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	2019
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)	5MSBU	~4.5MSBU	8+5(+5)=13(+5)[1] MSBU	~7.7MSBU
Data storage capacity	(Gbytes)	7TB	~6.2TB	8+16=24TB	~8.5TB

[1] An additional amount of 5 MSBU is pending to be allocated, once resources are made available for the second half of 2019.

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 have reconsidered planned simulations for this year for scientific reasons on the basis of our joint CMEMS SE project, and therefore we have resubmitted an additional resources request. This request exceeded the maximum allowed additional resources per year, which is 10 MSBU. For the first half of the year 2019, a total amount of 5 MSBU have been added to our account. We sincerely hope that, when more resources are available, another 5 MSBU will be added for the second half of 2019.

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

We wish to continue the SP, which is linked to a joint Copernicus project for the years 2018-2020.

List of publications/reports from the project with complete references

Vervatis, V., P. De Mey-Frémaux, S. Sofianos, N. Ayoub, M. Kailas, Stochastic Coastal/Regional Uncertainty Modelling II (SCRUM2): Consistency, reliability, probabilistic forecasting and contribution to CMEMS ensemble data assimilation, CMEMS Service Evolution Kick of Meeting, Mercator Ocean, Toulouse, FR, April 3, 2018.

Vervatis, V., P. De Mey-Frémaux, N. Ayoub, S. Ciavatta, R. Brewin, J. Karagiorgos and S. Sofianos (2018), Ensemble consistency analysis in coastal ecosystem: the Bay of Biscay paradigm, GODAE OceanView, COSS-TT, ICM6, pp. 38.

Vervatis, D. V., P. De Mey-Frémaux, et al., Which observations are fit for the validation of regional ocean model ensembles? Answers based on array modes, in prep.

De Mey-Frémaux, P., N. Ayoub, A. Barth, B. Brewin, G. Charria, F. Campuzano, S. Ciavatta, M. Cirano, C. Edwards, I. Federico, S. Gao, I. Garcia Hermosa, M. Garcia Sotillo, H. Hewitt, L. Hole, J. Holt, R. King, V. Kourafalou, Y. Lu, B. Mourre, A. Pascual, J. Staneva, E. Stanev, H. Wang, and X. Zhu (in press). Model-observations synergy in the coastal ocean. In press, Frontiers in Marine Science, OceanObs'19 Community White paper.

De Mey-Frémaux, P., N. Ayoub, M. Ghantous, and F. Toublanc, 2018: Downscaling hydrodynamique ensembliste dans les modèles côtiers. Final meeting, AMICO project (French Environment Ministry / CNRS), Paris, 23 Nov. 2018.

Pierre De Mey-Frémaux participation in the Mercator Ocean International Data Assimilation (MOIDA) workshop, Toulouse, 5-6 Sept. 2018. Co-chairing of a session and participation to summary.

Pierre De Mey-Frémaux participation in the CMEMS STAC meeting, Toulouse, 18-19 Oct. 2018. Short presentation on R&D directions for circulation models and DA for the global and regional ocean, including references to ongoing SE projects.

De Mey-Frémaux, P., V. Vervatis, B. Lemieux-Dudon, J. Karagiorgos, N. Ayoub, and S. Sofianos, 2019 : Generation, Consistency and Reliability of Ensemble-based Regional Physical/Biogeochemical Uncertainty Estimates. 27th IUGG General Assembly, Montréal, Canada, July 8-18, 2019.

Summary of R&D activities and 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.

Objectives

This project chiefly aims at strengthening CMEMS in the areas of regional/coastal ocean uncertainty modelling, Ensemble consistency verification, Ensemble probabilistic forecasting, and Ensemble data assimilation. Specifically, the work is based on stochastic modelling of ocean physics and biogeochemistry, in the context of Ensemble Data Assimilation (EDA) forecasting systems, and includes methods suitable to assess the suitability of Ensembles for probabilistic forecasting.

As a first step, we produce short- to medium-range Ensembles over repeated periods to mirror the MFC operational practices. This permits to take into account the Ensemble spin-up period (up to 10 days) and gives access to performance categorization based on the age of errors within a given forecast lead time (up to 20 days).

