 3D parametrized diffusion processes and the last term denotes all biological processes affecting the concentration of tracer C including the Sources Minus Sinks (SMS), such as for example respiration and death in phytoplankton growth and decay.

Fig. 1 shows the surface Chl Ensemble spread for all three Ensembles calculated in this project. In general, the spread is largest in Ens-3 (where both physical and bgc perturbations were applied). This is due to the fact that bgc uncertainties result from errors in physics and in bgc, in variable proportions depending on the location: compare e.g. the English Channel to the mesoscale field of the South Armorican slope current.

In this work, we discuss preliminary results during the second year of the CMEMS marine project SCRUM. In specific, we focus on the consistency analysis and data assimilation impact experiments incorporating Oceancolour data and ecosystem model variables.

Consistency analysis

The ArM ("Array Modes") toolbox developed at LEGOS as part of the European SANGOMA project (https://sourceforge.net/p/sangoma/code/HEAD/tree/tools) 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.

The toolbox works in the space of Array Modes, which are in essence how the array views model error. They are the eigenmodes of the scaled Representer Matrix (although they are calculated in practice as the singular modes of the scaled matrix of Ensemble anomalies). They allow to implement pattern-dependent consistency analysis, and also scale-dependent consistency analysis since it is common that the dominant array modes capture the larger scales, with scales decreasing with increasing rank.

The ArM-CA consistency analysis tool projects the problem of statistical consistency between innovation and Ensemble statistics onto array-space along the basis of array modes, implements a systematic consistency criterion for each modal rank, and provides an overall consistency criterion involving a user-defined tolerance. The mathematical concepts are explained in Charria et al. (2016) and Lamouroux et al. (2016).

On Fig. 2, we show Hovmöller diagrams of variations in time of the ArM spectrum, including Chl innovation consistency check against Ens-1, Ens-2, and Ens-3, projected onto 39 array modes. The product ID is 009_093 (http://marine.copernicus.eu/). White areas depict inconsistent array modes. Values smaller than 1, i.e. dark blue, depict poorly observable array modes (below the observational noise). Representer matrix (RM) spectra exhibit strong time variations, with peaks possibly corresponding to "differential blooms" across the Ensembles, associated with specific spatial patterns. The spectra also show a fairly long statistical bgc spin-up time, of the order of 3 months. Once past the bgc spin-up, e.g. in March/April, between 5 and 10 array modes show a fair degree of pattern consistency.

For SST and SSH, consistency is verified for the larger-scale patterns (not shown). However, for chlorophyll, we saw that consistent patterns were not always large-scale, and that the scale of inconsistent patterns was not always so different from the scale of consistent ones. In Ens-2, the physics are not perturbed, so bgc uncertainties are passively advected around. The patterns are of course more complex, and probably more realistic, in Ens-3: it seems difficult to attribute those error patterns to specific physical or biogeochemical processes.

Data assimilation impact experiments

In this step, we carry out the EnKF analysis step with SDAP, using multiple observations together. For this, we have developed a large part of the interface between SDAP and NEMO-PISCES: what is still missing for full EnKF is the model correction post-analysis. Considering this, we are getting closer to the perspective of a coupled Ensemble-based DA-ocean/biogeochemical system.

In this section, we illustrate the multivariate impact of both Temperature and Chl variables in the state vector, by assimilating two observational networks simultaneously, namely the OSTIA-SST and the Ocean Colour Chl (CMEMS Product IDs: SST_GLO_SST_L4_NRT_OBSERVATIONS _010_001;OCEANCOLOUR_GLO_CHL_L4_REP_OBSERVATIONS_009_093). In addition, we investigate the convergence of covariances and its impact on the increment analysis, incorporating different ensembles (i.e. Ens1 vs. Ens3) and ensemble sizes (10 vs. 40 members).

