

Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation

Guillem Candille, Jean-Michel Brankart, Pierre Brasseur

▶ **To cite this version:**

Guillem Candille, Jean-Michel Brankart, Pierre Brasseur. Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation. Ocean Science, 2015, 11 (3), pp.425-438. 10.5194/os-11-425-2015 . insu-01203709

HAL Id: insu-01203709

https://insu.hal.science/insu-01203709

Submitted on 23 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ocean Sci., 11, 425–438, 2015
www.ocean-sci.net/11/425/2015/
doi:10.5194/os-11-425-2015
© Author(s) 2015. CC Attribution 3.0 License.

Ocean Science
Open Access

Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation

G. Candille, J.-M. Brankart, and P. Brasseur

CNRS, LGGE, 38041 Grenoble, France

Correspondence to: G. Candille (candille.guillem@hydro.qc.ca)

Received: 9 October 2014 – Published in Ocean Sci. Discuss.: 4 December 2014
Revised: 18 April 2015 – Accepted: 7 May 2015 – Published: 4 June 2015

Abstract. A realistic circulation model of the North Atlantic ocean at 0.25° resolution (NATL025 NEMO configuration) has been adapted to explicitly simulate model uncertainties. This is achieved by introducing stochastic perturbations in the equation of state to represent the effect of unresolved scales on the model dynamics. The main motivation for this work is to develop ensemble data assimilation methods, assimilating altimetric data from past missions Jason-1 and Envisat. The assimilation experiment is designed to provide a description of the uncertainty associated with the Gulf Stream circulation for years 2005/2006, focusing on frontal regions which are predominantly affected by unresolved dynamical scales. An ensemble based on such stochastic perturbations is first produced and evaluated using along-track altimetry observations. Then each ensemble member is updated by a square root algorithm based on the SEEK (singular evolutive extended Kalman) filter (Brasseur and Verron, 2006). These three elements – stochastic parameterization, ensemble simulation and 4-D observation operator – are then used together to perform a 4-D analysis of along-track altimetry over 10-day windows. Finally, the results of this experiment are objectively evaluated using the standard probabilistic approach developed for meteorological applications (Toth et al., 2003; Candille et al., 2007).

The results show that the free ensemble – before starting the assimilation process – correctly reproduces the statistical variability over the Gulf Stream area: the system is then pretty reliable but not informative (null probabilistic resolution). Updating the free ensemble with altimetric data leads to a better reliability with an information gain of around 30 % (for 10-day forecasts of the SSH variable). Diagnoses on fully independent data (i.e. data that are not assimilated, like temperature and salinity profiles) provide more contrasted results when the free and updated ensembles are compared.

1 Introduction

One of the challenges in ocean data assimilation is to faithfully describe the uncertainty of the ocean state estimates using observations and models. Another challenge is to deal with a wide range of spatial and temporal scales (Haines, 2010). On the one hand, the models cannot resolve all the small scales that may have a significant impact on the larger scales in the ocean circulation. On the other hand, this complex system is sparsely constrained – in space and time – by the observations network, e.g. the spatial observation only monitors the surface of the ocean. This situation results in a very partial description of the chaotic processes characterizing the ocean circulation. These conditions tend to invalidate the theoretical assumptions of linearity and Gaussianity (Nichols, 2010) on which classic data assimilation methods (Talagrand, 2010; Kalnay, 2010) are based on. To improve the efficiency of data assimilation in geophysics, ensemble-based methods have been developed (e.g. Evensen, 1994; Burgers et al., 1998; Evensen, 2003). Ensemble methods are designed to describe the evolution of the probability density function (pdf) of the ocean and thus provide a useful way to represent the uncertainties associated with complex systems. These uncertainties mainly come from the unresolved scales by the model, and from the interactions between the model and the external forcings (e.g. the atmospheric forcing). To

account for these uncertainties, the model cannot be considered as deterministic and must be transformed into a probabilistic model. To be more explicit, let us denote $\mathcal{A}$ the space of the scales resolved by the model and $\mathcal{B}$ the space of all the unresolved scales and all the interactions between the model and the external forcings. The probabilistic approach allows objective comparisons between the model (in $\mathcal{A}$) and the observations (in $\mathcal{A} \cup \mathcal{B}$) by providing sufficient conditions to invalidate the model. This approach also considers the model as a weak constraint to data assimilation problems by including the explicit description of model uncertainties.

Stochastic parameterizations are one of the most convenient ways to explicitly introduce the uncertainties in the models. This approach has been commonly used for 15 years in weather ensemble forecasting (Buizza et al., 1999) as well as in other atmospheric fields (Palmer et al., 2005). In oceanography, the models are still usually considered as deterministic and the stochastic/probabilistic approach is pretty new in this domain (Brusdal et al., 2003; Sakov et al., 2012; Kitsios et al., 2013; Porta Mana and Zanna, 2014; Yan et al., 2014). A previous study on a realistic model (Brankart, 2013) shows that the explicit stochastic parameterization of the unresolved scales in the computation of the equation of state has a clear impact on the larger dynamical circulation scales.

Following the results from Brankart (2013), the current study relies on the stochastic parameterization of the equation of state to perform an ensemble simulation and to design a data assimilation system based on altimetric observations. The main goal of this experiment is to provide a characterization of the uncertainty associated with the eddy dynamics observed in the Gulf Stream area (North Atlantic basin). The use of altimetric data for that purpose is motivated by the positive impact on the description of the oceanic circulation obtained by integrating altimetric data (like Jason-1 along-track data) into a deterministic assimilation process with a previous SEEK filter (e.g. Verron et al., 1999; Testut et al., 2003). However, the new feature here is the probabilistic aspect of the experiment, which is central to this paper. As already mentioned, considering an ensemble can be an efficient way to describe uncertainties related to the ocean circulation. But to extract useful information from this ensemble, validation must also be performed in a probabilistic way. For that purpose the reliability and resolution properties are introduced, following the approach developed in the meteorological community (e.g. Toth et al., 2003). This will enable objective comparisons between the ensemble simulation and observations.

Section 2 describes the realistic oceanic model configuration as well as the design of the stochastic perturbations used to build the ensemble. Preliminary qualitative diagnostics of the ensemble are also presented in this section. Section 3 introduces probabilistic validation concepts and probabilistic measures – also called scores. At this stage, first objective comparisons are performed between the model and the observations, and the uncertainty is quantified. Section 4 presents the ensemble 4-D assimilation scheme using the altimetric along-track data from Jason-1 and Envisat satellites, and shows probabilistic evaluation of the assimilation process. Finally, some concluding remarks and discussions are presented in Sect. 5.

2 Model configuration and ensemble simulation

2.1 Model configuration

For this study, we need a model able to reproduce the eddy dynamics in a realistic context that can be observed by the altimetric measurements. Moreover, the model computer cost must be tractable to enable ensemble experiments.

The realistic model used for this experiment is the North Atlantic DRAKKAR configuration of NEMO (Nucleus for European Modelling of the Ocean) version 3.4 (called NATL025, Barnier et al., 2006). This model has a free surface formulation and the prognostic variables are temperature, salinity, zonal and meridional velocities, and sea surface height. The model domain covers the North Atlantic Basin from 20° S to 80° N and from 98° W to 23° E. The horizontal resolution is 0.25°, which is considered as eddy-permitting in the mid-latitudes where the Rossby radius of deformation is about 50 km. Lateral mixing of momentum and tracers is modelled with a biharmonic operator. Vertical mixing is modelled by the TKE (turbulence kinetic energy) turbulence closure scheme, and convection is parameterized with enhanced diffusivity and viscosity. The forcing fluxes are calculated through bulk formulations, using the ERA40 atmospheric forcing fields (Uppala et al., 2005). Buffer zones are defined at the southern, northern and eastern boundaries (which are closed), with restoring to Levitus climatology (Levitus et al., 1998).

The presented study covers years 2005/2006 (when two altimetric satellites, in situ observations and external forcings are simultaneously available) and is mainly focused on the description of the Gulf Stream dynamics. For comparisons against observations and for assimilation process, the NEMO_OBS module (Bouttier et al., 2012; Lea et al., 2012) is used to compute the model equivalent at observation time and location. Actually, NEMO_OBS projects – by linear interpolations – the model outputs into the observation space at the exact observation time and location. The observations are the Argo profiles (temperature and salinity variables, hereafter denoted T/S) provided by the UK-MetOffice (Ingleby and Huddleston, 2007) and the along-track altimetric data from Jason-1 and Envisat satellites (produced by Ssalto/Duacs and distributed by AVISO, with support from CNES). In order to get comparable data between the SLA (sea level anomalies) observations provided by the satellites and the SSH (sea surface height) model output, NEMO_OBS removes the 7-year mean sea surface – averaged from 2002 to 2008 – computed using one integration of the stochastic ver-

sion of the model simulation (see next Sect. 2.2). In the following, SSH denotes both SLA observations and SSH model outputs which are both anomalies computed with their own mean sea level.