In addition, we examine the SCRUM2 and the previous SCRUM (cf. technical reports of the previous project account SPGRVERV) Ensembles focusing on Ensemble consistency and probability scores (and associated statistical properties). Data from Thematic Assembly Centers (TAC) and arrays from the CMEMS catalog are used in this task, including Sentinel 3A data.

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 the ECMWF Ensemble Prediction System; 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 and probabilistic tools described here were implemented within SDAP and interfaced with NEMO.

R&D Status

Production of Ensembles: Following NEMO Ensemble capabilities with MPI double parallelization implementation, we are now able to perform stochastic approaches based on SPPT and SPUF methods, as well as incorporating atmospheric Ensembles (i.e. ECMWF-EPS). For this task, we have made extensive changes in the STO/ directory and in fldread.F90 module inside the NEMO trunk. Initialization techniques with account for the age of errors and forecast lead time, mirroring the error dynamics encountered in operational forecasting systems and hence relevant for those systems. The SCRUM2 novelties for Ensemble generation (in comparison to the previous project SCRUM) are:

- Performing Ensembles with the European Center for Medium-Range Weather Forecasts Ensemble Prediction System (ECMWF-EPS; 50 members). The goal using multivariate atmospheric Ensembles is to obtain a controlled (without DA) ocean Ensemble spread and focus on variables and timescales of particular interest to CMEMS.
- Revisiting the stochastic approaches investigated in SCRUM to optimize further the ensemble strategies for coastal/regional configurations.

Two perturbation approaches are followed in parallel: the stochastic implementation introduced in the previous project SCRUM and further developed here, and the ocean response to atmospheric Ensembles, i.e. using the ECMWF-EPS. For the stochastic approach in SCRUM2, we use the SPPT scheme with an increased uncertainty for the wind from 0.3 to 0.4 st.dev., and for the SMS(C) from 0.6 to 0.8 st.dev., in order to deliberately inflate the Ensemble spread. For the same reason, we use simultaneously the SPUF stochastic method sampling gradients from the T/S state vector performing random walks. In brief, the main simulations completed so far are briefly the following:

- A Control Run (hereinafter CR) was performed using the ECMWF-CTRL forcing (9km resolution). An upgrade compared to the SCRUM project is that we have used nutrient inputs from the atmosphere and in the river mouths to maintain a background of biological activity in periods of limited primary production.
- A seasonal-range Ensemble (hereinafter Ens4) of 50 members was carried out incorporating the ECMWF-EPS system (18km resolution), focusing on the period between December, 2016 to June, 2017. In order our approach to be general, we have integrated the ECMWF-EPS atmospheric forcing to be handled by the NEMO enhanced MPI capabilities for double parallelization in the spatial domain and the Ensemble dimensions, and in particular, we have made changes in the fldread.F90 NEMO module.

A summary of the Ensembles performed so far, and the ones currently generated and scheduled till the end of the project, is provided in the Table 1 below. In addition, we have included Ens-0 (Vervatis et al., 2016) and Ens-1 (previous project SCRUM; Vervatis et al., submitted 2019) as reference Ensembles.

Ensemble name	**Contents**
Ens-4	Seasonal-range Ensemble incorporating the atmospheric **ECMWF-EPS** system
Ens-5	Stochastic time-chunked Ensemble based on **SPPT** and **SPUF** methods incorporating the **ECMWF-CTRL** system
Ens-6	Combining the **Ens-4** and **Ens-5** Ensembles, incorporating the atmospheric **ECMWF-EPS** system, in the form of a time-chunked Ensemble
Ens-7	As **Ens-4** in the form of time-chunked Ensemble
S-	Several sensitivity tests (in the form of small Ensembles) that may include restart inflation techniques using the **ASM** modules of **NEMO**
	Reference Ensembles
Ens-0	Seasonal-range Ensemble performing **stochastic** modelling of the **wind forcing** using **EOFs** (Vervatis et al., 2016)
Ens-1	Seasonal-range Ensemble performing **stochastic** modelling based on the **SPPT** method, perturbing several variables (Vervatis et al., in review 2019)

Table 1: Ensemble generation in this project

Innovation statistics based on Sentinel-3A and in-situ products: A NetCDF I/O data loader has been developed to incorporate CMEMS altimetry data (along-track with gaps), as well as different in-situ platforms and observatories, streaming quality-verified data from CMEMS TACs.

