

## Estimation and Impact of Nonuniform Horizontal Correlation Length Scales for Global Ocean Physical Analyses

ANDREA STORTO, SIMONA MASINA, AND SRDJAN DOBRICIC

*Numerical Applications and Scenarios Division, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy*

(Manuscript received 11 March 2014, in final form 20 May 2014)

### ABSTRACT

Optimally modeling background-error horizontal correlations is crucial in ocean data assimilation. This paper investigates the impact of releasing the assumption of uniform background-error correlations in a global ocean variational analysis system. Spatially varying horizontal correlations are introduced in the recursive filter operator, which is used for modeling horizontal covariances in the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) analysis system. The horizontal correlation length scales (HCLSs) were defined on the full three-dimensional model space and computed from both a dataset of monthly anomalies with respect to the monthly climatology and through the so-called National Meteorological Center (NMC) method. Different formulas for estimating the correlation length scale are also discussed and applied to the two forecast error datasets. The new formulation is tested within a 12-yr period (2000–11) in the  $1/2^\circ$  resolution system. The comparison with the data assimilation system using uniform background-error horizontal correlations indicates the superiority of the former, especially in eddy-dominated areas. Verification skill scores report a significant reduction of RMSE, and the use of nonuniform length scales improves the representation of the eddy kinetic energy at midlatitudes, suggesting that uniform, latitude, or Rossby radius-dependent formulations are insufficient to represent the geographical variations of the background-error correlations. Furthermore, a small tuning of the globally uniform value of the length scale was found to have a small impact on the analysis system. The use of either anomalies or NMC-derived correlation length scales also has a marginal effect with respect to the use of nonuniform HCLSs. On the other hand, the application of overestimated length scales has proved to be detrimental to the analysis system in all areas and for all parameters.

### 1. Introduction

It is acknowledged that one of the most crucial tasks in developing data assimilation systems is the design of the background-error covariances and their estimation. Usually, variational data assimilation systems define a *climatological* background-error covariance matrix, unlike Kalman filters (e.g., the ensemble Kalman filters), which propagate forward in time the background errors. However, four-dimensional variational data assimilation systems implicitly evolve the background-error covariances within the assimilation window. Over recent years, the meteorological variational assimilation community is devoting efforts in allowing the background-error covariances to be flow dependent (Isaksen et al. 2007; Raynaud et al. 2011).

Modeling horizontal correlations is generally one of the most computationally expensive tasks within the variational data assimilation minimization. It is responsible for propagating over the domain the informative content of the observations. Background-error vertical and horizontal covariances in oceanographic models require a gridpoint formulation as a result of the presence of the bathymetry, which may significantly complicate the problem (Gaspari et al. 2006). Gridpoint filters, such as the diffusion equation operator (Weaver and Courtier 2001) and the recursive filters (Purser et al. 2003a,b), represent an efficient way of accounting for spatial correlations while preserving the spatial discontinuities typical of ocean models. The latter method has been shown to be faster in regional applications (Dobricic and Pinardi 2008) with respect to both explicit filters, which require a long time stepping, and implicit filters. However, Mirouze and Weaver (2010) have recently shown the close analogy between the two approaches. The implementation of recursive filters on massively parallel computers

---

Corresponding author address: Andrea Storto, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Viale Aldo Moro 44, I-40127 Bologna, Italy.  
E-mail: andrea.storto@cmcc.it

requires more care than diffusion operators, because of nontrivial algorithms needed for the parallelism of the recursive filter, although successful implementations have been recently shown.

Many global ocean variational data assimilation systems (e.g., Weaver et al. 2005; Storto et al. 2011) do not currently account for local variations of the background-error horizontal correlations in a sophisticated way. This represents a major simplification in the analysis system, as forecast errors propagate differently depending on large-scale (e.g., by latitudinal bands; Derber and Rosati 1989) and local regime characteristics. Although both recursive filters and diffusion operators have been theoretically extended to include smoothed variations of the correlation length scales (Purser et al. 2003b; Gaspari et al. 2006; Mirouze and Weaver 2010), no practical estimation and impact study of the locally varying correlation length scales have been so far provided at the global scale. Zhou et al. (2004) were able to show the positive impact of the local variations of the horizontal correlation length scales (HCLSs) in the tropical Pacific; however, because of the coarse resolution of their model configuration, they were able to analytically construct the horizontal correlations of the background-error covariance matrix, which is unfeasible for high-resolution global applications. Other formulations, like in Cummings (2005) and Carton et al. (1996), simplify the local dependence of the correlation length scales as a function of latitude, derived from the first baroclinic Rossby radius of deformation or innovation statistics, respectively. The former approach reduces in first approximation to a latitudinal dependence of the correlation radius, provided that the geographical variability of the Rossby radius of deformation is dominated by its inverse dependence on the Coriolis parameter (Chelton et al. 1998), thus neglecting any information on the local variability of ocean state and the local dynamical regimes. The latter instead relies on the amount of available observations; therefore, it can appear questionable at global scale, characterized by a very irregular observation coverage.

The present study details the implementation of locally varying horizontal correlation length scales in the global implementation of software OceanVar (Dobricic and Pinardi 2008), a three-dimensional variational data assimilation system used for both regional operational analysis (Pujol et al. 2010) and retrospective analyses at both global scale (Storto et al. 2011) and regional scale (Adani et al. 2011). The paper is structured as follows: the OceanVar data assimilation system is briefly recalled (section 2), methods and results of the estimation of inhomogeneous correlation length scales are presented in section 3, and the new correlation length scales

are validated (section 4); section 5 discusses the main achievements.

## 2. The analysis system

This section briefly describes the analysis system used in this study. The resolution of the model is of about  $0.5^\circ$  in both zonal and meridional directions with 50 vertical height levels; the grid configuration is called ORCA05L50 and the horizontal meshing avoids pole singularities through a tripolar geometry (Madec and Imbard 1996). The resolution of the model varies from about 26 km at high latitudes to 55 km near the equator for both zonal and meridional directions. The ocean general circulation model (OGCM) is the Nucleus for European Modelling of the Ocean (NEMO), version 3.2 (Madec et al. 1998), coupled with the Louvain-la-Neuve Sea Ice Model, version 2 (LIM2), sea ice model (Fichefet and Morales Maqueda 1997). Surface boundary conditions are obtained through the Common Ocean Reference Experiment (CORE) bulk formulation (Large and Yeager 2004) by using 3-hourly turbulent and daily radiative and freshwater fluxes from the Interim European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-Interim) atmospheric reanalysis (Simmons et al. 2007; Dee et al. 2011). Shortwave radiations are modulated to reproduce daily cycles as formulated by Bernie et al. (2007). The freshwater input from the rivers is taken from the monthly climatology of Dai and Trenberth (2002). Since the system is conceived mostly for reanalysis applications, a number of corrections to the atmospheric forcing fields are applied, as explained in Storto et al. (2012), who also details the OGCM configuration.

The data assimilation step consists of minimizing the usual three-dimensional variational data assimilation (3DVAR) cost function  $J$ , in its incremental formulation (Courtier et al. 1994), with respect to the state vector  $\mathbf{x}$ :

$$J = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[\mathbf{d} - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)]^T \mathbf{R}^{-1}[\mathbf{d} - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)], \quad (1)$$

where  $\mathbf{x}^b$  is the background model state;  $\mathbf{B}$  is the background-error covariance matrix;  $\mathbf{d}$  is the vector of innovations (or misfits), namely, the vector containing the differences between the observations and the model equivalents;  $\mathbf{H}$  is the observation operator; and  $\mathbf{R}$  is the observational error covariance matrix. Preconditioning of the background term of the cost function in OceanVar implies that a change of variable is performed prior to the cost function evaluation, by means of which the new control variable  $\mathbf{v}$  is chosen, such that  $\mathbf{x} - \mathbf{x}^b = \mathbf{V}\mathbf{v}$  with the background-error covariance matrix  $\mathbf{B}$  equal to

$\mathbf{B} = \mathbf{V}\mathbf{V}^T$ . Therefore, the optimization problem of Eq. (1) simplifies to

$$J = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}(\mathbf{d} - \mathbf{H}\mathbf{V}\mathbf{v})^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{V}\mathbf{v}). \quad (2)$$

The root square background-error covariance matrix is decomposed into a sequence of linear operator  $\mathbf{V} = \mathbf{V}_h\mathbf{V}_v$ , which separately accounts for horizontal correlations and vertical covariances. The operator  $\mathbf{V}_v$  consists of 10-mode bivariate vertical EOFs from the eigen decomposition of monthly anomalies with respect to the monthly climatology from a previous simulation, as in Bellucci et al. (2007). The operator  $\mathbf{V}_h$  acts on the vertical profiles of temperature and salinity, and models horizontal Gaussian covariances by means of an application of a first-order recursive filter, on an extended domain with duplicated fields and observations in the extension zone to ensure cyclic conditions to the boundaries of the global domain. More details of the formulation are given by Dobricic and Pinardi (2008), while the global ocean implementation can be found in Storto et al. (2011). Details of the recursive filter implementation are given in the appendix. Note that here the focus is on the estimation and assessment of nonuniform background-error correlations, which may be achieved also with other formulations of the horizontal correlations. For instance, Weaver and Mirouze (2013) developed an extension of diffusion operators for the case of inhomogeneous length scales, but did not indicate a practical method to estimate them.