2.2 Model uncertainties

Before building a NATL025-based ensemble required for data assimilation, we have to wonder how to represent the uncertainties in the model and how to simulate the impacts of the unresolved small scales on the larger-scale circulation. Brankart, 2013, has shown that the unresolved scales in the nonlinear seawater equation of state represent a major source of uncertainties in the computation of the large-scale horizontal density gradient (from T/S large scale fields), and the impact of these uncertainties can be simulated by random processes representing unresolved T/S fluctuations.

Following these conclusions, the NATL025-based ensemble is built by introducing stochastic perturbations in the equation of state. In practice, a single nonperturbed integration is first performed from Levitus (1998) to 1 January 2005 to spin up the model state. Then a 96-member ensemble of perturbed simulations is run for 6 months with the following stochastic formulation of the equation of state:

$$\rho = \frac{1}{2}\Big\{\rho\big[T+\Delta T, S+\Delta S, p_o(z)\big] + \rho[T-\Delta T, S-\Delta S, p_o(z)]\Big\}, \quad (1)$$

where $p_o(z)$ is the reference pressure depending on the depth, and ΔT and ΔS are a set of T/S perturbations defined as the scalar product of the respective local T/S gradients with random walks ξ:

$$\Delta T = \xi \cdot \nabla T \qquad \text{and} \qquad \Delta S = \xi \cdot \nabla S, \quad (2)$$

ξ are produced by first-order autoregressive processes (AR-1) with a 10-day decorrelation timescale, and horizontal and vertical standard deviations σ_s equal to 1.4 and 0.7 grid points, respectively. ξ are uncorrelated over the horizontal and fully correlated along the vertical. These stochastic parameters are chosen to produce an ensemble spread that is large enough for our purpose while keeping the model numerically stable. Nevertheless, in order to avoid numerical instabilities, limiting factors are introduced (1.5 σ_s) on the perturbations and the time step of the stochastic model is divided by 4 compared to the time step of the classical model (600 s instead of 2400 s). Such an ensemble is thus built to spread mostly over areas with strong gradients and where the equation of state is strongly nonlinear, for instance in the Gulf Stream area. In practice, the ensemble simulation is performed on a massively parallel computer, which enables to produce a relatively large ensemble within a reasonable clock time. The size of the ensemble (96 members) is chosen in order to satisfy several factors. Without considering the numerical cost, the larger the size of the ensemble, the more accurate the descriptions of the pdfs and the covariance matrices. The ensemble size is then a compromise between the numerical constraints and the accuracy of the pdfs and covariance matrices associated with the ensemble. Moreover, we also have to take into account the saturation – depending on the ensemble size – of the probabilistic measures (see Sect. 3) with which the ensemble is evaluated.

These perturbations are designed to represent the major part of the uncertainties on the large-scale horizontal density gradient. We thus expect an impact on the mesoscale circulation that is observable by altimetric data. But we cannot reasonably expect that these perturbations can simulate all kinds of uncertainties that significantly influence the thermohaline circulation of the North Atlantic. This ensemble is designed to well-describe the mesoscale circulation with altimetric data, not to compensate for any model deficiencies in the description of the thermohaline structure of the ocean.

We now present a qualitative description of the 96-member ensemble running for 6 months (without any assimilation process). Figure 1 shows the variability of the dynamical fields (SSH) depending on the stochastic perturbations (only 6 members are shown here). This figure illustrates that the large-scale patterns are similar in all members, but every member presents a different eddy pattern: as expected, the largest variability of the ensemble is mostly located around the Gulf Stream front. Note that the introduction of the stochastic perturbations can produce gravity waves as observed on the right bottom panel.

Figure 2 summarizes the ensemble spread observed in Fig. 1 by showing the standard deviation after 6 months of growing stochastic perturbations: as expected, the ensemble spread is larger in the most active areas, especially around the Gulf Stream front.

Figure 2 also shows the verification areas and local points we used for the diagnostics in the rest of the paper. The main verification area (colour hexagon) focuses around the Gulf Stream. Most of the subtropical gyre is removed from the verification area because ensembles have almost no spread there and very weak altimetric signals are observed (compared to the signals in the Gulf Stream area). A smaller area is defined in the eastern Gulf Stream region (polygon Z) where the free-run ensemble spread is the largest after 6 months of integration. Finally, two points are selected to illustrate local diagnostics: A = 291.5° E, 35.5° N and B = 321.25° E, 45.5° N.

Local examples (points A and B) of time series of the ensemble spreads are shown in Fig. 6 (cyan curves) in next section. On that figure we mainly see differences in amplitudes and saturation times between the two locations: large spread and fast saturation in A, and smaller spread and slower saturation in B. Globally, the ensemble spread saturation over the whole Gulf Stream area is not really reached after 6 months. Nevertheless, the free run has been stopped because the global amplitude of the standard deviation is of the same order of magnitude as the unbiased RMSE (root mean

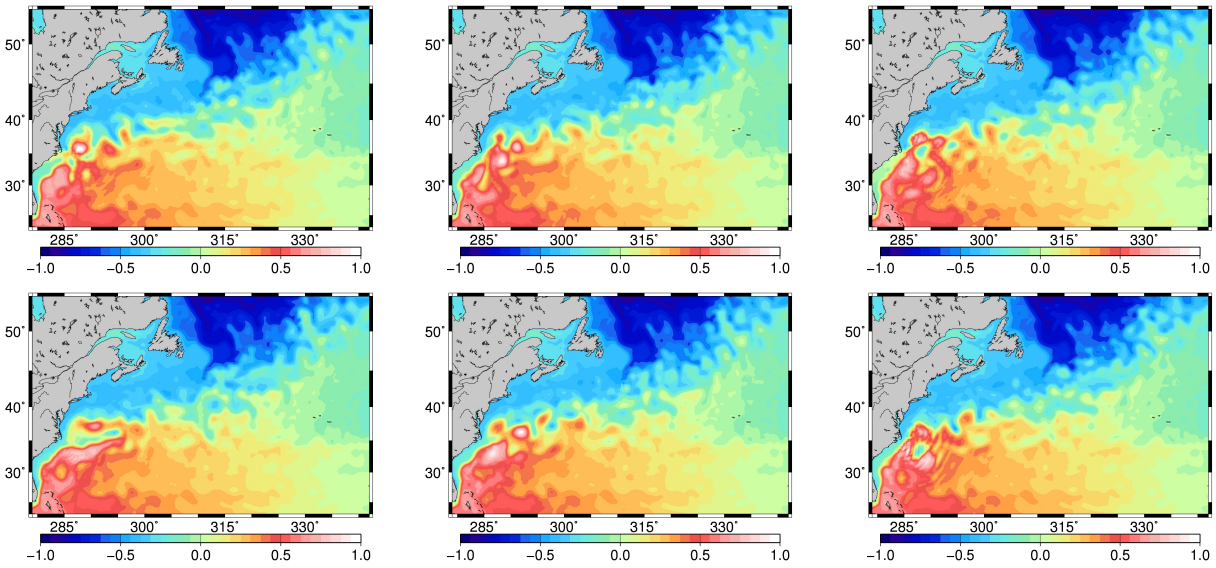


Figure 1. Six members from the free-run 96-member ensemble SSH (in m); snapshots for 9 July 2005.

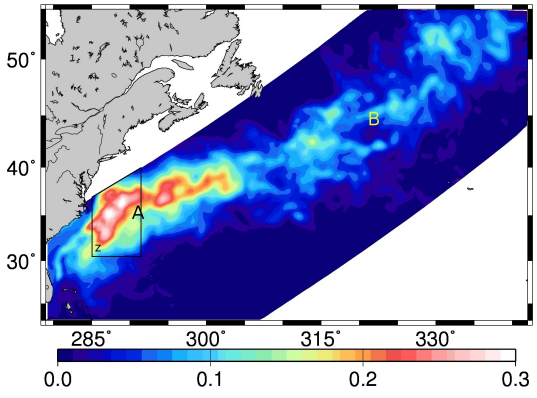


Figure 2. SSH standard deviation (in m) for 9 July 2005. The figure also shows the verification area (in colour) used for the numerical diagnostics together with the smaller focus area Z. Points A and B are particular locations used for diagnostics discussed elsewhere in the paper.

square error; not shown) between the ensemble mean and the GLORYS2V1 reanalysis. The GLORYS2V1 reanalysis is produced for the 1992–2009 period with NEMO global configuration at 0.25°, assimilating the T/S profiles, sea surface temperature (SST) and along-track SSH data (Ferry et al., 2012). This provides an estimation of the true state of the oceanic circulation that can be directly compared to the NATL025 outputs. In this way, we obtain a first estimation of the (stochastic) model error (ensemble mean against reanalysis) that is approximatively equal to the ensemble standard deviation after 6 months of simulation. For instance, averaged over the eastern Gulf Stream region (polygon Z on Fig. 7) and the sixth month of the integration, the model error is equal to 0.17 m and the standard deviation to 0.16 m.