In Figs. 3 and 4, we compare the correction fields based on Ensemble covariances from 40-member Ens1 and 40-member Ens3 respectively, assimilating both SST and Chl. The most significant impact is observed in the area of the English Channel (EC) for the surface Chl. In all other variables the changes in the analyses are moderate while being locally large, with the most significant change for the SSH being in the shelves and the EC. If we use fewer Ensemble members (e.g. 10 members; not shown), the correction patterns are less smooth because covariances are calculated from partially converged statistics. Finally, if we assimilate both observational networks SST and Chl, instead of only SST (not shown), the changes in the analyses are mainly observed on the shelves and the EC. The latter confirms the capability of Chl assimilation to increase the increment values, mostly at small scales, for all variables.

Figures

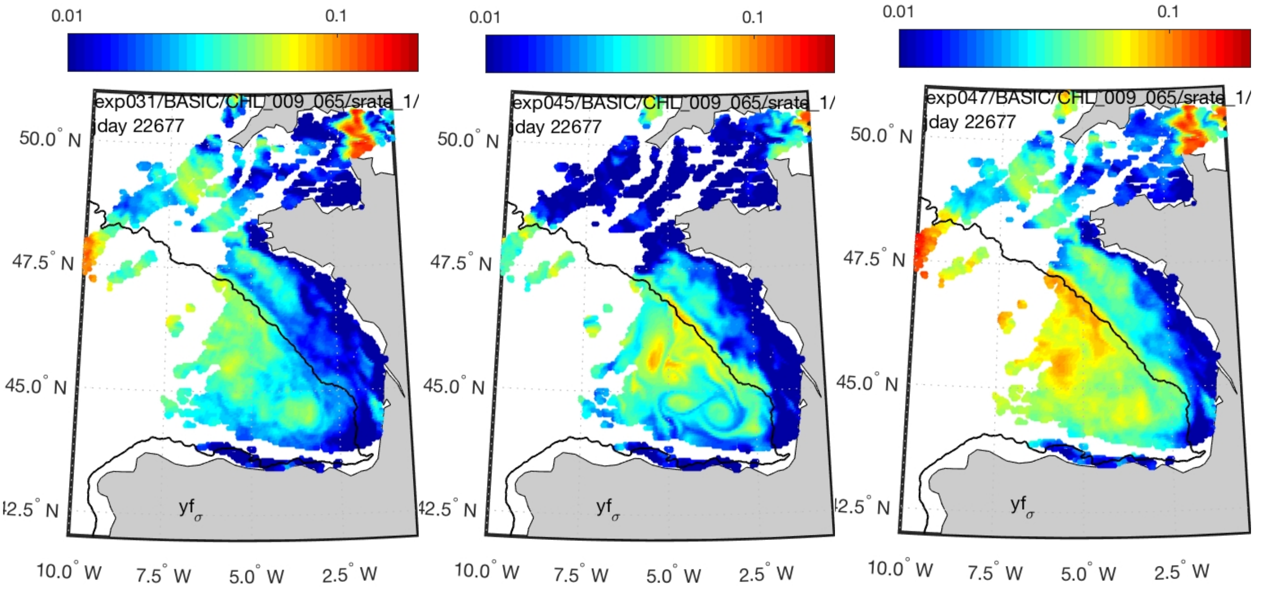

Fig. 1: Surface Chl 40-member ensemble spread (mg/m3) in data space (Product ID: 009_065). (left) Ens-1, (middle) Ens-2, (right) Ens-3. Date: 20120202.