The observational data consist of expandable bathythermographs (XBT); buoy reports (BUOY); sea-station reports [CTD or temperature, salinity, current report (TESAC)]; Argo floats, all of them from the quality-checked EN3 dataset (Ingleby and Huddleston 2007); and along-track sea level anomaly (SLA) observations calibrated and distributed by Collecte Localisation Satellites (CLS) and Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO; Le Traon et al. 1998). Prior to the 3DVAR minimization, the data assimilation system also performs (i) a climatology check that rejects observations whose departure from the *World Ocean Atlas* monthly climatology (WOA2009; Locarnini et al. 2010; Antonov et al. 2010) is greater than a certain threshold (currently 9.0 K and 3.0 psu for temperature and salinity observations, respectively); (ii) a background-quality check to reject observations whose square departure from model equivalents is too large (namely, thrice the sum of the observational and background-error variances); and (iii) a horizontal data thinning procedure for sea level anomaly observations to remove reports too close to each other, provided that SLA observational errors are assumed to be spatially uncorrelated. In this

analysis setup, only the temperature and salinity states are updated on a weekly basis by the data assimilation.

The analysis system also comprises a nudging scheme for relaxing sea surface temperature (SST) to daily SST observations from the Advanced Very High Resolution Radiometer (AVHRR) and the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) instruments mapped onto a regular grid (Reynolds et al. 2007) and a nudging scheme that assimilates sea ice concentration data from the dataset of Cavalieri et al. (1999).

### 3. Estimation of horizontal correlation length scales

#### a. Error dataset for background-error correlation length scales' estimation

Estimating the background-error correlation length scales requires a dataset of error realizations, from which, in general, covariances can be calculated. The choice of this dataset is not trivial, because the *true errors* are in practice unknown. Several strategies for representing forecast errors for background-error covariances estimation can be found in the data assimilation literature. Zhou et al. (2004) review some strategies for calculating HCLSs. All of them are derived from observational dataset, and, in particular, from the tropical Pacific mooring array. However, for a three-dimensional temperature and salinity global ocean estimation, the recourse to observational data to estimate HCLSs appears improper, as the observational coverage is strongly irregular (and often missing) on both the horizontal and vertical domains, and it strongly differs among the ocean parameters. Therefore, it seems more appropriate to rely on the model space-based dataset. We also argue that, especially near the surface, differences in the spatial variability of freshwater (runoff, precipitation) and air-sea heat fluxes should be taken into account. It can be achieved by calculating separately the correlation length scales for temperature and salinity. This choice, however, appears less crucial in the deep ocean, dominated by buoyancy-driven variability.

Commonly adopted approaches in ocean data assimilation either represent forecast errors through anomalies with respect to a long model simulation, called climatological anomalies (e.g., Bellucci et al. 2007), or mimic these errors through differences between ocean forecasts valid at the same time but initialized at different times, the so-called National Meteorological Center (NMC) method (Parrish and Derber 1992). Typically, the forecast initialized earlier has lower quality, and the differences between the two forecasts may reproduce the quality difference

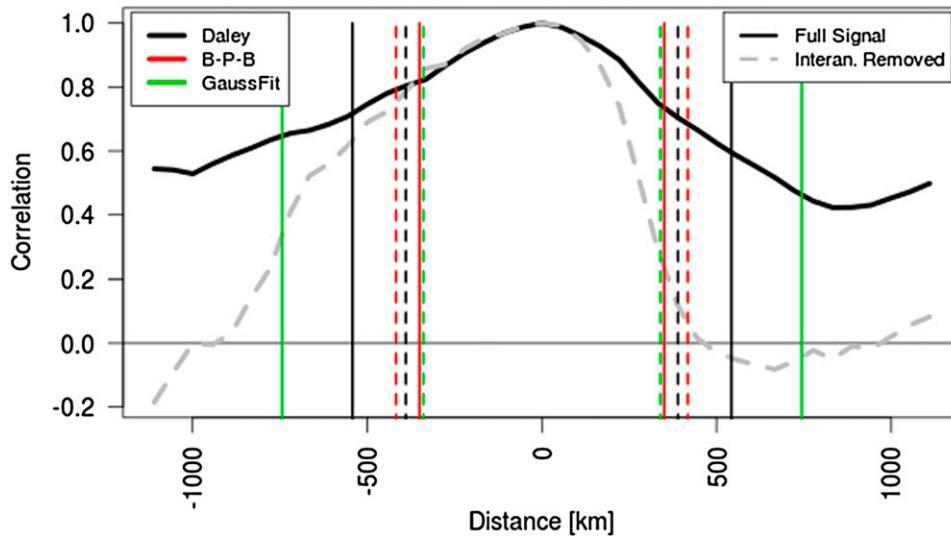


FIG. 1. Correlation vs distance plot for a model grid point in the tropical Pacific Ocean from a dataset of climatological anomalies. Correlations are shown for both the full (black solid line) and the detrended signal (gray dashed line). Also shown is the correlation length scale corresponding to different estimation formulas: Daley's formulation, the BPB, and the Gaussian fit. For these, the correlation length scale is plotted with full (solid) and detrended (dashed) signal.

because of both an additional assimilation step and longer model integration. The main advantage of the former is that it does not rely at all on the observing network coverage. However, such a dataset tends by construction to depict the ocean variability, which does not necessarily correspond to the ocean forecast errors. On the other hand, the NMC method may succeed in representing forecast errors, while the resulting background-error covariances may be affected by the observing network coverage (Berre et al. 2006). For example, in data-sparse areas, NMC-derived background errors may be underestimated, because the successive initialization does not impact the background fields over those areas.

In the remainder of this article, we will use climatological anomalies (CA method) and the NMC methods for the correlation estimations. Other strategies, such as the derivation of the background-error covariances from ensemble simulations (Belo-Pereira and Berre 2006, hereinafter BPB; Storto and Randriamampianina 2010; Storto et al. 2013), will be investigated in future studies.

The dataset of errors was extracted from a previous data assimilation experiment relative to the period 1993–2010. This experiment had a slightly different setup for the ocean model and the data assimilation systems than the one presented in section 2, and it used a globally constant value for the correlation length scales, equal for temperature and salinity. For the CA method, the error dataset is formed by the monthly means, from which the monthly long-term average was subtracted. For the NMC method, the dataset is formed by the differences of

forecast valid at the same time but initialized either 7 or 14 days before, consistent with the weekly frequency of the analysis scheme. In both cases, the error fields were grouped by season, in order to separately estimate a correlation for each of the four seasons.

A further preprocessing strategy applied to the error fields is their detrending: low-frequency signals (e.g., interannual) may in fact artificially overestimate the correlation length scales. This is of less importance in the case of NMC statistics, although large-scale biases of the atmospheric forcing or model parameterizations might affect the statistics. We therefore remove these signals, since the time scales typical of the analysis corrections are much shorter (weekly), and such low-frequency signals must not be accounted for during the analysis step. An example is provided in Fig. 1, which shows a correlation–distance curve for a tropical Pacific Ocean grid point, with (solid black line) and without (dashed gray line) the removal of the trend when CA statistics are used. Clearly, the correlation shape coming from the full signal is much broader and does not vanish for long distances, because low-frequency signals (e.g., El Niño–Southern Oscillation) may affect the correlation estimation over these areas. Simple detrending filters out these signals and narrows the length scales.

#### *b. Estimation formulas for correlation length scales*

In this section we detail the strategy for the nonuniform HCLSs estimation. The recursive filter previously introduced is able to mimic Gaussian correlations in a very

efficient way. When the shape of horizontal correlations is assumed to be symmetric, a measure of the correlation length scale can be obtained from the second derivative of the correlation function ( $\rho$ ) along the point by using Daley's correlation length scale Daley (1991, p. 110):

$$L_d^2 = -\frac{1}{\nabla^2 \rho(0)}. \quad (3)$$

Recently, BPB proved that under the hypothesis of local homogeneity and differentiability assumptions, such a formulation can be approximated along a generic direction  $d$  as

$$L_d^2 = \frac{E[(\varepsilon - \bar{\varepsilon})^2]}{E[(\partial_d \varepsilon - \bar{\partial}_d \varepsilon)^2] - \partial_d [E[(\varepsilon - \bar{\varepsilon})^2]]}, \quad (4)$$

where  $\partial_d$  is the spatial derivative along the generic  $d$  direction,  $E$  is the expected value operator,  $\varepsilon$  is the error vector, and  $\bar{\varepsilon}$  is its mean. Pannekoucke et al. (2008) suggested also that such a derivation can be used for any arbitrary direction. The reader is referred to BPB for the derivation of Eq. (4). Assuming that the horizontal correlation has a Gaussian shape, correlation length scales can also be derived by calculating the least squares fit of the Gaussian correlation function as a function of the separation distance (Barker et al. 2004) or by further simplification to Daley's correlation length scale (Pannekoucke et al. 2008). Pannekoucke et al. (2008) also report a discussion on different formulations to estimate the correlation length scales.