Now, the impact of the stochastic perturbations on 3-D variables – T/S profiles – is investigated. Figure 3 shows an example of ensemble profiles at locations A and B. As expected, a larger spread is observed at location A, as compared to location B. In all cases, the spread decreases with depth and becomes negligible below 1000 m depth. Moreover, the spread on T and S mostly corresponds to a lifting and lowering of the water column which is certainly appropriate to assimilate altimetric observations in eddy active regions (Cooper and Haines, 1996). As a result, we can also anticipate that this ensemble is not appropriate to control other features of the vertical structure of the ocean.

To conclude this section, it is important to stress that such an ensemble simulation contains a lot of useful information about the model dynamics. Many other diagnostics could directly benefit from a probabilistic point of view, as for instance the local fluxes through the main straits, the local mixed layer depth, and the meridional overturning circulation. This is however out of the scope of this overview which aims at introducing the ensemble assimilation experiment. More important to us is the comparison between this ensemble simulation and the real-world observations (as provided by Jason-1, Envisat satellites and Argo floats). The purpose of the next section is thus to provide a quantitative probabilistic evaluation of the ensemble against these observations.

3 Ensemble validation

3.1 Probabilistic concepts

The ensemble evaluation issues are well-known and referenced in other communities like atmospheric sciences and meteorology (e.g. Toth et al., 2003), but are quite new in the

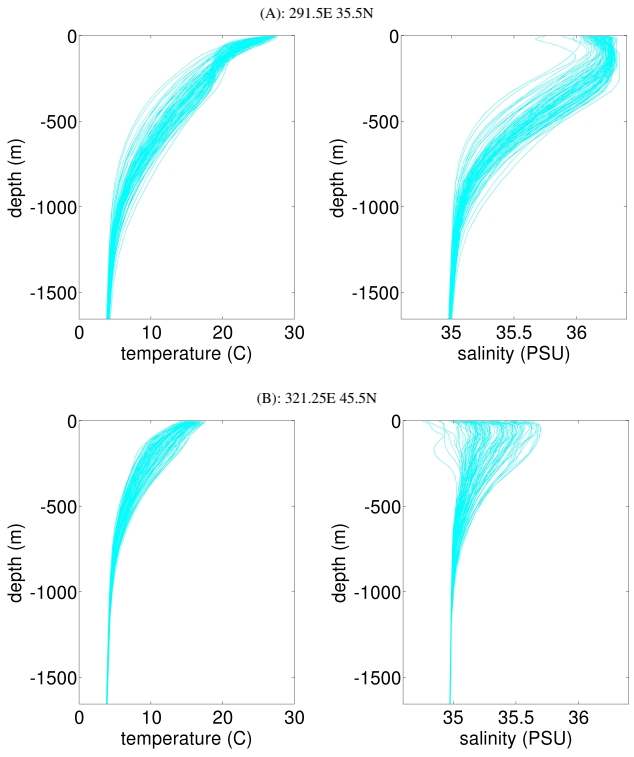


Figure 3. Local profiles, 96-member ensemble for temperature (left column) and salinity (right column), 9 July 2005.

field of oceanography. This subsection first introduces the basic concepts of probabilistic validation.

The main concern here is to introduce the specific criteria for evaluating the quality of a probabilistic system. Let us consider the following statement produced by an example of probabilistic system: "there is 30 % probability that the Northern Sea Route is free of ice". Assuming the event "free of ice" is unambiguously defined, neither its observed occurrence nor its nonoccurrence can be legitimately used to validate or invalidate the produced ensemble. Unlike a deterministic system, the validation of a probabilistic system cannot be performed over a single case (or realization). One must use a statistical approach, based on a sufficiently large set of realizations. Meaning this validation requires an aggregation of a large set of independent realizations of the considered process. After accumulating independent realizations of the probabilistic system, two probabilistic properties (attributes) can be measured: the reliability and the resolution.

In the example cited above, one has to wait until the 30 % probability is produced by the system a number of times. Then one can first check the proportion of actual observed occurrences of "free of ice". If that proportion is equal or close to 30 %, the probabilistic system can be considered as statistically consistent. If, on the contrary, that proportion is significantly different from 30 %, the system is statistically inconsistent. One condition for the validity of a probabilistic system is therefore the statistical consistency between produced probabilities and observed frequencies of occurrence of the event under consideration. This property of statistical consistency is called reliability. More generally, the reliability is the system's ability in producing pdfs in agreement with the associated observed pdf, i.e. the distribution of the observed variable when a given pdf is produced.

The reliability attribute is a necessary condition to have a skillful probabilistic system, but it is not a sufficient property. Actually, every system can be calibrated, i.e. it can be transformed into a reliable system by replacing the produced pdfs by the associated observed pdfs over a given verification set, and by applying this correction to the pdfs produced by the system over the subsequent verification set (under stationary assumption of the system). Also, if one knows the climatological distribution of the observed variables over a given verification data set, a system producing this distribution for each realization of the verification data set would be obviously reliable but it would provide no other useful information than the climatology (no need to integrate a complex numerical model to obtain this result). Here, the term "climatology" refers to either the distribution of the observations accumulated over a long past period or the distribution of the observations associated with the considered verification data set (in practice, the climatology often refers to the second option). For instance, one knows the climatological frequency of an ice-free Northern Sea Route is 2 months a year (occurrence $\approx 16\,\%$). If a probabilistic system produces the 16 % probability every day, it is reliable if one can evaluate its performance over a year, but it cannot provide any information about the seasonal (for instance) variability of that probability of occurrence. In other words, a climatological system would be perfectly reliable without providing any additional useful information. To determine if one has a skillful probabilistic system, another attribute is then needed.

The resolution is the system's ability to discriminate the distinct observed situations; this property is closely related to the information content and the entropy (e.g. Roulston and Smith, 2002). If the system is reliable, the resolution is also referred to as the sharpness which measures the spread of the produced pdfs. The resolution can then be seen as the spread of the associated observed pdfs. The sharper the associated observed pdfs compared to the climatological pdf, the better the resolution. In other words, the resolution is the additional information, compared to the climatology, that can be potentially extracted from the probabilistic system.

In summary, a skillful probabilistic system must satisfy both reliability and resolution criteria.

3.2 Practical probabilistic validation

Before introducing the different scores used to evaluate the reliability and the resolution, we briefly describe the way the probabilistic systems are validated in practice.

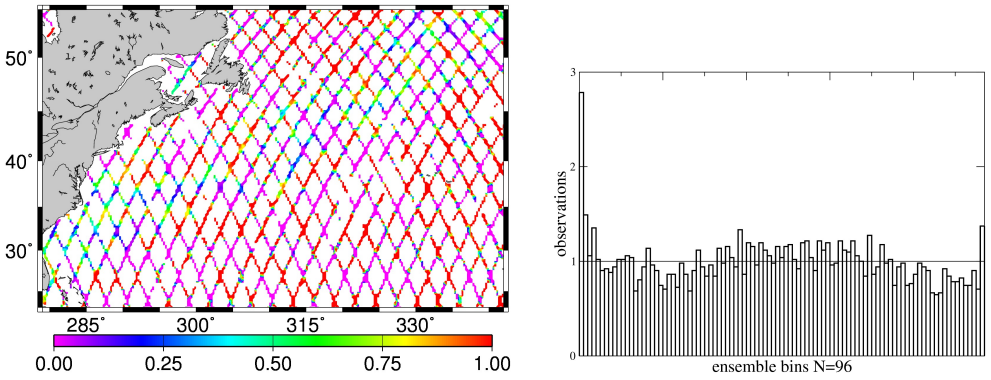


Figure 4. SSH Jason-1 tracks, local rank of the observations over 10-day Jason-1 cycle from 29 June to 9 July 2005 (left panel) and rank histogram over the Gulf Stream verification area (right panel).

As mentioned in the previous section, the validation of a probabilistic system must be statistically performed by accumulating large enough independent realizations of this system. For instance, in the following, the monthly diagnostics are computed by aggregating the data – for each variable – over the whole month and the chosen area for the same diagnosis. In this case, the climatology is the distribution of the available observations over this month and this area. This could lead to a meaningless diagnosis if the aggregated data are too heterogeneous over the chosen period and/or region. To evaluate the same probabilistic system, considering one verification data set or two separate sub-data sets could lead to very different conclusions in terms of reliability due to possible bias compensations, and in terms of resolution due to different climatological distributions (Hamill and Juras, 2006; Candille et al., 2007).