Here are the main statistical and probabilistic diagnostics developed in the project:

- We assess the consistency of SCRUM/SCRUM2 Ensembles using an expanded set of diagnostics. New approaches take into account correlated observational errors in the consistency analysis. One of them aims at serving as a guide towards updating the R matrix.
- We develop and test probabilistic forecast scores based on Ensemble forecasts. This requires defining the occurrence of "events" based on "stake variables" as well as scale dependency. We also started to examine how the scores vary depending on the age of errors (mimicking forecast range dependency).

Account for correlated observational errors: Work has only started on this topic. We illustrate with a real case (HF radars; not shown) the impact of correlated observational errors on the definition of the basis of array modes over which the ArM-CA criterion is calculated. We have also implemented in SDAP an analysis tool to examine the error budget at the level of individual observations, evidencing misspecifications of R which are very common in coastal regions because of the presence of fronts.

Probabilistic forecast scores: This is a fairly new area in ocean sciences. Several authors in the meteorological literature (e.g. Murphy, 1973, 1993; Toth et al., 2003) have evidenced the two most important attributes of probabilistic forecasts:

- Reliability: this is the consistency between prior probabilities (in our case, from the Ensemble) and observed frequencies of the occurrence of "events", as a function of range (the "age of errors", which we can now attain with the new time-chunked stochastic scheme used in this project).
- Resolution: this is the ability to separate cases when an event occurs or not, as a function of range.

The most widely used probabilistic scores are the Brier score, DRPS and CRPS (Discrete and Continuous Ranked Probability Scores; e.g. Toth et al., 2003). Several probabilistic score decomposition algorithms are documented in the literature. These decomposition algorithms are all meant to measure two major EPS skills: Reliability and Resolution. However, none of these decompositions are fully equivalent. Importantly, Reliability and Resolution are sensitive to the chosen stratification of the ensemble probability bins2. Moreover, the Reliability (and the nature of the probabilistic event itself) is either defined with respect to fixed thresholds (eg, Brier and DRP from Candille et Talagrand, 2005) or rather measured with respect to the dispersion of the ensemble (eg, the Hersbach-Lalaurette CRPS decomposition from Hersbach, 2000). The Hersbach-Lalaurette CRPS decomposition defines a weaker condition for reliability compared to the DRPS from Candille and Talagrand, 2005. These differences are far to be anecdotal. When defining a probabilistic event of interest, they must be thoroughly considered at the light of the specific characteristics of the EPS (eg, ensemble spread, biases,...).

We worked on robust implementations of the Brier score and of the CRPS and validated them with a quasi-reliable testbed on Ens-1. We now trust our implementation of the scores even if potential

[2] Extra terms must be introduced in the Brier decomposition when the stratification of the ensemble probability bins is not chosen as the finest possible (Stephenson et al, 2007).

improvements are identified. We are now in position to examine the reliability and resolution properties with regionalization over SCRUM2 Ensembles (including time-chunked as a function of the age of "forecast" errors, mimicking the forecast range dependency). Short-term advances will include the definition of events and/or typical scales, accounting for underpopulated bins (algorithmic extension), and other points. We are interested in the verification of Ensemble-based probabilistic forecasts, typically issued for an interval of a variable or other type of event.

<u>Towards NEMO-SDAP coupling:</u> The structure of the SDAP code is being modified to permit NEMO to fully control the double MPI parallelization, including (1) domain decomposition, and (2) Ensemble operation. A demo based on the NEMO-GYRE configuration has been implemented to illustrate first steps towards a correction and restart strategy, using the NEMO double parallelization capabilities and the ASM/ modules inside the trunk.

<u>Preliminary results</u>

In Figs.1-2, we depict preliminary results for the time-chunked Ensemble Ens5 of 10 members over the same period with Ens4, for the upper-ocean variables SST and CHL in model space. The Ens5 SST spread is small during winter at about ~0.1 °C, though values are increased for the time-chunk Ensembles during spring at about ~0.5 °C. The latter result suggests that under the current stochastic protocol, we are able to obtain a large spread in medium-range Ensembles, with a forecast lead time of a few days/weeks. This is also true for CHL, even during winter. Those results are encouraging in the sense of having a large model spread, without having to perform long spin-up Ensemble simulations, as for example the seasonal-range Ensembles in SCRUM.