To exemplify the use of the formulas previously introduced, we report in Fig. 1 the correlation length scales for a tropical Pacific Ocean grid point using Daley's formula, the approximation of Eq. (4), and the Gaussian fit. This exercise is repeated also when the signal is detrended, and it provides the input for some considerations: Daley's formula is more sensitive to the signal removal procedure than the BPB formula, since it relies on the second derivative of the correlations, while the latter on the first derivative of the error fields. It also generally provides more spatial variability of the correlation length scales. In the case of signal removal, the two solutions are generally very similar. The BPB formula is in general very stable. Daley's correlation formula generally provides results very similar to those of Eq. (4), although the resulting correlation length scales appear noisier (not shown). In its basic methodology, the Gaussian fit may inaccurately estimate distribution tails, because remote correlations may be noisy. On the other hand, a more sophisticated application of the Gaussian fit at each model point that accounts for the remote noise could become computationally very expensive. As the BPB formula and the Daley's formula give very similar

estimates and the BPB formula is much computationally faster than Daley's formula, in experiments we will apply only the BPB formulation and the Gaussian fit procedure for the actual estimation of the HCLSs. To satisfy the symmetry assumption for the correlation length scales, for each grid point the correlation length scale is the arithmetic mean of the two values to the right and left of the grid point. The correlations thus computed are successively low-pass filtered to avoid too large local gradients of the correlation length scales. This filtering allows the correlation length scales to have gradients comparable to those of the meridional grid resolution and respects the assumption of local homogeneity of Eq. (4), by letting the correlations vary slowly and smoothly in space (Weaver and Mirouze 2013).

### c. Results of the estimation

Results of the correlation length scales' estimation are summarized in this section. In Fig. 2 we show the zonal HCLSs averaged in the first 100 m of depth for temperature (left panels) and salinity (right panels) from the CA and NMC methods (top and middle panels, respectively) using the BPB estimation, and from the CA method using the Gaussian fit. Mesoscale areas such as the Kuroshio, the Gulf Stream, the Agulhas Current, and the Antarctic Circumpolar Current (ACC) regions are characterized by short correlation length scales, generally below 100 km, for all parameters and methods. While all the methods show the signature of the tropical circulation along with the large correlations in correspondence to the Arctic and Antarctic subpolar currents, they importantly differ in absolute values and in the physical process, yielding the longest correlation length scales. The CA method, relying on the natural variability of the ocean state rather than the forecast errors, emphasizes the length of the correlations in proximity to the equator. The NMC generally exhibits larger correlations, with maximum values in correspondence of the South Pacific and subpolar gyres, as reported also by Zhou et al. (2004), who used innovation-based computations. This may indicate that correlation length scales are effectively very long in such an upwelling region of the Pacific Ocean. The use of the Gaussian fit procedure instead of the BPB formula in practice amplifies the length scales without impacting significantly the spatial patterns. The ratio between the correlation length scales and the spatial resolution (see the appendix) has the minimum with values less than 2 in the Gulf Stream, the Kuroshio, and the ACC, and the maximum with values greater than 5 in the equatorial region.

It is also possible to note very similar (neutral to slightly longer) correlation length scales for salinity with respect to temperature in the case of CA correlations. On the other

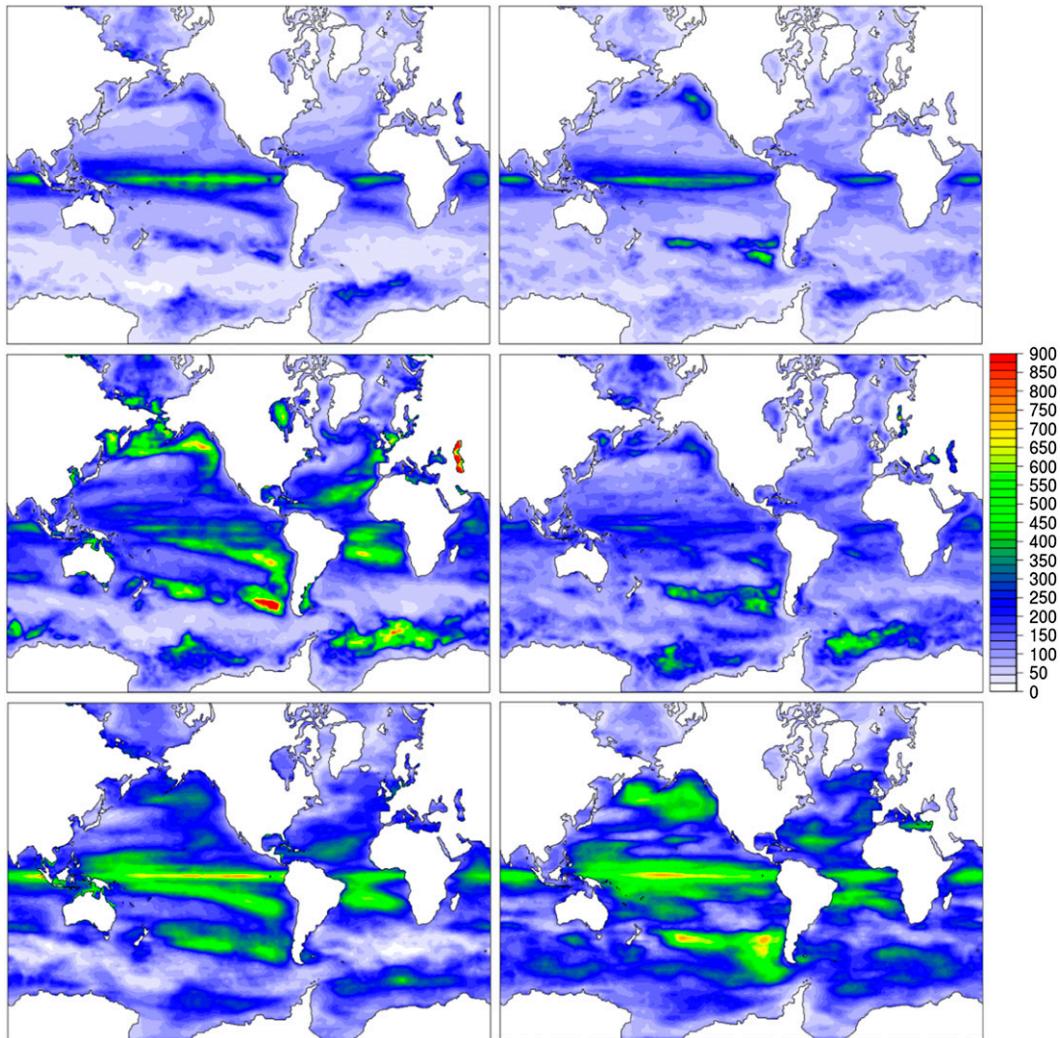


FIG. 2. Vertically averaged 0–100-m zonal correlation length scales (km) for (left) temperature and (right) salinity using: (top) the CA method and the BPB formula as explained in the text; (middle) using the NMC method and the BPB formula; and (bottom) using the CA method and the Gaussian fit.

hand, salinity correlations in the NMC method appear shorter than temperature ones. This feature seems to be mainly affected by a smaller number of salinity observations, and not by the dynamics of the system.

The correlation length scales for deeper layers are reported in Fig. 3, which in particular shows the zonal length scales averaged between 100 and 800 m of depth. For the temperature length scales, all the methods agree on presenting patterns very similar to those in the first 100 m of depth, although the length scales are shorter. On the other hand, while still showing similar patterns near the surface, salinity length scales do not decrease that much with depth, suggesting that its spatial variability is rather uniform throughout the first 800 m of depth.

Figure 4 depicts the 0–100-m averages of temperature and salinity ratio between zonal and meridional

correlation length scales for the same HCLSs computation of Fig. 2. The figure thus shows the anisotropy of the correlation length scale fields. Close to the equator, zonal correlation scales are 2 times longer than the meridional scales. This is in accordance with similar results given by Meyers et al. (1991) and Carton et al. (2000), who also report how meridional correlations do not change as much as the zonal correlations across the equator. Values at the equator are close to 2.28, as suggested by Derber and Rosati (1989). Note also that maps of the ratios reveal the importance of the anisotropic formulation of the HCLSs over coastal areas, as horizontal length-scale correlations perpendicular to the shorelines systematically are much shorter than those parallel. At the northern boundary of the Antarctic Circumpolar Current, as well as in the Pacific subpolar