3.3 Probabilistic scores

We first check the reliability of the ensemble described in Sect. 3.1 by introducing the rank histogram (Anderson, 1996) and the reduced centred random variable (RCRV; Candille et al., 2007). Let us consider that the simulated pdf is represented by an ensemble of size N. For each realization of the system, the N ensemble members are ranked in increasing order, thereby defining $N+1$ intervals (or bins). Then we compute the rank of the verification (observation) within these bins, and the rank histogram is built by accumulation over all available realizations – assumed independent – of these ranks. The ensemble is reliable, i.e. the verification is statistically indistinguishable from the ensemble if it falls with equal probability in each of the $N+1$ intervals and then shows a flat rank histogram (uniform distribution of the ranks). A nonuniform rank histogram characterizes a lack of reliability of the system. For instance, an ensemble with many outliers, meaning that the verification values fall outside the ensemble, presents a U-shape rank histogram and is called underdispersive. On the other hand, an overdispersive ensemble, presenting a bell-shape rank histogram, means the verification values too often fall inside the ensemble. Also, a positive (negative) bias is characterized by the fact that the bins of the left (right) side of the histogram are overpopulated compared to the bins of the right (left) side of the histogram, i.e. the ensemble tends to over(under)estimate the verification value.

In Figure 4, the reliability of SSH is verified against along-track altimetric observations from Jason-1 during the 10-day cycle from 29 June to 9 July 2005. The left panel shows the observation ranks. Many outliers are observed, especially in the subtropical gyre. This is mainly due to the very small spread of the ensemble in this area (see Fig. 2) combined to an important bias in the simulation (ensemble mean minus observation). Actually, the ensemble spread is so small in this area that the outliers only reflect this local model bias. For the rank histogram construction (right panel), the statistics are only accumulated over the Gulf Stream verification area (see Fig. 2) to avoid aggregating too heterogeneous data from the frontal region and the gyre. Graphically, we observe a weak positive bias (asymmetric rank histogram to the left) and a slight underdispersion. Note that observational error is taken into account in the rank histogram construction by adding a Gaussian random noise to the ensemble members (with standard deviation consistent with the observation error used in the assimilation process, $\sigma_o = 10$ cm; see Sect. 4).

To numerically assess the lack of reliability graphically observed on the rank histogram, we introduce the RCRV as follows. For each realization of the system, the RCRV is defined as

$$y = \frac{o - m}{\sigma}, \tag{3}$$

where m and σ are the mean and the standard deviation of the simulated pdf and o the observed value. Note that the observation error σ_o is simply introduced in y by considering $\sigma = \sqrt{\sigma_{\text{ens}}^2 + \sigma_o^2}$. The system is reliable if the mean of y over all realizations of the probabilistic system is null and its standard deviation is equal to 1. In this way, the reliability is decomposed into a (normalized) bias $b = E[y]$ and a dispersion

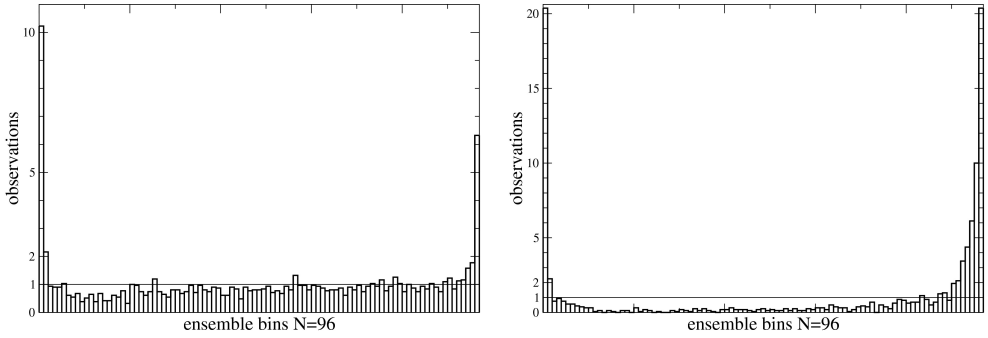


Figure 5. Global rank histogram of 0–200 m layer for temperature (left panel) and salinity (right panel); same period and same area as in Fig. 4.

$d^2 = E[y^2] - b^2$. For the example in Fig. 4, the normalized bias and the dispersion from the RCRV are equal to $b = -0.1$ and $d = 1.15$, respectively, i.e. the system has a weak positive bias (10 %) and a slight underdispersion (15 %).

These two diagnostics – the RCRV and the rank histogram – only measure the reliability of the system. We need to use other scores evaluating the resolution to get a full probabilistic assessment of the skill of the ensemble.

The continuous rank probability score (CRPS; introduced by Stanski et al., 1989) measures the global skill of a probabilistic system by evaluating both reliability and resolution. It can be seen as a generalization of the absolute error. It is based on the square difference between the produced cumulative distribution functions (cdfs) of a univariate variable x and the corresponding cdf of the observation:

$$\mathrm{CRPS} = E\left[\int_{\mathbb{R}} \left(F_\mathrm{p}(x) - F_o(x)\right)^2 \mathrm{d}x\right], \tag{4}$$

where F_p is the cdf associated with the produced pdf, F_o the cdf associated with the observation (a simple Heaviside distribution if no observation error is considered), and $E[\cdot]$ is the average of the integrals over the whole verification data set. Unlike the reliability scores presented above, the CRPS has the dimension of the verification variables (for instance expressed in metres for SSH, in Kelvin for temperature, etc). The CRPS can be decomposed into the reliability/resolution parts in many different ways. Here for practical and numerical reasons (see Candille and Talagrand, 2005), the decomposition described by Hersbach, 2000) is chosen:

$$\mathrm{CRPS} = \mathrm{Reli} + \mathrm{Resol}. \tag{5}$$

The term Reli can be seen as the expected value of the absolute error and the term Resol as a correction factor that measures the sharpness of the probabilistic system. These scores are negatively oriented, i.e. the reliability part (Reli) is null for a reliable system and the resolution part (Resol) goes from 0 for a perfect deterministic system to $\mathrm{Unc} = \int_{\mathbb{R}} F_c(x)\,(1 - F_c(x))\,\mathrm{d}x$ for a useless and noninformative system. F_c is the climatological cdf associated with the verification data set, and Unc – called uncertainty part of the CRPS – is the reference value of the CRPS only based on the variability of the verification data set. A value of Resol larger than Unc indicates that the system is poorer than climatology. Evaluated through the CRPS, a skillful probabilistic system must satisfy two criteria: Reli = 0 and Resol $\ll$ Unc. From the resolution criterion, a measure of the potential gain compared to the climatology can be defined as follows:

$$\mathcal{G} = 1 - \frac{\mathrm{Resol}}{\mathrm{Unc}}. \tag{6}$$

Considering the case in Fig. 4, the uncertainty associated with the along-track SSH of Jason-1 is Unc = 0.07 m, and the CRPS and its decomposition are equal to $0.076 = 0.005 + 0.071$. This indicates that the system is poorly informative with $\mathcal{G} > 1$ in this example. We can then conclude that – after 6 months of stochastic perturbations integration – the ensemble produced by the system tends to the climatology of the verification data set: pretty good reliability but poor resolution. In summary, we have here an example of a full probabilistic diagnosis on the SSH produced by the system against Jason-1 along-track observations. Of course, such SSH diagnostics can be performed against Envisat or both satellites.

The goal of the assimilation (next section) is then to improve the information contained in the system (i.e. improvement of the resolution), while keeping the system as reliable as possible (i.e. without deteriorating reliability).

Now, we evaluate the skill of the system for the T/S profiles against Argo observations. In this case, the statistics are accumulated during 10 days (Jason-1 cycle), over the Gulf Stream verification area and over the first 0–200 m layer of the ocean (observations from different depths are considered independent). The observation error is estimated from 0–200 m observation errors based on the Ingleby and Huddleston (2007) study about the quality control of T/S profile data: $\sigma_o = 0.9$ K for temperature and $\sigma_o = 0.17$ psu for salinity (these values take measure and representativeness errors into account).

Figure 5 shows the rank histograms for both temperature and salinity. Unlike the SSH rank histogram, a strong under-

Table 1. RCRV scores for 0–200 m layer temperature and salinity; same period and area as in Fig. 4.

	Bias	Dispersion
Temperature	−0.07	1.9
Salinity	0.50	2.5

Table 2. Same as Table 1 for CRPS.