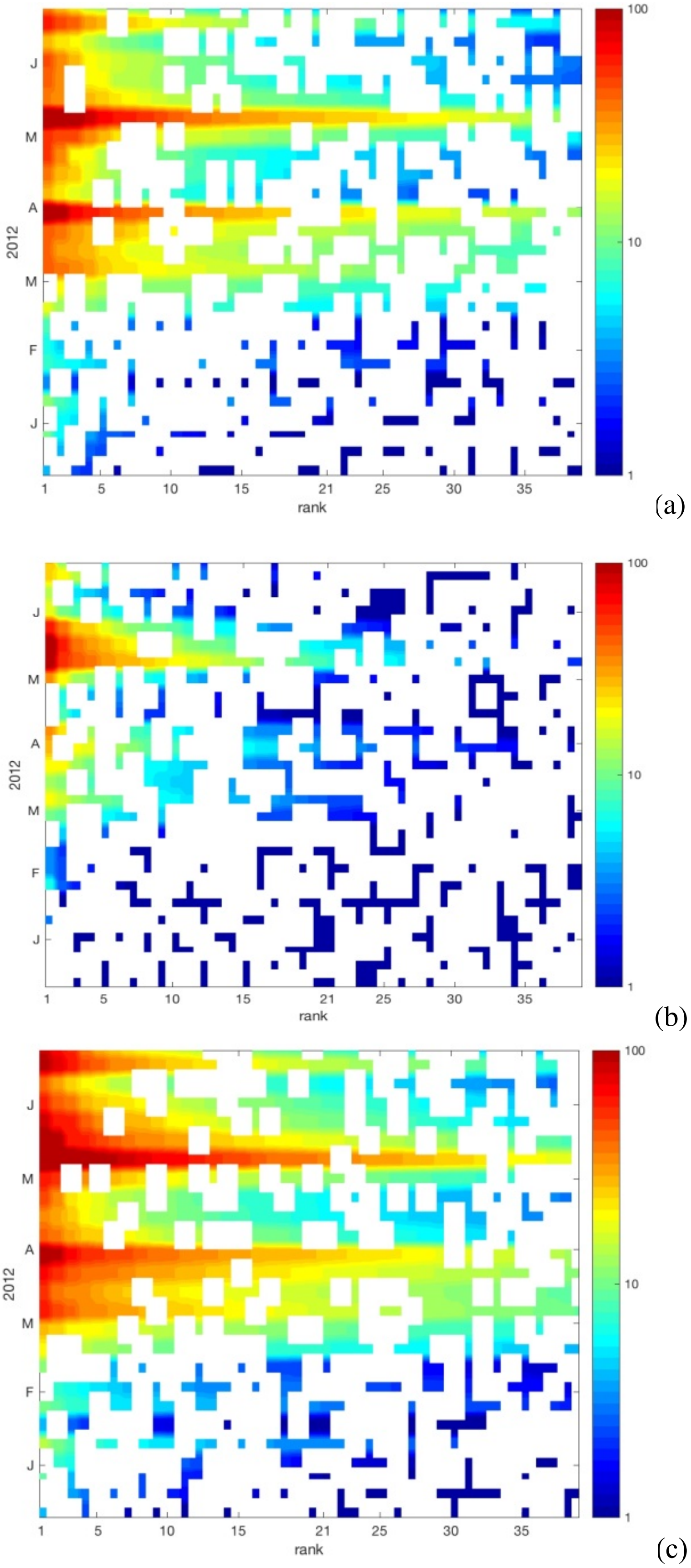

Fig. 2: (a) Hovmöller diagram of variations in time of the ArM spectrum, including a CHL gridded data innovation consistency check against Ens-1, projected onto 39 array modes. Product ID: 009_093. White areas depict inconsistent array modes. Colorbar: RM spectrum (values smaller than 1, i.e. dark blue, depict array modes below the observational noise), (b & c) same for Ensembles Ens-2 & Ens-3, respectively.