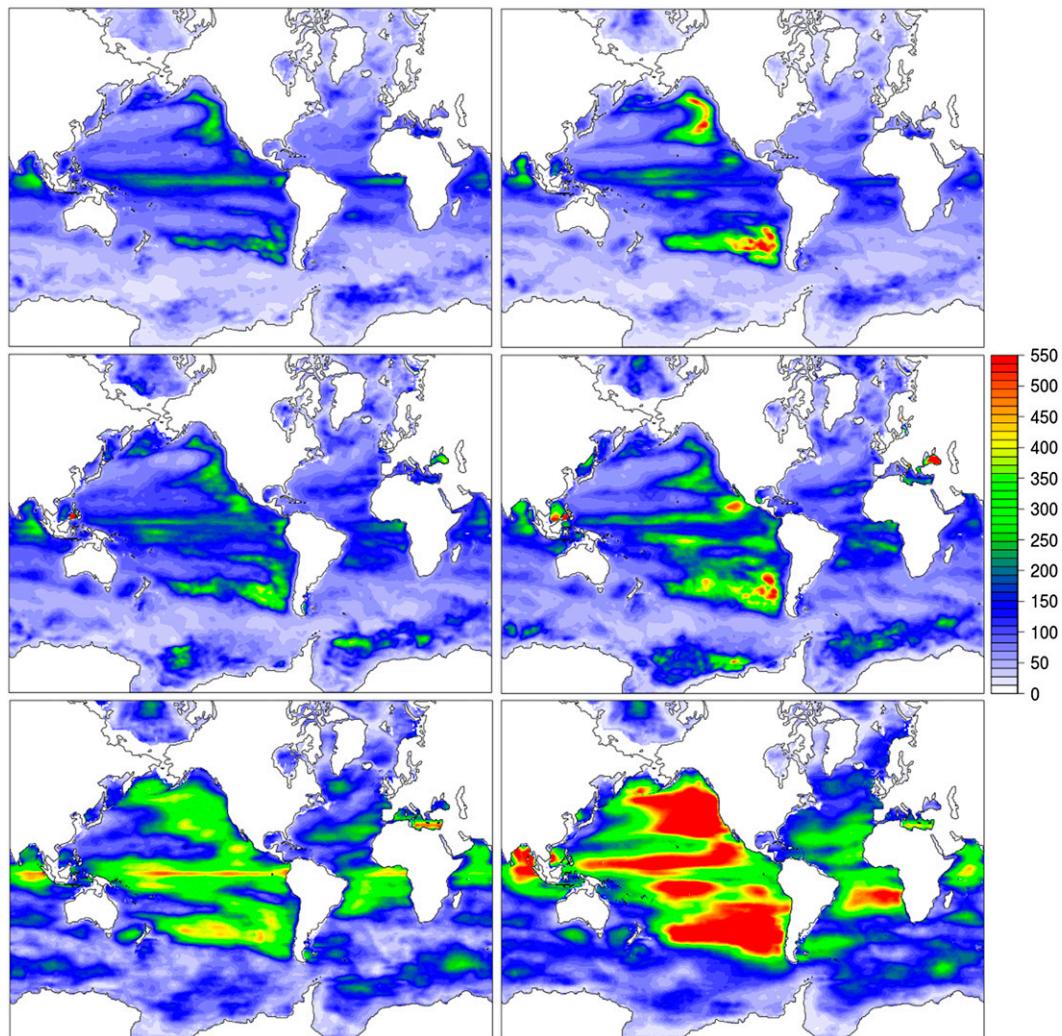


FIG. 3. As in Fig. 2, but for the vertically averaged 100–800-m zonal correlation length scales. Note that contour intervals and color scales are different from Fig. 2.

gyre, meridional correlations last longer than the zonal ones, except for the Gaussian fit computation. This may be explained by the fact that over these regions there are strong zonal currents. Thus, any displacement of the polar fronts will lead to meridional correlations length scales longer than the zonal ones. On the other hand, the Gaussian fit procedure partly smoothes out this feature, as it accounts for the tail of the correlation distribution more than Eq. (4). Interestingly, the CA and NMC methods are very consistent in representing the spatial distribution of the ratio, when the BPB formula is adopted. The differences between the ratios are also almost identical for temperature and salinity, suggesting that they are explained only by the ocean circulation patterns. The ratio reduces toward the unit with depth (not shown), being much closer to one below 300 m and almost one below 1200 m of depth.

Temperature correlation length scales generally decrease with the depth, and have a nearly symmetric shape with the largest values at the equator. This is shown in Fig. 5 for winter and summer zonal averages. Note also local maxima at around 30°S and 30°N, in correspondence of the centers of the subtropical gyres. While we have shown that surface correlations respond to large-scale atmospheric patterns, such as the intertropical and South Pacific convergence zones, it is not clear whether the shortening of the correlations with depth is real or an artifact of depths dominated by low-frequency variability, which may in turn be expected to have long spatial scales. This feature may be explained by the weakening of the atmospheric forcing and is probably linked to the methods (CA and NMC), which primarily rely on the model or analysis increments' variability, respectively, thus underestimating the correlations

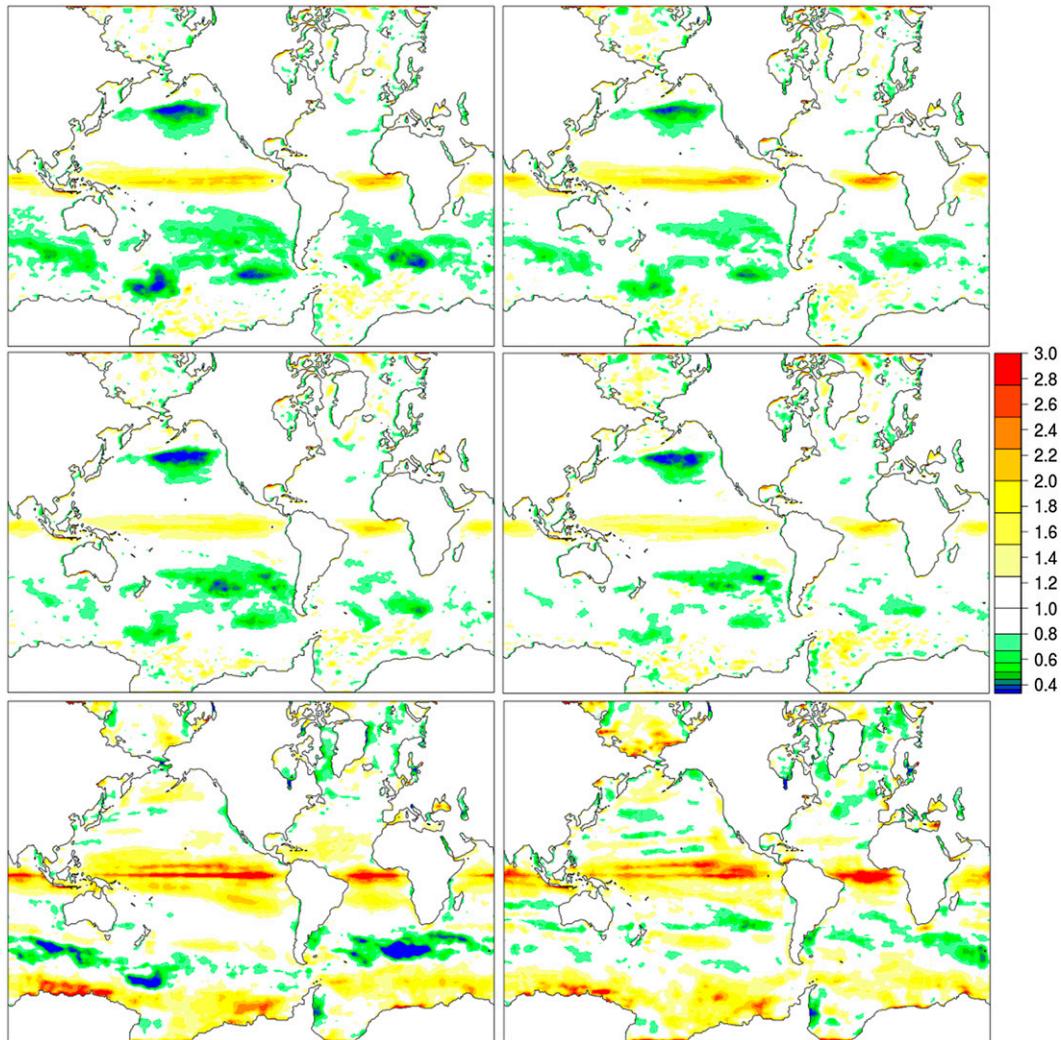


FIG. 4. As in Fig. 2, but for the vertically averaged 0–100-m ratio of zonal and meridional correlation length scales.

in the deep and bottom waters. Although this feature might be avoided by means of for example, vertical normalization of the correlation length scales, this drawback does not seem critical for data assimilation applications, provided that there are only a very few observations below 1200 m of depth, and the definition of background-error covariances below this depth are unimportant.

Another interesting aspect in Fig. 5 is the seasonality of the near-surface tropical correlation length scales (i.e., 0–30 m of depth), which is mainly due to the northward shift of the intertropical convergence zone (ITCZ) from winter to summer, confirming the importance of the seasonal dependence of the correlations, especially for the tropical areas.

To appreciate the impact of the spatially varying correlation length scales, we performed idealized single-observation data assimilation experiments, where the use

of the horizontal correlations, calculated as previously described, is compared to the use of uniform correlation length scales, here computed as the global averages of the nonuniform correlations. These tests (not shown) indicate that the use of nonuniform correlation scales in areas of strong mesoscale activity produces a much narrower analysis correction with respect to the uniform case. The magnitude of the correction is very large for both correlation cases, as a result of the large background-error standard deviations typical of these regions. Therefore, the analysis correction remains large in absolute value but it is very localized. On the contrary, areas with longer correlations (e.g., in the tropical Pacific Ocean) are characterized by much broader but smaller analysis increments, as a consequence of longer correlation length scales and small background-error variances.