	CRPS	Reli	Resol	Unc
Temperature	1.12	0.28	0.84	3.44
Salinity	0.32	0.13	0.19	0.86

dispersion is here observed for both variables and a negative bias is noticeable for salinity (asymmetry to the right side of the rank histogram).

The diagnostics related to the RCRV, shown in Table 1, numerically confirm the graphical diagnostics from the rank histograms. The system is strongly underdispersive for both profile variables and salinity is also strongly negatively biased.

Table 2 shows the global CRPS score, its reliability/resolution decomposition and the uncertainty associated with each verification data set. The potential gain compared to the uncertainty is $\mathcal{G} = 76\,\%$ and $\mathcal{G} = 78\,\%$ for temperature and salinity, respectively. These large values of the gain are mainly due to the large heterogeneity of these variables (horizontal and depth) over the verification domain. This leads to a very large spread of the climatological pdf used to compute the uncertainty, while the ensembles are able to capture this spatial variability, even with a very poor reliability. This highlights one of the crucial points of the statistical verification (see Sect. 3.2): in order to get a verification data set large enough, we sometimes need to aggregate many heterogeneous data.

Even if the potential reduction of uncertainty seems important (verification sample size issue), the system can hardly be considered as skillful for T/S profile variables considering the strong lack of reliability, especially the large underdispersion.

This kind of diagnostics for T/S profiles was expected since the stochastic perturbations of the model were not designed to explore the major uncertainties of the vertical structure of the ocean (see Sect. 2.2). On the other hand, it has been shown that the perturbations nevertheless produce a good representation of the uncertainties in the surface circulation, which is something that we can expect to be well-described with altimetric observations.

Using the prior ensemble described in this section, an assimilation scheme was implemented for the altimetric along-track observations from Jason-1 and Envisat missions. A significant improvement in the surface circulation (SSH) is expected, without negative impact on the thermohaline structure of the ocean.

4 4-D ensemble data assimilation experiment

The prior ensemble built in the previous section tends to reflect the climatological SSH variability of the flow over the Gulf Stream area. What could be the benefit from an assimilation process with altimetric data?

4.1 Methodology

In this study, only altimetric data are assimilated and the focus is on the eddy dynamics over the Gulf Stream. The observations are along-track data coming from two different satellites: 10-day cycle Jason-1 mission (≈ 350 km inter-track distance at the Equator) and 35-day cycle Envisat mission (≈ 80 km inter-track distance at the Equator). The assimilation cycles are defined to fit with the 10-day cycle of Jason-1 and are then performed within 10-day assimilation windows $[t_k, t_{k+10}]$. The update (see details below) is performed at the middle of the assimilation window t_{k+5} with all the observations and the model equivalent (i.e. model outputs projected to the exact observation times and locations) contained in the 10-day assimilation window. Increments are then computed for each ensemble member as explained below. The increment is then introduced into the model using the incremental analysis update (IAU) algorithm (Ourmières et al., 2006): a 10-day integration run from t_k by injecting fractions of the increment step by step all along the assimilation window. The full increment is thus introduced in the system at the final time t_{k+10} of the assimilation cycle. The IAU ensemble at the final time of the cycle provides the initial conditions for the forecast ensemble of the next cycle. The forecast ensemble trajectories are thus discontinuous from one cycle to the next, while the IAU ensemble trajectories remain continuous (and avoid possible numerical shocks which would occur if the increments were fully injected at one single time).

The assimilation scheme is a Kalman-filter-based method. The SEEK filter (Brasseur and Verron, 2006) is applied with a localization (equivalent to LETKF; Bishop et al., 2001). This is a square root algorithm mixed with an ensemble methodology (Burgers et al., 1998) where each member is individually updated. The ensemble approach enables bypassing the linearity assumption because each ensemble member is propagated by the nonlinear model $\mathcal{M}$ and directly projected into the observation space by operator $\mathcal{H}$ (via module NEMO_OBS) for the update. On the other hand, the Gaussian assumption still remains for the ensemble update but could easily be relaxed by anamorphosis transformations (not done here; Brankart et al., 2012).

Each member i of the forecast ensemble of vector state $\boldsymbol{x}$ is written as

$$\boldsymbol{x}_i^{\mathrm{f}} = \overline{\boldsymbol{x}^{\mathrm{f}}} + \delta\boldsymbol{x}_i^{\mathrm{f}}, \tag{7}$$

where $\overline{\boldsymbol{x}^{\mathrm{f}}}$ is the ensemble mean and $\delta\boldsymbol{x}_i^{\mathrm{f}}$ the associated anomalies which define the columns (with factor $1/\sqrt{N-1}$) of the forecast square root covariance matrix $\mathbf{S}^{\mathrm{f}}$ (covariance matrix $\mathbf{C}^{\mathrm{f}} = \mathbf{S}^{\mathrm{f}}\mathbf{S}^{\mathrm{f}T}$). Each ensemble member is projected into the observation space by $\mathcal{H}$ so that $\overline{\mathcal{H}\boldsymbol{x}^{\mathrm{f}}}$ is the forecast ensemble mean in observation space and $\delta\left(\mathcal{H}\boldsymbol{x}_i^{\mathrm{f}}\right)$ are the associated anomalies defining the columns (with factor $1/\sqrt{N-1}$) of the forecast square root covariance matrix in the observation space $\mathcal{H}\mathbf{S}^{\mathrm{f}}$.

The ensemble mean is then updated with a square root algorithm (Brankart et al., 2011), without requiring observation perturbation. The update is therefore computed in the eigenspace of

$$\boldsymbol{\Gamma} = (\mathcal{H}\mathbf{S}^{\mathrm{f}})^T\mathbf{R}^{-1}(\mathcal{H}\mathbf{S}^{\mathrm{f}}) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1}, \tag{8}$$

where $\mathbf{U}$ and $\boldsymbol{\Lambda}$ are, respectively, the unitary matrix and the diagonal matrix with the eigenvectors and eigenvalues of $\boldsymbol{\Gamma}$ (with no truncation). Note that the observation error covariance matrix $\mathbf{R}$ is diagonal: $\mathbf{R} = \sigma_o^2 * \mathbf{I}$ with $\sigma_o = 0.1$m (value taking account of both measure and representativeness errors). The transformation into the eigenspace of $\boldsymbol{\Gamma}$ is needed to make the algorithm compatible with the localization process (see localization parameters below). The analysis ensemble mean $\overline{\boldsymbol{x}^{\mathrm{a}}}$ is then updated as follows:

$$\overline{\boldsymbol{x}^{\mathrm{a}}} = \overline{\boldsymbol{x}^{\mathrm{f}}} + \mathbf{S}^{\mathrm{f}}\mathbf{U}(\mathbf{I}+\boldsymbol{\Lambda})^{-1}\mathbf{U}^T(\mathcal{H}\mathbf{S}^{\mathrm{f}})^T\mathbf{R}^{-1}\left(\boldsymbol{y}_o - \overline{\mathcal{H}\boldsymbol{x}^{\mathrm{f}}}.\right). \tag{9}$$

$\overline{\boldsymbol{x}^{\mathrm{a}}}$ thus depends on the innovation $\boldsymbol{y}_o - \overline{\mathcal{H}\boldsymbol{x}^{\mathrm{f}}}$, the observation error covariance $\mathbf{R}$, and the anomalies expressed both in model and observation spaces: $\delta\boldsymbol{x}_i^{\mathrm{f}}$ and $\delta\left(\mathcal{H}\boldsymbol{x}_i^{\mathrm{f}}\right)$.

In a second time, each ensemble anomaly i is updated as follows:

$$\delta\boldsymbol{x}_i^{\mathrm{a}} = \sqrt{N-1}\left(\boldsymbol{S}^{\mathrm{f}}\boldsymbol{U}(\boldsymbol{I}+\boldsymbol{\Lambda})^{-\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{U}^T\right)_i, \tag{10}$$

so that each updated ensemble member $\boldsymbol{x}_i^{\mathrm{a}}$ can be rebuilt from the updated ensemble mean and its associated updated ensemble anomaly:

$$\boldsymbol{x}_i^{\mathrm{a}} = \overline{\boldsymbol{x}^{\mathrm{a}}} + \delta\boldsymbol{x}_i^{\mathrm{a}}. \tag{11}$$

The increments $\delta\boldsymbol{x}_i = \boldsymbol{x}_i^{\mathrm{a}} - \boldsymbol{x}_i^{\mathrm{f}}$ are then computed and are introduced into the model using the IAU method in order to produce a continuous updated ensemble. It is also important to mention that the stochastic version of the model is used for the assimilation cycle (same as for the free-run integration). Actually, the stochastic perturbations are sufficient to avoid the collapse of the ensemble by ensuring an appropriate spread.