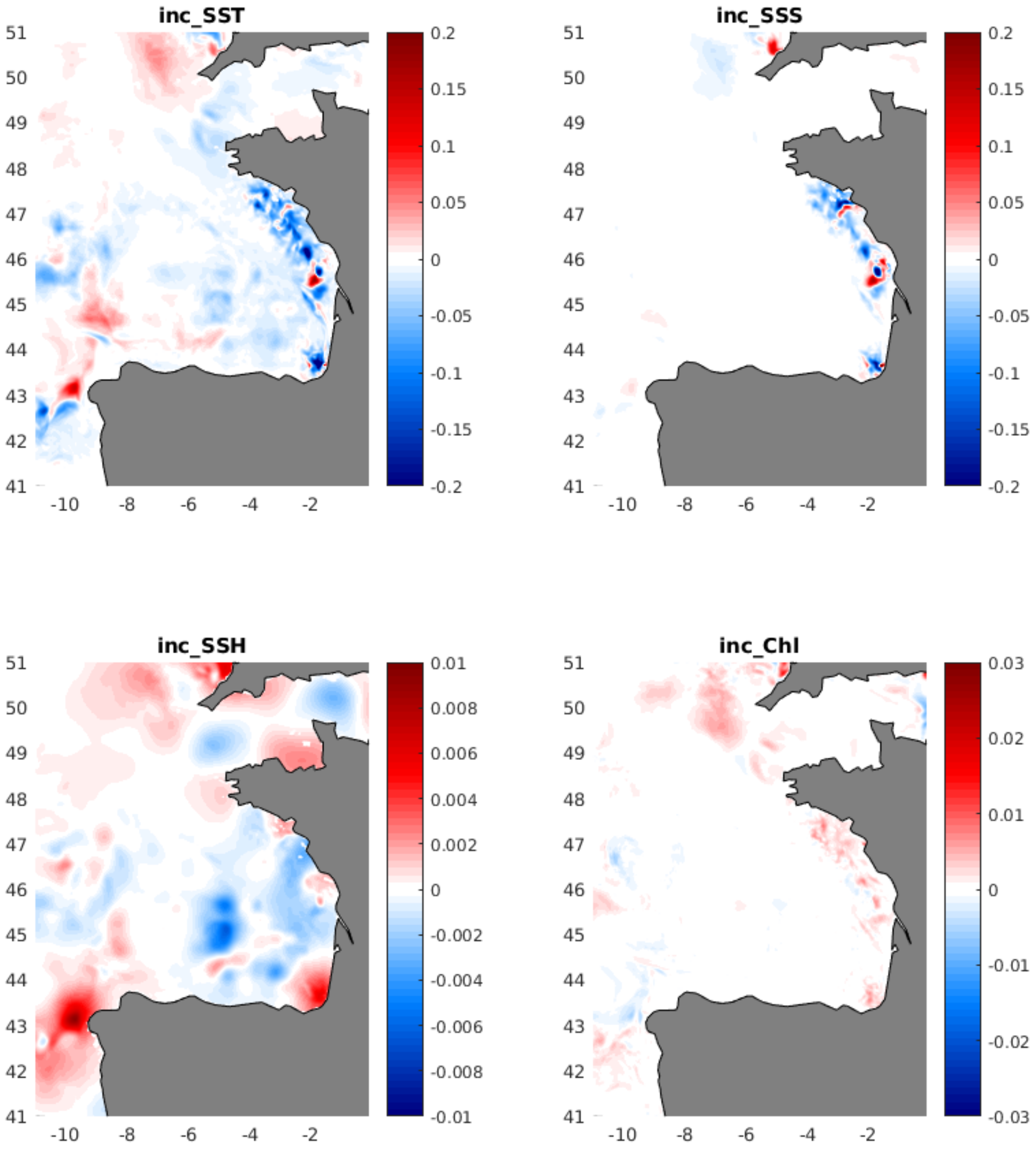

Fig. 3: Incremental analysis of the first member assimilating OSTIA-SST and Ocean Colour Chlorophyll on February 1, 2012. Ensemble covariances are calculated from Ens1 40 members. Correction fields on SST, SSH, SSS, Chl (cf. subplot titles "Inc_").