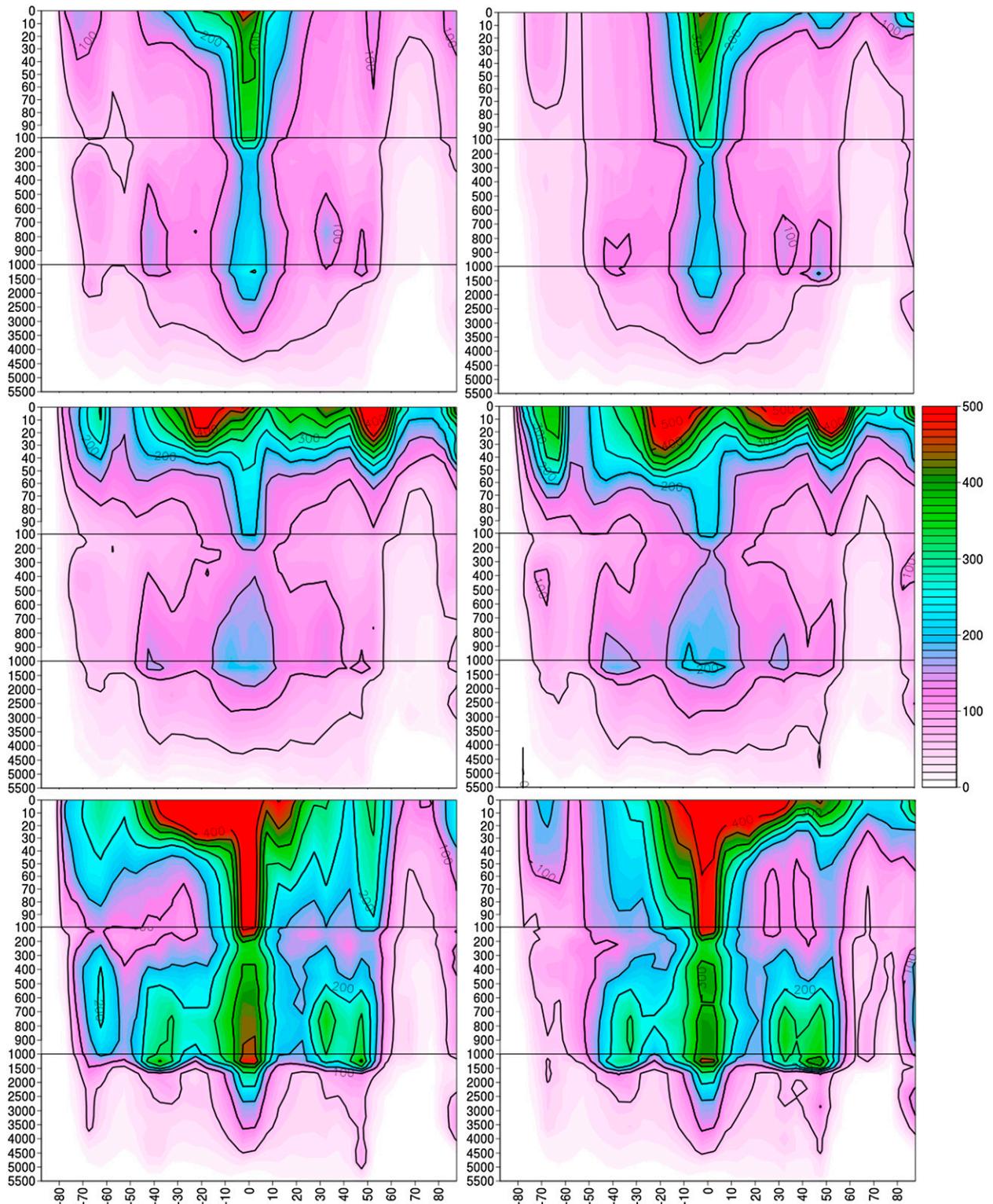


FIG. 5. Zonal averages of zonal temperature correlation length scale (km) as a function of depth for (left) winter and (right) summer time statistics. Length scales from the (top) the CA method and the BPB formula, (middle) NMC method and the BPB formula, and (bottom) the CA method and the Gaussian fit.

TABLE 1. List of experiments performed with associated features of the horizontal correlation length scales used.

Expt name	Correlation type	Error dataset	Estimation formula
CTR	No assimilation	—	—
CR1	3D nonuniform	CA statistics	BPB
CR2	3D nonuniform	NMC statistics	BPB
CR3	3D nonuniform	CA statistics	Gaussian fit
CN1	Vertical global averages	CA statistics	BPB
CN2	Vertical global averages	NMC statistics	BPB
CN3	Vertical global averages	CA statistics	Gaussian fit

#### 4. Impact of nonuniform correlation length scales

In this section we evaluate the impact of nonuniform correlation length scales with respect to the experiments with horizontally uniform correlation length scales. The list of experiments is reported in Table 1. The experiments with uniform correlation length scales (CN1, CN2, and CN3) use, as a uniform value for each model level, the global spatial average from the experiments with nonuniform correlation length scales (CR1, CR2, and CR3, respectively). Note that the length scales vary vertically in CN1–CN3.

Figure 6 reproduces the globally averaged correlation length scales used in the experiments CN1–CN3. The experiments are designed to investigate (i) the impact of nonuniform versus uniform correlation length scales, (ii) the impact of the different statistics considered (CA vs NMC), and (iii) the different estimation methods [Eq. (4) vs Gaussian fit]. We did not combine the Gaussian fit methodology with the NMC statistics, since we found the Gaussian fit amplifies the HCLSs values without changing the patterns. Furthermore, the impact of the tuning of

a global value of the correlation length scale can be also appreciated through the relative comparison between CN1, CN2, and CN3. From Fig. 6 it is easy to see that global averages do not show any strong seasonal dependence. The temperature NMC statistics lead to correlation length scales approximately 2 times longer than CA statistics in the first 20–50 m of depth, while they are very similar below. This is related to the fact that the small-scale spatial variability borne by the air–sea heat fluxes in the CA method is not felt by the NMC method, which relies on different initializations in order to reproduce forecast errors. The figure also confirms that the salinity length scales do not decrease with depth, unlike the temperature length scales. A local peak in the globally averaged length scales is often visible at around 1000 m of depth. At such a depth, long correlations are found in correspondence of the eastern tropical Pacific Ocean (not shown) and respond more to numerical effects of the HCLSs computation (e.g., ratio between variance and spatial variability) rather than to physical processes.

The 12-yr experiments start on the 1 January 2000 and last until the end of 2011, thus allowing an evaluation of

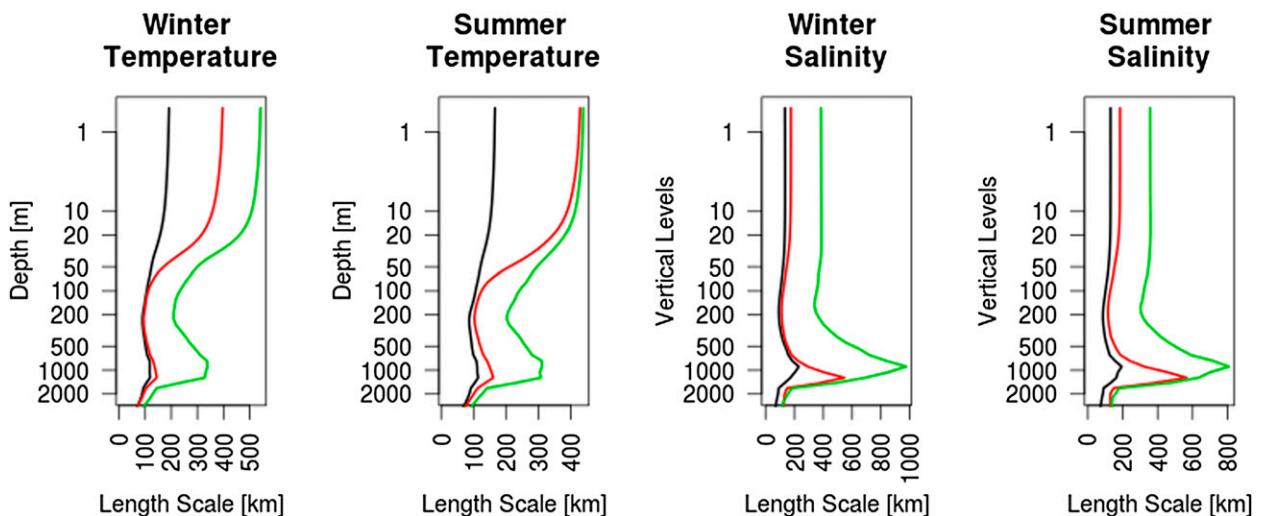


FIG. 6. Global averages of correlation length scales from CA statistics (black lines) and NMC statistics (red lines) using the BPB formula, and CA statistics using the Gaussian fit (green lines) for temperature and salinity in winter and summer. The depth axis is logarithmic. These global profiles are also used in the CN1–CN3 experiments with uniform correlation length scales, as explained in the text.

TABLE 2. Performance statistics of the 3DVAR assimilation system. The table shows the Euclidean Norm of the cost function gradient at the first iteration averaged over the simulation period, the percentage reduction of the Euclidean cost function Norm with respect to the first iteration at 10, 20, and 25 iterations and the number of iterations required by the minimizer to converge averaged over the simulation period.

Expt name	Avg Norm of $J$ gradient (1st iteration)	Percentage reduction of $J$ gradient Norm (10th, 20th, 25th iteration)	Avg No. of iterations
CR1	$1.146 \times 10^{+6}$	5.84, 1.15, 0.41	24.3
CR2	$1.296 \times 10^{+6}$	9.74, 1.14, 0.85	26.0
CR3	$1.591 \times 10^{+6}$	7.18, 1.42, 0.96	28.4
CN1	$1.382 \times 10^{+6}$	5.41, 0.92, 1.30	24.4
CN2	$1.468 \times 10^{+6}$	5.97, 1.44, 1.34	25.9
CN3	$1.676 \times 10^{+6}$	5.27, 1.02, 1.13	27.9

the analysis system performances with full deployment of Argo floats. All the experiments are initialized by means of the same initial conditions, valid at 1 January 2000 from a previous data assimilation experiment.

#### a. Performance of the data assimilation system

As a first result, we review here the performance of the data assimilation system. Table 2 reports for each experiment, the Norm of the cost function gradient averaged over the simulation period for the first minimizer iteration and as a relative percentage reduction at the 10th, 20th, and 25th iterations with respect to the first iteration. The Norm of cost function gradient provides valuable information on both the minimization speed and the accuracy of the 3DVAR solution. The gradient Norm it is smaller at both the first and last iterations (represented here by the 25th iteration for simplicity) in the case of nonuniform correlation length scales, suggesting that they lead to a better accuracy of the 3DVAR analysis. The gradient cost function reduction at the end of the minimization (represented here by the 25th iteration) is more effective when correlations are inhomogeneous. Note also that the Norm in the experiments CN1 and CN3 does not decrease from iteration 20 to 25, while it decreases only very slightly for experiment CN2, implying that the convergence of the minimizer for the last iterations is prevented by the uniform correlation length scales. This confirms the importance of using short correlations in areas of large spatial gradients, the denial of which causes the 3DVAR minimizer to suboptimally reach the minimum of the cost function, given the larger smoothing in these areas.