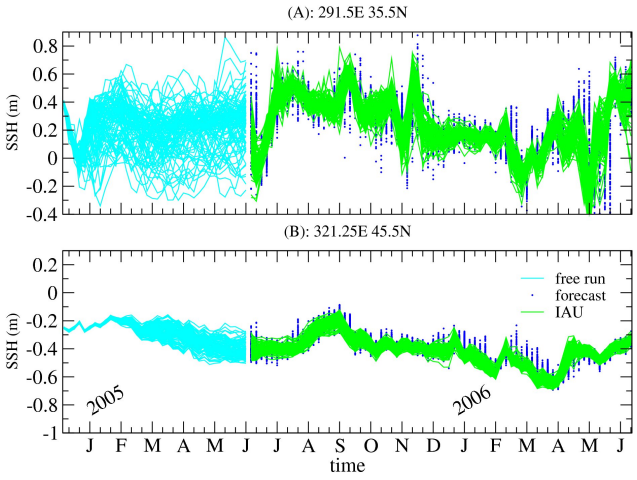


Figure 6. SSH time series of the ensemble (free run, forecast and IAU) at locations **(a)** and **(b)**.

To avoid the spurious effect of inaccurate long-range correlations the update is also performed with a localization algorithm: the local assimilation areas are limited by a radius of 4.5° (≈ 450 km at 30° N) and the observations' influences are defined by Gaussian functions with standard deviation of 1.5° (≈ 150 km at 30° N). The localization is applied on a state vector as described in Brankart et al., 2011.

As an illustration of the assimilation process described above, local time evolutions of the 96-member ensemble are shown in Fig. 6 for 18 months (6 free-run months and 12 assimilation months).

As already mentioned, we observe the saturation of the spread of the free-run ensemble (the cyan curves stop spreading) with different amplitudes and saturation timescales depending on the locations. This saturation shows that the local climatological variabilities are reached. The updated ensembles (blue dots for the forecast ensemble and green curves for the IAU ensemble) present a noticeable spread reduction. Also, they present a temporal variability globally included in the climatological envelop defined by the free-run ensemble saturation. This kind of ensemble behaviours could foreshadow an improvement of the probabilistic resolution without degrading the reliability. We also note the spread reduction (blue dots outside the green curves area) and a slight bias correction (asymmetry of the blue dot outliers) with the IAU ensemble compared to the forecast one.

The examples in Fig. 6 are a first qualitative evaluation of the updated ensembles. The next subsection presents more quantitative and probabilistic evaluations of these ensembles.

4.2 Assimilation results

We only focus this study on SSH (Jason-1/Envisat altimetric data sets) and T/S profiles (Argo buoys network).

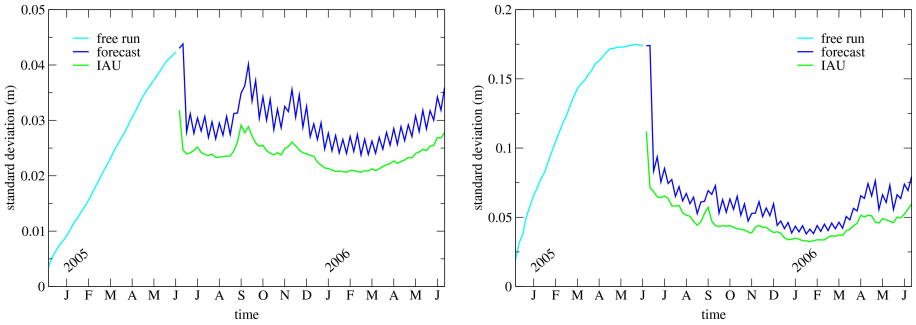


Figure 7. SSH time series of standard deviations for free-run, forecast and IAU ensembles over the Gulf Stream verification area (left panel) and the Z focus area (right panel).

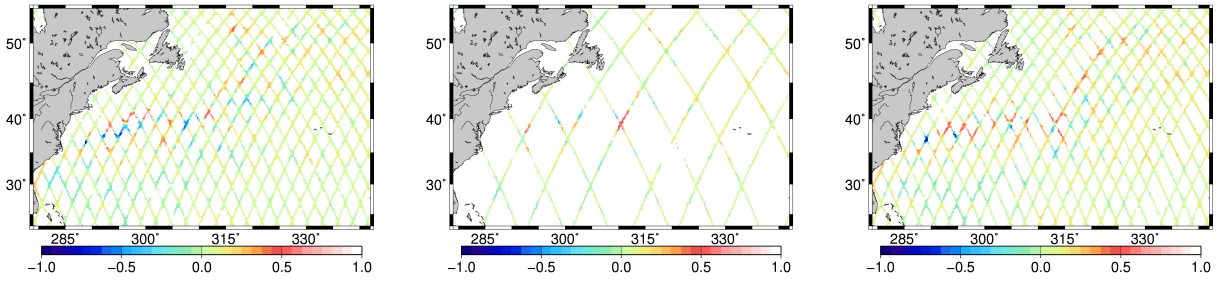


Figure 8. SSH Jason-tracks coverage for three successive 10-day cycles from 7 September to 7 October 2005.

In order to generalize the first diagnostics suggested by Fig. 6, Fig.7 shows the time series of the ensembles standard deviation averaged over the Gulf Stream and Z-focus areas.

In this figure we can see first that the saturation of the standard deviation is reached in the Z-focus area but not over the global Gulf Stream area. The main reason for this is that the Gulf Stream verification area contains locations where the perturbations grow very slowly (e.g. near the subtropical gyre). The main point here is the reduction of the standard deviations observed with the introduction of the altimetric corrections. This reduction is effective from the first assimilation cycle. After that, the averaged standard deviation of the forecast and IAU solutions is globally stabilized by the subsequent assimilation cycles. We also notice the clear reduction of the standard deviations of the IAU ensemble as compared to the forecast. This means that the stochastic perturbations are strong enough to produce a significant spread within 10 days and thus to avoid the ensemble collapse. The 10-day oscillations of the standard deviation is the result of the discontinuity of the forecast ensembles.

As mentioned just before, the altimetric updates tend to reduce and stabilize the standard deviation, except the spurious increase observed around September 2005 (especially over the large verification area). This event is caused by a lack of observed data resulting from missing Jason-1 tracks in September 2005, as shown in Fig. 8.

This occasional Jason-1 failure shows the impact of the satellite coverage on the updated ensembles, resulting in an increase of the ensembles standard deviation. That may sound obvious, but it shows that the accuracy of the correction is very sensitive to the number of available altimetric data.

The same kind of behaviour is observed on the T/S variables (not shown): the standard deviation increases with the free-run ensemble and is reduced and stabilized by means of altimetric corrections. Nevertheless, the difference between the IAU and forecast ensembles standard deviations is much smaller along the T/S profiles, simply because these variables are not assimilated (unlike in Yan et al., 2014).

4.3 Probabilistic diagnostics for SSH

After studying the general characteristics of the ensemble, we now present the probabilistic diagnostics introduced in Sect. 3. We first investigate the reliability property of SSH through the bias and the dispersion related to the RCRV. In Fig. 9, three sets of curves are shown: dashed curves for verification against Jason-1 data, dotted curves for verification against Envisat data and solid curves for verification against both satellites' data. Note also that – for all the presented probabilistic scores in this section – the statistics are accumulated over 1 month (different from Sect. 3) and over the Gulf Stream area. In the interpretation of the results, it is also very important to remark that the IAU ensemble is checked against observations that have been used to compute the increments. In this case, the ensemble system and the observations are not independent. For the SSH variable, the 10-

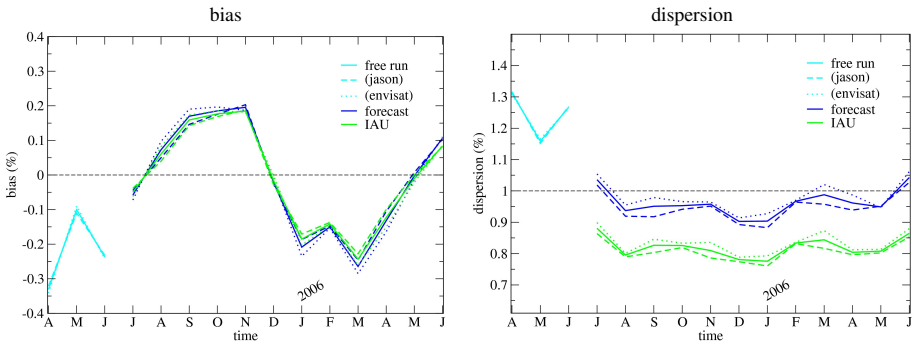


Figure 9. SSH bias (left panel) and dispersion (right panel) from RCRV.