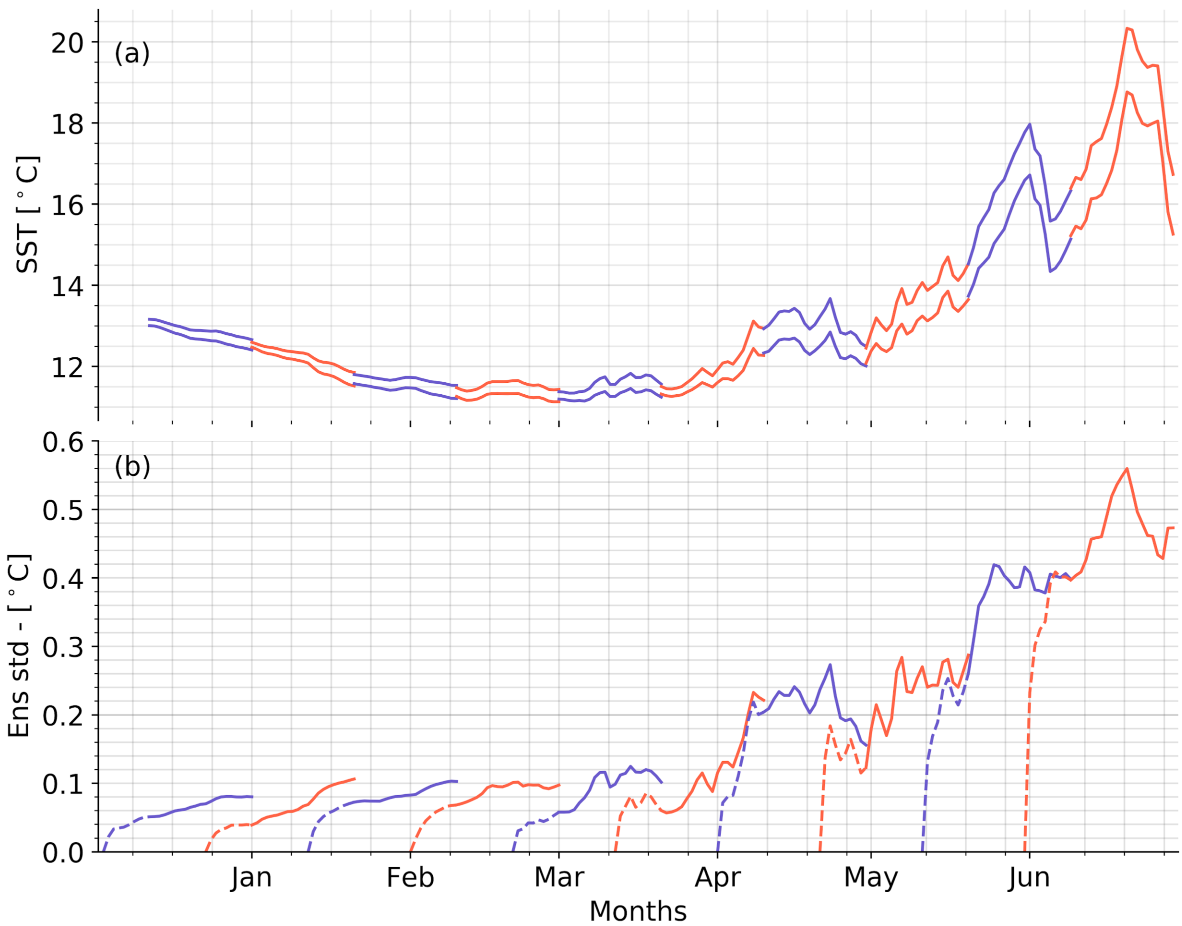

Fig.1 (a) Time-chunked stochastic Ensemble Ens-5 envelope for 10 members during Dec.2016 to Jun.2017 and (b) spread for the SST in model space, after a spin-up period of 10 days (overlapped in each time-chunk) and a Usable Period with forecast lead time of 20 days.