The average number of 3DVAR iterations needed by the minimizer to reach the minimization criterion is also reported in Table 2. The use of uniform or nonuniform correlation length scales does not lead to appreciable differences (values of CR1 vs CN1, CR2 vs CN2, and CR3 vs CN3), although with the increase of the correlations (e.g., from CN1 to CN2 and CN3), there is an increase in the average number of iterations, as longer

correlations prevent the minimizer to reach a quick convergence.

To better understand the impact of the nonuniform correlation length scales, we study the different analysis increments' standard deviation between CN1 and CR1 throughout the experimental period. This is reported in Fig. 7 in terms of zonal averages as a function of depth, for the temperature analysis increments only. The most relevant effect of using nonuniform HCLSs is visible at the midlatitudes, around 40°S and 40°N, where their use decreases the analysis increments' standard deviation. This decrease is found in correspondence of the areas with the largest mesoscale activity (not shown), that is, the Gulf Stream and the Kuroshio regions in the Northern Hemisphere, and in correspondence of the Agulhas and Falkland Currents in the Southern Hemisphere. In these regions, the use of uniform HCLSs has the effect of low-pass filtering the analysis increments as a result of the overestimated correlation length scales, spreading out and increasing the data assimilation corrections. As a consequence, the variability borne by the data assimilation increases in those regions, since the background fields get farther from the observations. Similar conclusions were found also for the salinity analysis increments.

#### b. Verification

We present in this section the main verification results from the experiments. Figure 8 shows, for a number of selected regions, the percentage root-mean-square error (RMSE) gain (RMSG) for all the in situ and altimetry observations with respect to the control (CTR) experiment, defined as

$$\text{RMSG} = 100 \frac{\text{RMSE}_{\text{CTR}} - \text{RMSE}_K}{\text{RMSE}_{\text{CTR}}}, \quad (5)$$

where  $K$  is the  $k$ th experiment of Table 1. The figure also reports the same statistics for a number of pairs of experiments, and in particular the RMSG of CR1 with respect to CN1, of CR2 with respect to CN2, of CR3 with respect of

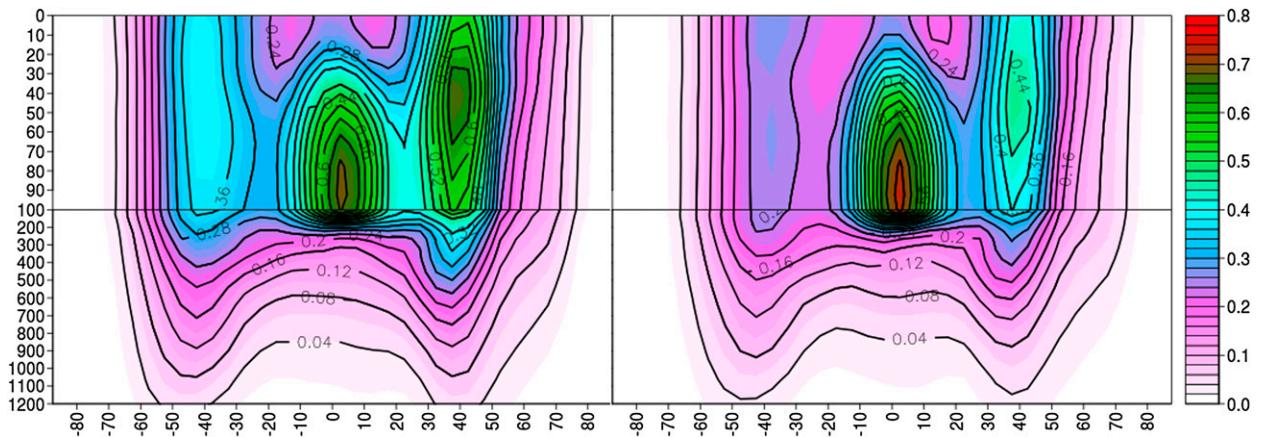


FIG. 7. Zonal averages of the standard deviation of the temperature analysis increments ( $^{\circ}\text{C}$ ) as a function of depth for the experiments (left) CN1 and (right) CR1.

CN3, and of CR1 with respect to CR2, calculated in the same way as Eq. (5). This allows us to appreciate the impact of both the nonuniform correlation length scales with respect to the uniform case and that of different correlation length-scale estimations. All the available observations are used without any quality check, in order to have an independent dataset of verifying observations.

For the temperature statistics, the CR1 RMSG with respect to the CTR experiment is significantly larger in all areas, except in the ACC region, where CR2 performs slightly better, probably because of longer correlations in close proximity to the Antarctic continent. For all the correlation estimation methods, the use of nonuniform length scales always improves the RMSE results. CN3, which has uniform and longer correlation length scales, exhibits the worst scores. Note that the gain with respect to the CTR experiment in the ACC is smaller than in other areas, because of the small amount of verifying observations. The largest improvements brought by the assimilation are in the tropical region, where the non-assimilation.

Experiment is found not to be able to correctly represent the thermocline depth and its east–west variations (not shown). The gain of the nonuniform HCLSs with respect to the respective uniform case is between 4% and 8% for the global ocean, with peaks in the Kuroshio Extension between 7% and 13%. Note that the gain is higher when CR3 is compared to CN3, as this latter experiment has low skill scores, which are therefore easy to beat. The differences in RMSE between CR1 and CR2 are generally very small and below 1% for all the areas.

The salinity scores show qualitatively similar results, except for the tropical areas. There the effects of using nonuniform length scales are rather negligible when the BPB formula is adopted (CR1 and CR2), while the CN3 experiment presents again the worst scores. Unlike the temperature and sea level results (below), the salinity

RMSE decrease with respect to the CTR experiment is larger in the Kuroshio and Gulf Stream regions than in other areas. This indicates that in these areas, probably because of the large uncertainty of the precipitation fluxes from the atmospheric reanalysis, the assimilation plays a crucial role in correcting the inaccuracies of the freshwater content evolution. The gain borne by the use of nonuniform HCLSs is between 2% and 8% for the global RMSE, peaking in the Gulf Stream region with values between 4% and 12%.

The large RMSG of the experiments with nonuniform HCLSs, with respect to those with uniform HCLSs, is even more visible in the verification against altimetry data. This is an important result, as sea level anomalies represent the proxy for the DEPTH-integrated density variations. The impact of the data assimilation is noticeable in the Tropical regions of the Atlantic and Pacific Oceans, and less effective in the ACC and in the Kuroshio and Gulf Stream regions. The experiment CR1 leads to better skill scores on all the regions investigated. In the Gulf Stream and Kuroshio regions, the RMSG of CN3 with respect to CTR is very small (about 5%), suggesting that a data assimilation system with largely overestimated correlation length scales is able to correct only very weakly the column-integrated seawater density. The gain of CR1 with respect to CN1 and of CR2 with respect to CN2 is high not only in the Kuroshio Extension (from 7% to 8%) and in the Gulf Stream (from 5% to 7%) but also in the tropical Pacific (from 5% to 7%), suggesting that the new formulation of the correlation length scales is particularly beneficial to the altimetry data assimilation even in tropical areas, as a result of the dependence of the sea level anomalies on the water-column-integrated density anomalies. Again, the impact of using a different error dataset (CR1 vs CR2) is found to be small although significant, while the use of the