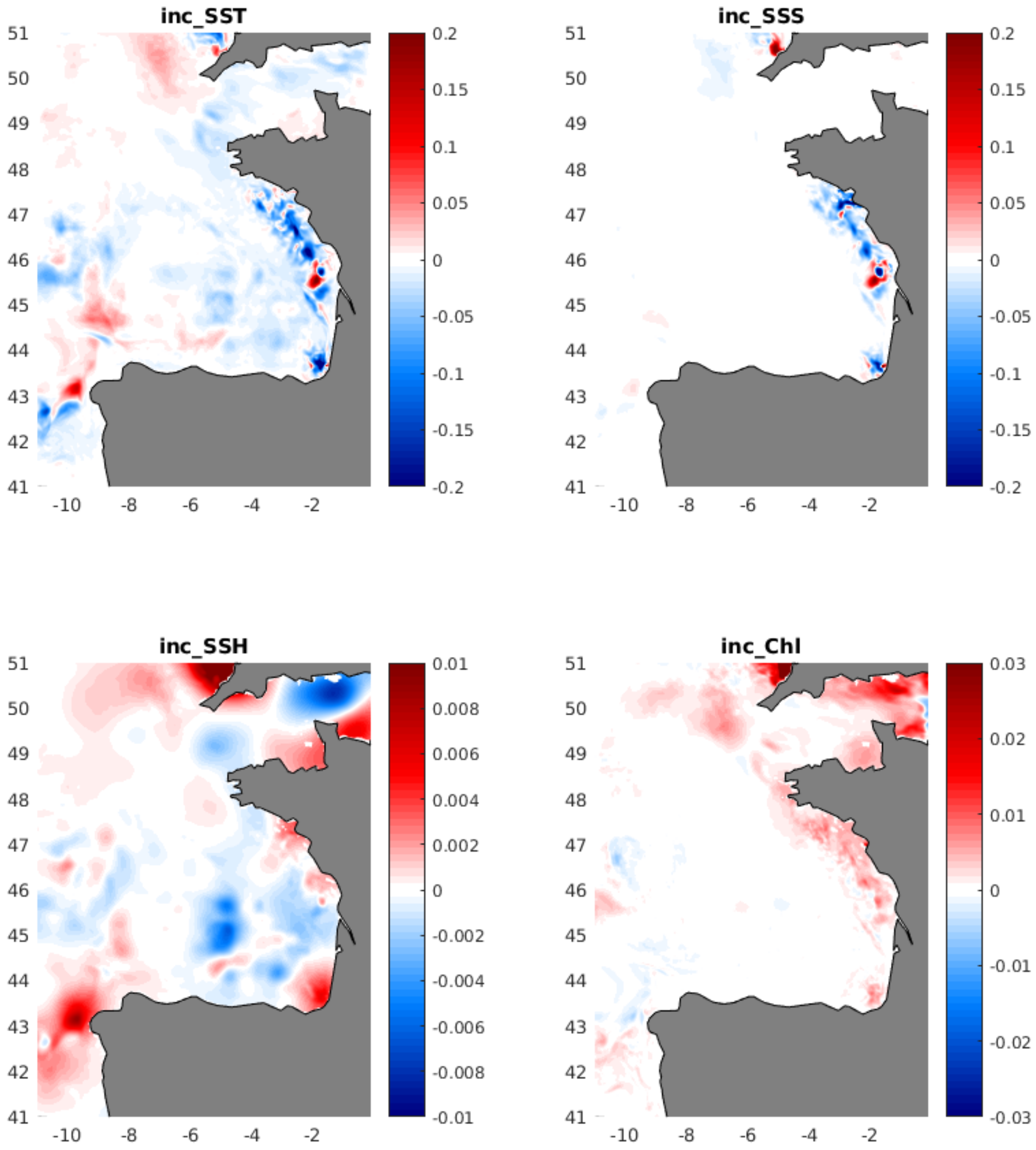

Fig. 4: Same as **Fig. 3**, for Ens3 40 members

References

Charria, G., Lamouroux, J. and P. De Mey, 2016: Optimizing observational networks combining gliders, moored buoys and FerryBox in the Bay of Biscay and English Channel. Journal of Marine Systems, http://dx.doi.org/10.1016/j.jmarsys.2016.04.003 .