Figure 3 focuses on the Brier score implementation. It shows the evolution of the Reliability diagram over Dec. 2011 (top panels) and March 2012 (bottom panels) as well as the frequency of occurrence of probability pi. Over this longer time range, two statements are confirmed: i) given the identified small spread of ENS1, most of the bins are underpopulated, ii) despite these large sampling errors, the quasi-reliable testbed shows some reasonable reliability skills as revealed by the quasi-linear trend in the Reliability diagram.

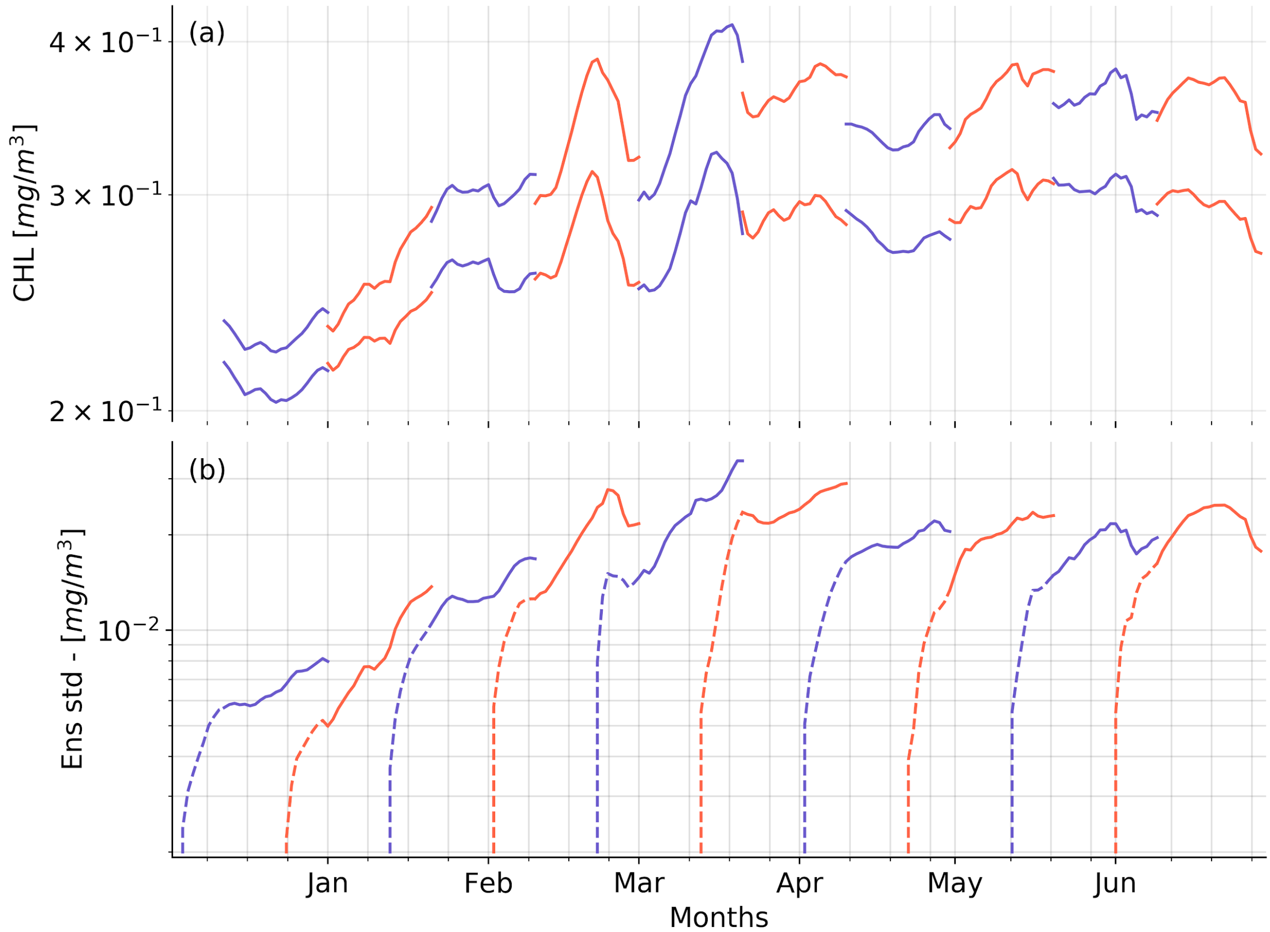

Fig.2 Same as Fig.1 for CHL.