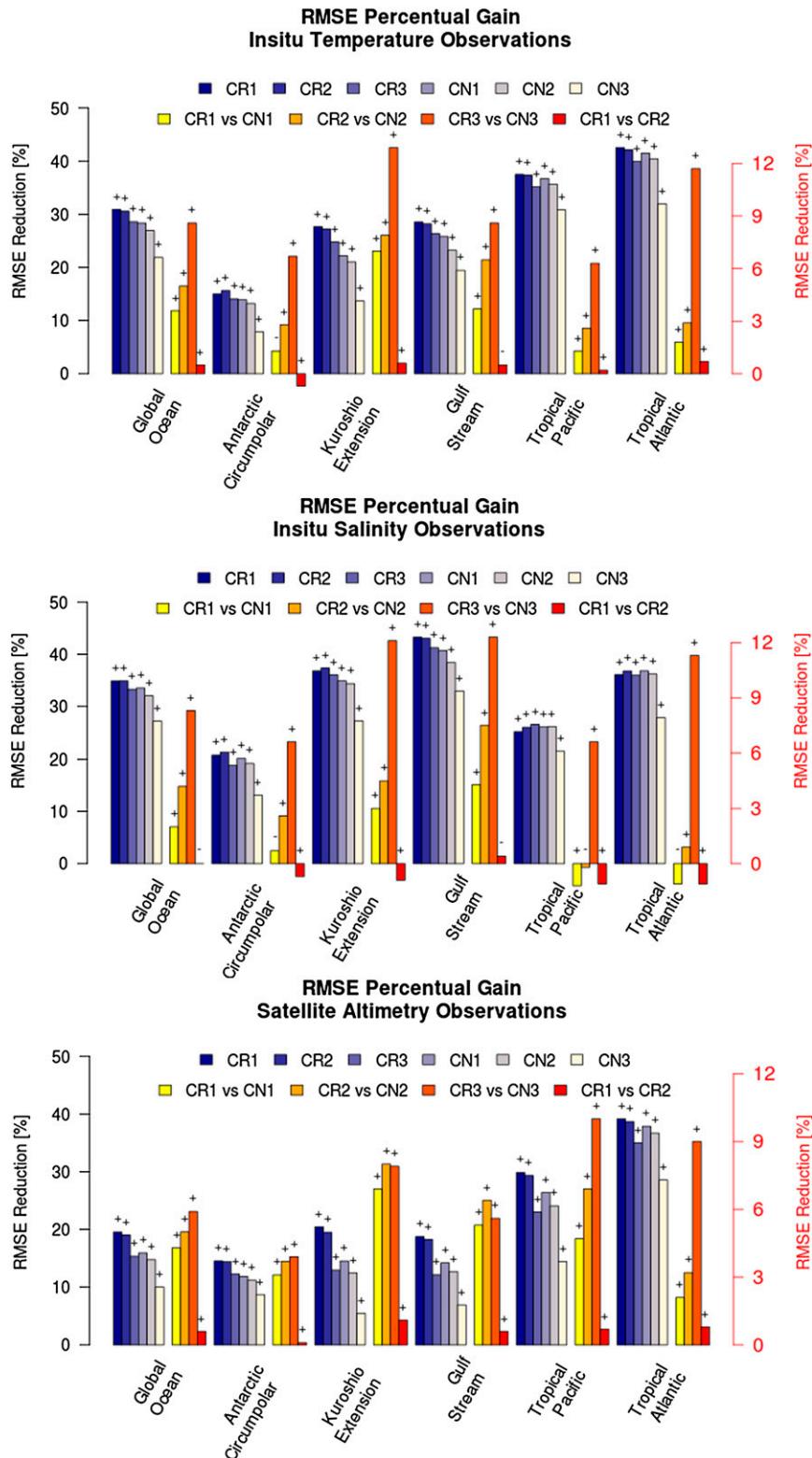


FIG. 8. Percentage gain (RMSE reduction) for misfits of (top) in situ temperature observations, (middle) in situ salinity observations, and (bottom) altimetry observations for a number of regions in the ocean. The reduction is calculated with respect to the non-assimilation experiment for all the experiments (blue to white bars, black left-side axis) and between several pairs of experiments (yellow to red bars, red right-side axis). The plus (minus) signs on top of the bars indicates that the RMSE difference is significant (nonsignificant) according to a 95% confidence level two-sided paired *t* test performed over the RMSE series.

Gaussian fit (CR3) is found to be detrimental with respect to CR1.

The RMSE profiles of temperature and salinity observations are shown in Fig. 9, only for the global ocean and the Kuroshio Extension cases. The global profile of temperature RMSE suggests a small impact of the correlation length scale near the surface (from 0 to about 20 m of depth). Below this depth, the RMSE increases because of the misplacement of the mixed layer thickness, and the impact appears more visible, with smaller RMSE from the experiments with nonuniform HCLSs, and CR1 showing the smallest values for the RMSE. The impact on the salinity is more appreciable in the first 50 m, where the RMSE error is larger, as a consequence of the larger air–sea freshwater fluxes’ variability, and experiment CR1 leads to better scores, followed by the other experiments with nonuniform HCLSs. In the Kuroshio region, the improvement borne by the nonuniform HCLSs is visible throughout the water column, with the better skill scores always provided by CR1, for both temperature and salinity misfits. Note that the RMSE for CN1 and CN2 is identical in practice, indicating that a small tuning (see Fig. 6) of the overestimated correlation length scales has no impact. Similar results apply also to the Gulf Stream region (not shown).

The sea level anomaly observing network provides rather uniform observational coverage. It is therefore possible to map the misfits onto a regular grid and compute their root-mean-square error. Figure 10 shows the differences between the RMSE of experiments CTR, CN1, CR2, and CR3 minus the RMSE of experiment CR1, used here as a reference. Note that the palettes and the contour ranges differ among the four figure panels, in order to better appreciate the RMSE differences. This allows a detailed spatial investigation of the impacts of the different correlation length scales. The first panel shows the impact of data assimilation with a positive RMSE decrease over the entire global ocean with peaks (up to 10 cm) corresponding to the North Atlantic (Gulf Stream region and subpolar gyre) and also within the ACC. The second panel shows the impact of the nonuniform HCLSs (CR1) with respect to the uniform HCLSs (CN1). The improvements cover almost all the global ocean and exhibit the largest values in correspondence of the areas of shorter correlation length scales (up to 3 cm), suggesting once again the importance of nonuniform HCLSs in those areas. The differences in RMSE between CR1 and CR2 exhibit a rather noisy behavior. Generally, a slightly positive impact of the NMC-derived correlations is found in the tropical region, while the CA-derived correlations seem to perform better in the subtropical regions. In conclusion, the use of the Gaussian fit that leads to longer correlations is found to be detrimental almost everywhere.

Note that the correlations used in CR1 in the equatorial Pacific are longer than the global average used in CN1 and the ones used in CR2 (Figs. 2 and 6). The RMSE of CR1 in that region is slightly greater than that of CN1 and CR2, suggesting the CA methods overestimate the correlations length scale in the equatorial region, relying on the natural ocean variability of this region.

### c. Impact on the ocean variability

We briefly review here the impact of the nonuniform HCLSs on the mean ocean variability. A confirmation of their positive impact is visible in the analysis of the eddy kinetic energy (EKE). Figure 11 displays the zonal averages of the EKE for the experiments presented above, along with the EKE from the Ocean Surface Currents Analyses–Real Time (OSCAR; Bonjean and Lagerloef 2002), which merges satellite altimetry and scatterometers to produce ocean surface velocities. For all the dataset, we define EKE as

$$\text{EKE} = \frac{1}{2}(\langle u^2 \rangle - \langle u \rangle^2 + \langle v^2 \rangle - \langle v \rangle^2), \quad (6)$$

where  $u$  and  $v$  are the monthly-mean ocean velocity components from either the OSCAR velocities or the experiments (taken at 15 m of depth), respectively; and the operator  $\langle \cdot \rangle$  is the 2000–11 average operator. In Fig. 11, the lines that correspond to the experiments are below the OSCAR dataset at all latitudes. This underestimation with respect to OSCAR may be primarily due to the horizontal resolution, which is  $1/2^\circ$  for the model and  $1/3^\circ$  for OSCAR. However, while the different HCLSs lead to an almost identical EKE near the equator, the mid-latitudes, dominated by the western boundary currents and the Gulf Stream and Kuroshio regions in the Northern Hemisphere (around  $40^\circ\text{N}$ ) and several eddy-dominated regions in the Southern Ocean, exhibit an increase in EKE in CR1 and CR2. Note that the EKE coming from the non-assimilation experiment is largely below all the other experiments, while using long uniform HCLSs (CN3) only partially increase the EKE. The data assimilation partially overcomes the relatively coarse resolution. The different behavior of the near-surface eddy kinetic energy suggests again that the use of uniform HCLSs prevents the midlatitude from developing their natural variability, and erroneously filters out the large spatial variability of the ocean circulation.

## 5. Summary and discussion

This paper documents a methodology to achieve local variations in the representation of the background-error horizontal correlations. This represents a major