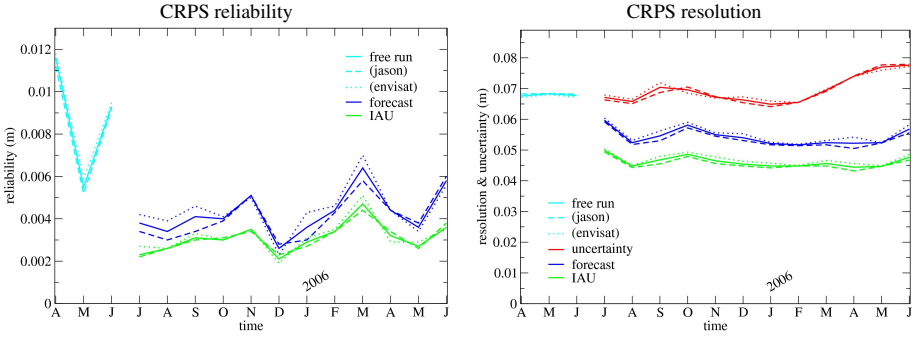


Figure 10. SSH: reliability (left panel) and resolution components of the CRPS compared to the uncertainty (right panel).

day forecast ensemble and observations are truly independent data.

The monthly time series (left panel) show no clear bias reduction compared to the free-run ensemble with altimetric corrections. The bias seems to present a seasonal cycle (not explained). Nevertheless, the IAU ensemble reduces the bias as compared to the forecast. This is due to the inbreeding between the IAU ensembles and the observations in this case. Regarding the dispersion (right panel), the underdispersion of the free-run ensembles is removed by the altimetric corrections. The forecast ensemble becomes almost perfectly dispersive while a slight overdispersion ($\approx 85\,\%$) is observed for the IAU ensemble (mostly due to the inbreeding with the observations).

Two results look contradictory at first sight: we observe the spread reduction of the IAU ensembles (Fig. 7) and at the same time the IAU ensembles overdispersion (Fig. 9 right panel). Actually, the spread of the ensemble is reduced, but also the bias against the observations so that the dispersion is finally degraded down to an overdispersive system.

We now investigate the global CRPS measure for the SSH variable. Figure 10 shows the reliability (left panel) and the resolution (right panel) components of the CRPS. The resolution part is compared to the uncertainty associated with the verification data set (this represents the climatological variability over the verification area and period).

The reliability – as measured by the CRPS – of the ensembles is improved by the altimetric corrections compared to the free-run ensemble. Also, the system becomes more informative after assimilation processes (better resolution), i.e. the ensemble system is more accurate in space and time for instance by correctly translating the eddies along the Gulf Stream front. For both components of the CRPS, the IAU ensemble performs better than the forecast (these scores are negatively oriented), partially due to the inbreeding between the IAU ensemble and the observations. The potential gain $\mathcal{G}$ compared to the uncertainty is shown in Table 3.

The resolution curves in Fig. 10 (right panel) show that the free-run ensemble has no resolution, i.e. $\mathcal{G} \approx 0\,\%$. If we only look at the independent verification data, i.e. the forecast ensemble, the potential gain we get with the assimilation process is up to 30 %.

For the SSH variable the assimilation process leads to an information gain with a reliability improvement. Basically, the uncertainty (not the uncertainty from the CRPS) on the 10-day forecast is reduced by 30 % and this information is reliable.

4.4 Probabilistic diagnostics for T and S

We now present the impact of the altimetric corrections on T/S profiles. The verification is performed against Argo observations. To investigate the probabilistic attributes of our

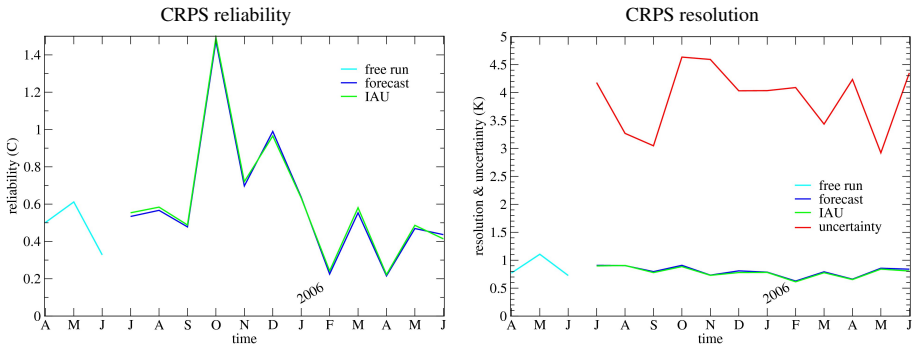


Figure 11. T profiles, reliability (left panel) and resolution components of the CRPS compared to the uncertainty (right panel).

Table 3. Potential gain $\mathcal{G}$ compared to the uncertainty from July 2005 to June 2006.

	July 2005	August 2005	September 2005	October 2005	November 2005	December 2005	January 2006	February 2006	March 2006	April 2006	May 2006	June 2006
Forecast	11 %	20 %	22 %	16 %	18 %	19 %	20 %	21 %	25 %	30 %	32 %	27 %
IAU	26 %	32 %	34 %	30 %	31 %	32 %	31 %	32 %	34 %	40 %	42 %	39 %

ensembles along T/S profiles, we show the CRPS reliability part and the information gain (CRPS resolution part) for temperature in the 0–200 m layer depth (Fig. 11).

Regarding T profiles, no noticeable difference can be detected in the score between the free-run ensemble and the updated ones. Furthermore, there is almost no difference in the score between the forecast and the IAU ensembles. Considering the potential gain, the same remark can be done here as for the free-run ensemble (see Sect. 3, Table 2).

On the other hand, the reliability CRPS part is very variable from month to month compared to the SSH diagnostics. This can be explained by the spatial distribution of the observed Argo profiles used for the verification: for SSH this distribution is pretty similar each month for the SSH (three full Jason-1 cycles and 1 full Envisat cycle) except in exceptional situations like the one shown in Fig. 8, but for the T/S profiles it is very variable depending on the Argo buoy locations. Actually, from one month to the other, very different areas are sampled by the Argo moving network. As a result, we can conclude that it is very difficult for the CRPS to measure any possible correction on T/S profiles.

The results presented above indicate that the assimilation process leads to a more skillful probabilistic system for a SSH 10-day forecast and then for surface currents. The altimetric corrections applied to the stochastic ensemble show clear positive impacts in terms of

- improvement of the reliability of the system forecast, i.e. the confidence in the description of the uncertainty associated with the state of the flow increases;

- improvement of the resolution, i.e. reduction of the uncertainty by 30 %.

5 Conclusions

One important source of uncertainty of the model – not the only one – is simulated by introducing stochastic parameterizations in the formulation of the equation of state (as in Brankart 2013). The effects of the unresolved scales on the ocean circulation are thus simulated, and a stochastic perturbations ensemble is produced with the stochastic NATL025-formulation. This ensemble is then objectively compared to altimetric observations (Jason-1/Envisat along-track data) and in situ T/S profile data (Argo buoys network). The present study shows that this ensemble correctly represents the climatological variability of the eddy dynamics over the Gulf Stream area, especially in the frontal regions. The ensemble system (without assimilation) tends to be reliable in those regions – even if it is globally underdispersive over the entire North Atlantic Basin – but provides no useful information on the mesoscale circulation (null probabilistic resolution associated with climatological system) as a result of the chaotic nature of the eddy flow. The underdispersion is also due to the fact that only one aspect of the model error is considered here. The ensemble is then updated by assimilating altimetric along-track data (Jason-1 and Envisat) through a full 4-D ensemble assimilation process: the covariance matrix is propagated by the ensemble and model equivalent of every member computed at the exact observation time and location. The assimilation makes the ensemble more reliable – the underdispersion is reduced – and more informative compared to climatology. The updated ensemble system is probabilistically more skillful, considering the reliable reduction of uncertainty by 30 %.

The experiment presented in this study shows promising results in terms of uncertainty, simulation and uncertainty reduction, but a number of aspects needs still to be improved.