Finally, the Figure 4 presents the Brier components, over Dec. 2011 (left panel) and March 2012 (right panel). Even though very preliminary, Figure 4 confirms that the Brier algorithm behaves as expected in the framework of a close to reliable system and for both months, the Reliability shows very small values, as expected for a reliable system. The potential Brier (ie, Brier for an ensemble system recalibrated to a fully reliable system) is therefore very close to the Brier itself. The potential Brier remains small compared to the Uncertainty: this is a desirable skill since it results in high Resolution values (this latter skill meaning that the conditional occurrence of the verifying observation when pi is issued is distinctly different from the climatological base rate and that we have an informative system). The changes in the Uncertainty during the month of December and to a lesser extent in March reflects the changes in the base rate, which in turns impact the Resolution (sum of the squared difference between the conditional frequency of occurrence of the verifying observation and the base rate). Indeed, the base rate and all the Brier components shown here are daily values (realizations are only sampled through space).

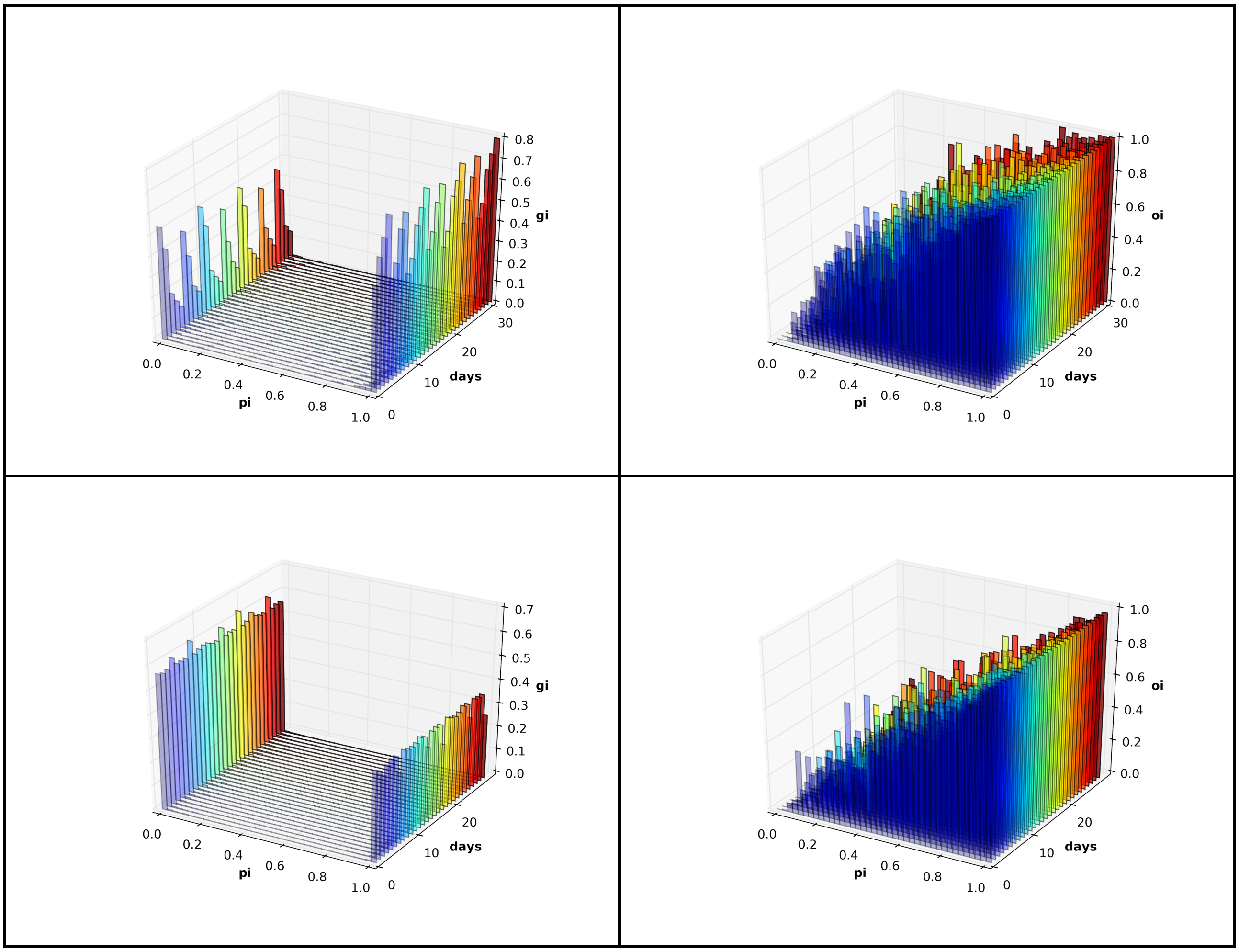

Fig. 3: Frequency of occurrence of ensemble probability (left) and Reliability diagram (right) from the Brier calculation in the framework of the synthetic verifying SST observation randomly drawn from a member of the ensemble (reliable EPS by construction) for Dec. 2011 (Top