Lamouroux, J., G. Charria, P. De Mey, S. Raynaud, C. Heyraud, P. Craneguy, F. Dumas and M. Le Hénaff, 2016: Objective assessment of the contribution of the RECOPESCA network to the monitoring of 3D coastal ocean variables in the Bay of Biscay and the English Channel. Ocean Dynam.,66(4):567-588, http://dx.doi.org/10.1007/s10236-016-0938-y .

List of publications/reports from the project with complete references

Conferences and colloquia

De Mey, P., G. Charria, J. Lamouroux, M.Le Hénaff, et M. Iskandarani, 2016 : Ensemble-based array performance assessment in coastal ocean models. SCRUM project kick-off meeting, Athens, 9-13 May 2016.

De Mey, P. et V. Vervatis, 2017: Estimation et validation ensembliste des incertitudes dans l'océan côtier. 2ème colloque de restitution TOSCA, Paris, 21-22 March 2017.

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

Vervatis, V., P. De Mey, M. Kailas, N. Ayoub and S. Sofianos (2017), Stochastic Coastal/Regional Uncertainty Modelling: a Copernicus marine research project in the framework of Service Evolution, Workshop in "Forced and Chaotic ocean variability: toward probabilistic oceanography", Grenoble, FR, April 20-21, 2017

Ghantous, M., N. Ayoub, V. Vervatis and P. De Mey (2017), Downscaling model errors in the Bay of Biscay, "Forced and Chaotic ocean variability: toward probabilistic oceanography", a workshop by invitation, Grenoble, FR, April 20-21, 2017

De Mey, P., V. Vervatis, 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

De Mey, P., V. Vervatis, N. Ayoub, M. Kailas, G. Charria, J. Lamouroux, Ch. Skandrani, M. Iskandarani, M. Le Hénaff, and S. Sofianos (2017), Array design and Ensemble consistency analysis, GODAE OceanView Joint DA-TT & OSEval-TT Meeting, CMRE, La Spezia, 11-13 October 2017

Vervatis, V., P. De Mey, M. Kailas, N. Ayoub and S. Sofianos (2018), Model uncertainties stemming from ocean-biogeochemical autoregressive processes: a high-resolution application for the Bay of Biscay, Abstract, Ocean Sciences Meeting, Portland, OR, 12-16 February 2018.

CMEMS meetings

Vervatis, V., P. De Mey, S. Sofianos, N. Ayoub and M. Kailas (2016-2018), Stochastic Coastal/Regional Uncertainty Modelling (SCRUM): Sensitivity, Consistency and potential contribution to CMEMS Ensemble Data Assimilation:

- CMEMS Service Evolution 2nd Coordination & Final-Term Meetings, Mercator Ocean, Toulouse, FR, February, 2018
- CMEMS Service Evolution R&D Copernicus Marine Week, Brussels, Belgium, September 25-29, 2017
- 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

Papers under review, in preparation

Vervatis, V. D., De Mey-Frémaux, P., Ayoub, N., Sofianos, S., Testut, C.-E., Kailas, M., Karagiorgos, J., and Ghantous, M.: Physical-biogeochemical regional ocean model uncertainties stemming from stochastic parameterizations and potential impact on data assimilation, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-31, in review, 2019.

Vervatis, V. D., P. De Mey et al., Which observations are fit for the validation of Ensembles? Answers from consistency analysis based on array modes, in prep.

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 are currently running the continuation of this Special Project for the years 2018-2020, named SCRUM2 (project account: spgrver2), linked also with a new awarded Copernicus marine project for the same period.