First, the reduction of uncertainty has only been proved for the SSH analysis and the 10-day forecast. No significant improvement could be objectively assessed for observed T/S profiles (using the RCRV or CRPS measures). A positive objective assessment is difficult to obtain for T and S because the observation coverage (Argo floats) is still sparse and quickly changes with time, because the computation of the scores thus requires aggregating observations with very heterogeneous statistics, and because any possible improvement of the T and S mesoscale structure is dominated by other sources of uncertainty (like the large-scale model bias). Other sources of uncertainty should be introduced to effectively assimilate these observations, for instance uncertainties in the atmospheric forcing (as in Yan et al., 2014), and uncertainties in the initial condition, uncertainties in the model parameterizations (like the vertical mixing, the mixed layer dynamics). Further work should be dedicated to design a probabilistic model that is already reliable for T and S (or at least quite dispersive enough) before starting data assimilation, as we did in this paper for altimetry. This may require spending considerable time and effort to fine tune the parameterization of the various sources of uncertainty. To simplify this task, automatic procedure to estimate unknown statistical parameters could be very helpful. Then, after assimilation has started, these procedures could continue their work, for instance, by using adaptative tuning of the parameters linked to model uncertainty (i.e. a kind of generalization of the adaptative estimator of model error covariances; see Dee, 1995). As we have tried to show in this paper, a correct simulation of model uncertainty is indeed necessary to produce consistent probabilistic ocean forecasts, to perform ensemble data assimilation and, most importantly, to obtain objective verification scores.

The Supplement related to this article is available online at doi:10.5194/os-11-425-2015-supplement.

Acknowledgements. This work is a contribution to the OSTST project – Toward the Next Generation of Altimeter Data Assimilation for Physical Ocean and Marine Ecosystem Monitoring and Prediction – supported by CNES, and to the EU project – Stochastic Assimilation for the Next Generation Ocean Model Applications (SANGOMA). This work received partial support from the European Community's Seventh Framework Program FP7/2007-2013 under grant agreements no. 283367 (MyOcean2) and no. 283580 (SANGOMA). The calculations were performed using NEC-SX8 from GENCI-IDRIS resources (project i20130111279).

Edited by: M. Meier

References

Anderson, J.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations, J. Climate, 9, 1518–1530, 1996.

Barnier, B., G., Madec, T., Penduff, J.-M., Molines, A.-M., Treguier, J., Le Sommer, A., Beckmann, A., Biastoch, C., Böning, J., Dengg, C., Derval, E., Durand, Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., McClean, J., and De-Cuevas, B.: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy permitting resolution, Ocean Dynam., 56, 543–567, 2006.

Bishop, H. C., Etherton, B. J., and Majumdar, S. J.: Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: theoretical aspects, Mon. Weather Rev., 129, 420–436, 2001.

Bouttier, P.-A., Blayo, E., Brankart, J.-M., Brasseur, P., Cosme, E., Verron, J., and Vidard, A.: Toward a data assimilation system for NEMO, Merc. Quart. Newsl., 46, 24–30, 2012.

Brankart, J.-M.: Impact of uncertainties in the horizontal density gradient upon low resolution global ocean model, Ocean Model., 66, 64–76, 2013.

Brankart, J.-M., Cosme, E., Testut, C.-E., Brasseur, P., and Verron, J.: Efficient local error parameterization for square root or ensemble Kalman filters: application to a basin-scale ocean turbulent flow, Mon. Weather Rev., 139, 474–493, 2011.

Brankart, J.-M., Testut, C.-E., Béal, D., Doron, M., Fontana, C., Meinvielle, M., Brasseur, P., and Verron, J.: Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations, Ocean Sci., 8, 121–142, doi:10.5194/os-8-121-2012, 2012.

Brasseur, P. and Verron, J.: The SEEK filter method for data assimilation in oceanography: a synthesis, Ocean Dynam., 56, 650–661, 2006.

Brusdal K., Brankart, J.-M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P. J., Dombrowsky, E., and Verron, J.: A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems, J. Mar. Syst., 40, 253–289, 2003.

Buizza, R., Miller, M., and Palmer, T. N.: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, Q. J. Roy. Meteor. Soc., 125, 2887–2908, 1999.

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, 1998.

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

Candille, G., Côté, C., Houtekamer, P. L., and Pellerin, G.: Verification of an ensemble prediction system against observations, Mon. Weather Rev., 135, 2688–2699, 2007.

Cooper, M. and Haines, K.: Altimetric assimilation with water property conservation, J. Geophys. Res., 101, 1059–1077, 1996.

Dee, D. P.: On-line estimation of error covariance parameters for atmospheric data assimilation, Mon. Weather Rev., 123, 1128–1145, 1995.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10143–10162, 1994.

Evensen, G.: The Ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, 2003.

Ferry, N., Parent, L., Garric, G., Bricaud, C., Testut, C.-E., Le Galloudec, O., Lellouche, J.-M., Drevillon, M., Greiner, E., Barnier, B., Molines, J.-M., Jourdain, N. C., Guinehut, S., Cabanes, C., and Zawadzki, L.: GLORYS2V1 global ocean reanalysis of the altimetric era (1992–2009) at meso scale, Mercator Quarterly Newsletter, 44, 29–39, 2012.

Haines, K.: Ocean data assimilation, in: Data assimilation: Making sense of observations, edited by: Lahoz, W., Khattatov, B., and Menard, R., Springer-Verlag, Berlin Heidelberg, 517–548, 2010.

Hamill, T. M. and Juras, J.: Measuring forecast skill: is it real skill or is it the varying climatology?, Q. J. Roy. Meteor. Soc., 132, 2905–2923, 2006.

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

Ingleby, B. and Huddleston, M.: Quality control of ocean temperature and salinity profiles: historical and real-time data, J. Marine Syst., 67, 158–175, 2007.

Kalnay, E.: Ensemble Kalman filter: current status and potential, in: Data assimilation: Making sense of observations, edited by: Lahoz, W., Khattatov, B., and Menard, R., Springer-Verlag, Berlin Heidelberg, 69–92, 2010.

Kitsios, V., Frederiksen, J. S., and Zidikheri, M. J.: Scaling laws for parameterisations of subgrid eddy–eddy interactions in simulations of oceanic circulations, Ocean Model., 68, 88–105, 2013.

Lea, D., Martin, M., Mogensen, K., Vidard, A., and Weaver, A.: Observation and Model Comparison in NEMO Ocean Engine v3.4. Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace, 27, Chapter 12, 2012.

Levitus, S., Boyer, T. P., Conkright, M. E., O'Brian, T., Antonov, J., Stephens, C., Stathopolos, L., Johnson, D., and Gelfeld, R.: World Ocean Database 1998. NOAA Atlas NESDID 18, US Government Printing Office, Washington DC, 1998.

Nichols, N. K.: Mathematical concepts of data assimilation, in: Data assimilation: Making sense of observations, edited by: Lahoz, W., Khattatov, B., and Menard, R., Springer-Verlag, Berlin Heidelberg, 13–39, 2010.

Ourmières, Y., Brankart, J.-M., Berline, L., Brasseur, P., and Verron, J.: Incremental analysis update implementation into a sequential ocean data assimilation system, J. Atmos. Ocean. Tech., 23, 1729–1744, 2006.

Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, F. J., Jung, T., and Leutbecher, M.: Representing model uncertainty in weather and climate prediction, Annu. Rev. Earth Pl. Sc., 33, 163–193, 2005.

Porta Mana, P. and Zanna, L.: Toward a stochastic parameterization of ocean mesoscale eddies, Ocean Model., 79, 1–20, 2014.

Roulston, M. S. and Smith, L.: Evaluating probabilistic forecast using information theory, Mon. Weather Rev., 130, 1653–1660, 2002.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R., and Korablev, A.: TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Sci., 8, 633–656, doi:10.5194/os-8-633-2012, 2012.

Stanski, H. R., Wilson, L. J., and Burrows, W. R.: Survey of common verification in meteorology, World Weather Watch Report No. 8, TD No. 358, World Meteorological Organization, Geneva, Switzerland, 114 pp., 1989.

Talagrand, O.: Variational assimilation, in: Data Assimilation: Making Sense of Observations, edited by: Lahoz, W., Khattatov, B., and Menard, R., Springer-Verlag, Berlin Heidelberg, 41–67, 2010.

Testut, C.-E., Brasseur, P., Brankart, J.-M., and Verron, J.: Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy permitting primitive equation model of the North Atlantic Ocean, J. Marine Syst., 40–41, 291–316, 2003.

Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and ensemble forecasts, in: Forecast Verification: a Practitioner's Guide in Atmospheric Science, edited by: Jolliffe, I. and Stephenson, D. B., Wiley, UK, 137–163, 2003.

Uppala, S. M., Kållberg, P. W., and Simmons, A. J., et al.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, 2005.

Verron, J., Gourdeau, L., Pham, D. T., Murtugudde, R., and Busalacchi, A.: An extended Kalman filter to assimilate satellite altimeter data into a non-linear numerical model of the tropical pacific: method and validation, J. Geophys. Res., 104, 5441–5458, 1999.

Yan, Y., Barth, A., Beckers, J.-M., Candille, G., Brankart, J.-M., and Brasseur, P.: Ensemble assimilation of ARGO temperature profile, sea surface temperature and altimetric satellite data into an eddy permitting primitive equation model of the North Atlantic Ocean, J. Geophys. Res.-Oceans, under review, 2015.