panels) and March 2012 (Bottom panels) with gi being the frequency of occurrence of the ensemble probability pi=i/N plotted against pi (Left panel)), and Reliability diagram with oi the conditional frequency of occurrence of the verifying observation when pi is issued plotted against pi (Right panel). Calculations use Ens-1.

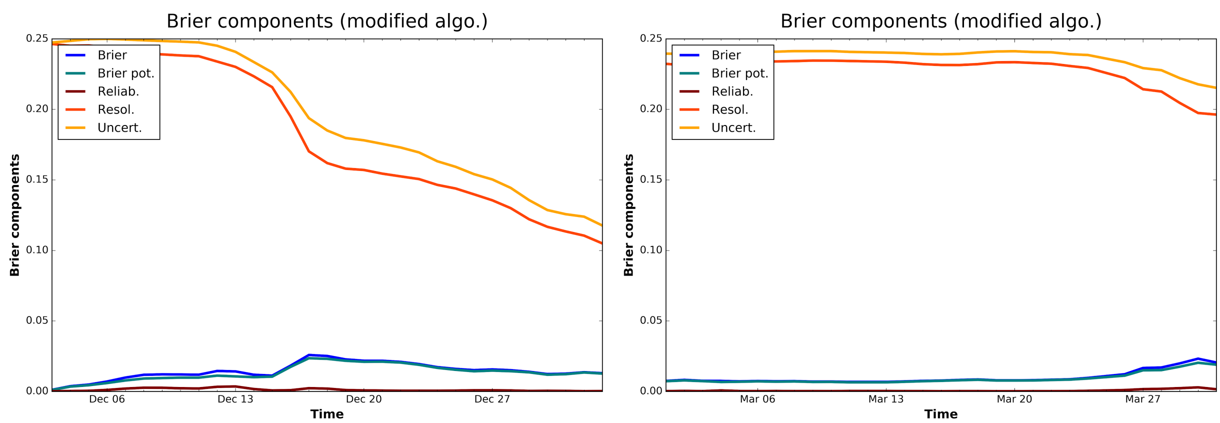

Fig. 4: Brier components in the framework of the synthetic verifying SST observation randomly drawn from a member of the ensemble for Dec. 2011 (left panel) and March 2012 (right panel). Calculations use Ens-1.

References

Candille G. and O. Talagrand. 2005. Evaluation of probabilistic prediction systems for a scalar variable. Quart. J. Roy. Meteor. Soc., 131, pp 2131–2150.

Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting, 15, pp 559–570.

Murphy, A.H., 1973: A new vector partition of the probability score. J. Appl. Met., 12, 534-537.

Murphy, A.H, 1993: What is a good forecast? An essay in the nature of goodness in weather forecasting. Weather and Forecasting, vol. 8, issue 2, pp. 281-293, https://doi.org/10.1175/1520-0434(1993)008%3C0281:WIAGFA%3E2.0.CO;2.

Toth, Z., O. Talagrand, G. Candille and Y. Zhu, 2003: Probability and Ensemble Forecasts. IN: I.T. Jolliffe and D.B. Stephenson, Forecast Verification: A Practitioner's Guide in Atmospheric Science, John Wiley and Sons, pp. 137-164.

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 Modelling, 100, 1-19, http://dx.doi.org/10.1016/j.ocemod.2016.01.003.

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