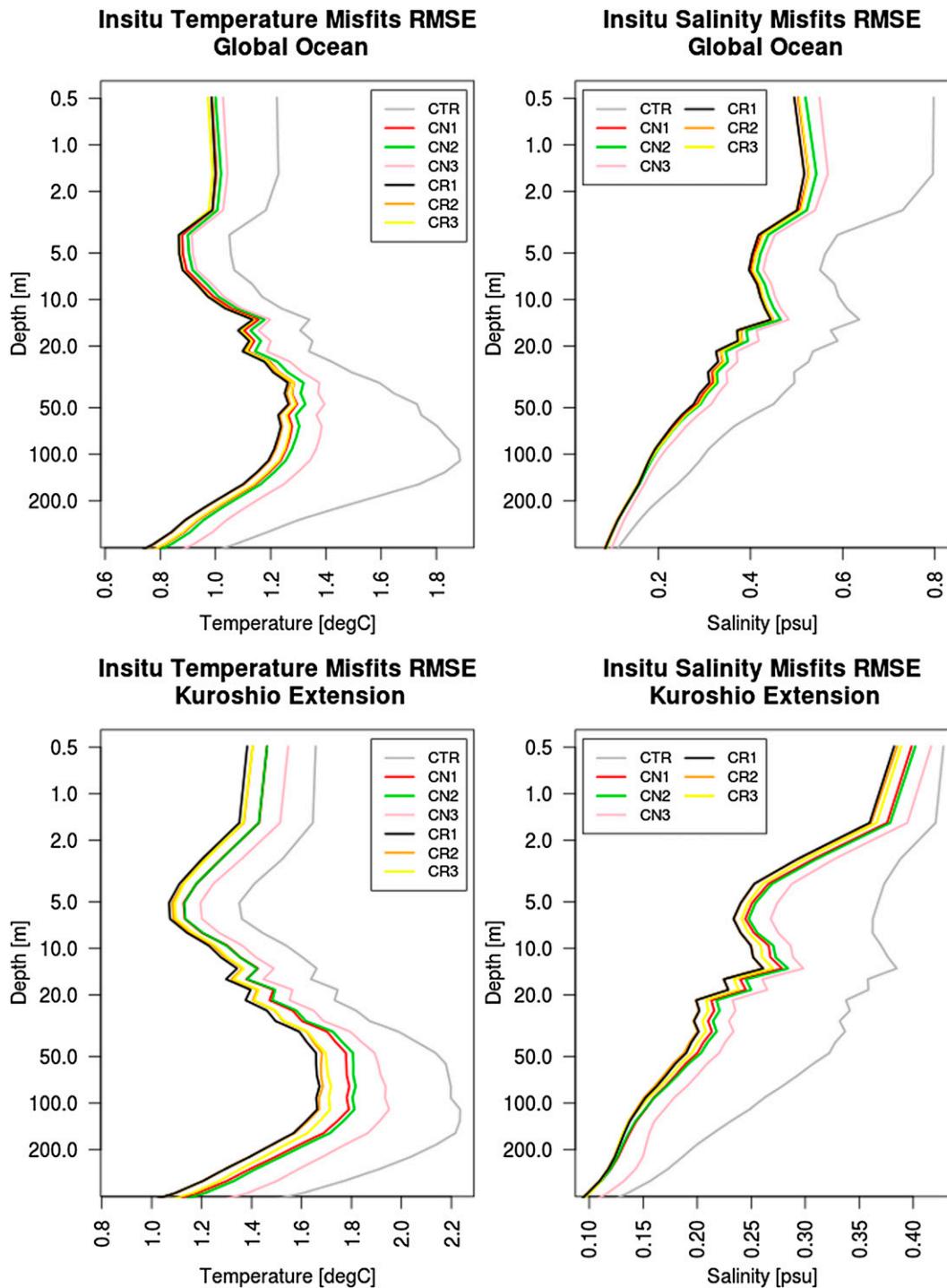


FIG. 9. Vertical profiles of RMSE of (left) temperature and (right) salinity in situ observations for the (top) global ocean and (bottom) Kuroshio Extension. The depth axis is logarithmic.

novelty in the global implementation of OceanVar, the ocean data assimilation system used at the Euro-Mediterranean Center on Climate Change [Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC)]

for both global reanalysis applications and global and regional operational forecasting.

A first-order recursive filter is used as a horizontal correlation operator in OceanVar. The recursive filter

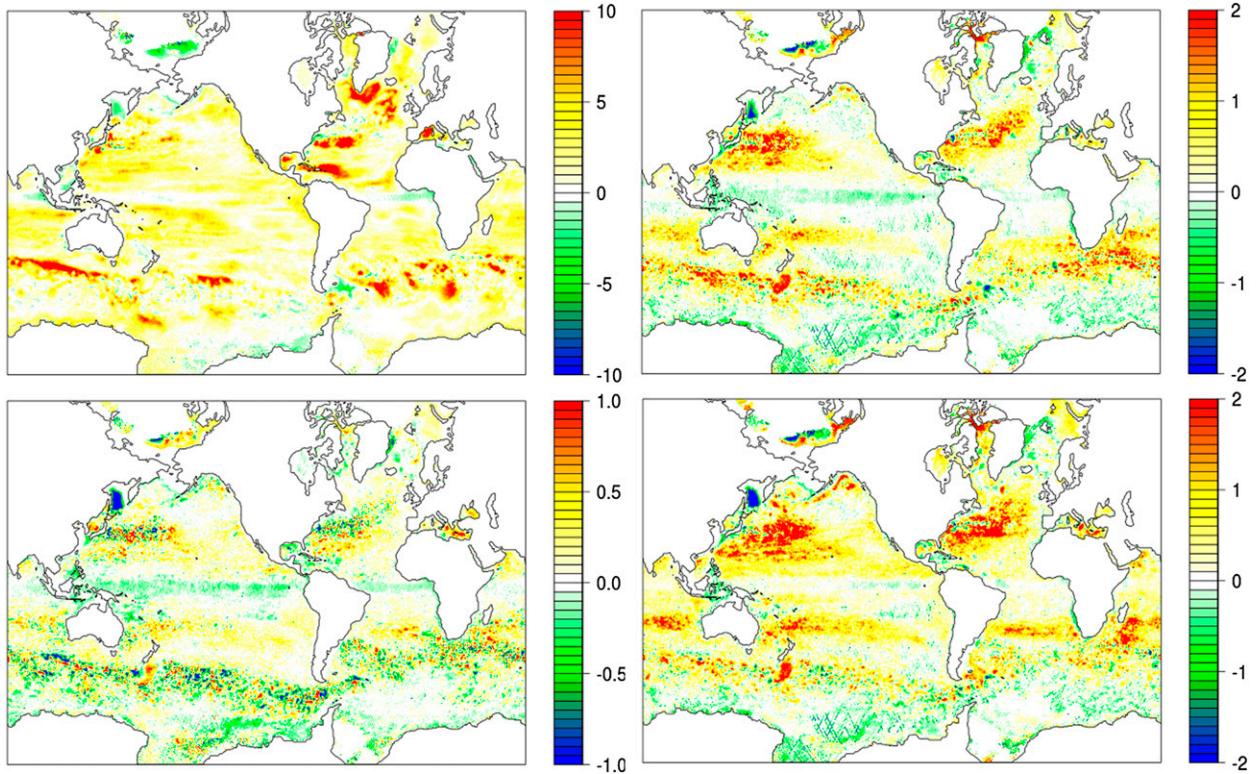


FIG. 10. Gridded differences of RMSE of SLA (cm) observations of (top) (left) CTR minus CR1 and (right) CN1 minus CR1; and (bottom) (left) CR2 minus CR1 and (right) CR3 minus CR1. Note that the contour intervals and color scales are different.

coefficients are then computed from the correlation length-scale dataset at full model resolution. This requires that correlation length scales vary slowly in the space (Purser et al. 2003b), and the use of a low-pass filter makes sure that this assumption is satisfied. Furthermore, gridpoint filters require the definition of normalization coefficients. Although this could be computationally costly, the use of lookup tables that contain normalization coefficients as a function of discrete values of model resolution and correlation length scales overcomes this issue.

This strategy is implemented in the CMCC global ocean analysis system at  $1/2^\circ$  of the horizontal resolution. We evaluate the impact of computing the correlation length scales in three different ways by changing either the error dataset used to extrapolate the length scales or the formula adopted to approximate the length scales [Eq. (4) vs Gaussian fit]: CA method vs NMC method, using anomalies with respect to the climatology, or the differences in weekly forecasts valid at the same time but initialized at different time. While innovation-based methods exist, which have been extensively exploited in atmospheric data assimilation, we argue that these methods are questionable in the ocean, where the coverage of the observations is very irregular in space and time

and differs between the OceanVar-analyzed variables (temperature and salinity) for most observing networks. Further studies should be devoted to understanding whether ensemble-derived correlation length scales are able to further improve the analysis system.

We may assume that in other oceanographic systems with a similar resolution, our method would produce similar correlation scales. On the other hand, even within our analysis system the correlation scales may become significantly different when the model resolution changes.

The results coming from the experiments presented in this work highlight the superiority of the nonuniform length scales. This appears particularly crucial for eddy-dominated areas, where the RMSE decreases by at least 7%, 4%, and 7% for temperature, salinity, and sea level anomalies, respectively. Furthermore, the ocean variability of these regions, in terms of eddy kinetic energy, is better captured when nonuniform correlations are used. Therefore, all the correlation length-scale formulations either uniform or based on latitudinal or Rossby radius dependence are not able to properly represent the spatial scales in these areas. For other areas, such as the tropical regions, the impact is still positive, but the percentage error decrease is not as large as in the eddy-dominated areas, resulting also from the fact that the

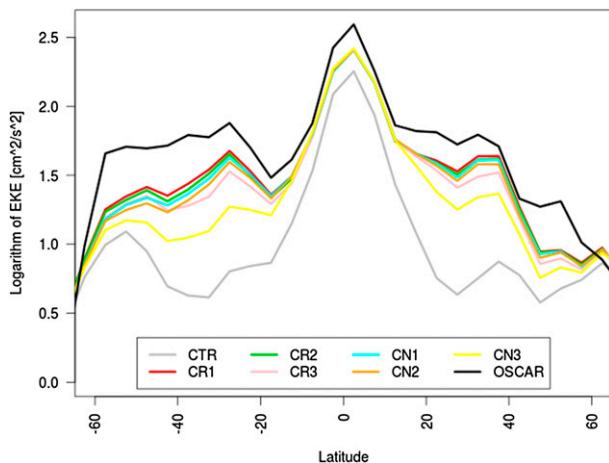


FIG. 11. Zonal averages of the logarithm of EKE for all the experiments and for the OSCAR dataset of near-surface currents.

analyses are more sensitive to variations of length scales for small values of the length scales themselves. All the methods agree in representing zonal correlations at least twice longer than meridional correlations in proximity of the equator. Furthermore, the signature of the ITCZs is clearly visible in the near-surface correlation length scales, making the seasonality of their patterns non-negligible (seasonal displacement of ITCZs) up to about 50 m of depth from the sea surface.

In direct comparison to the correlation length scales derived either from the climatological anomalies or through the NMC method, it is possible to see how the former draws the length scales from the spatial scales of the natural ocean variability rather than the forecast errors, showing, for example, the longest correlations near the equator, while the latter depends more on the observation coverage, for instance, exhibiting shorter length scales at depth or in the Southern Ocean. We have also reported a different vertical structure in case the two methods are used, with a decrease of temperature length scales more pronounced for the NMC statistics and longer than salinity, with salinity length scales kept rather constant with depth. However, in terms of verification skill scores, their relative RMSE differences are smaller than expected, indicating a slight superiority of the anomaly-derived correlations.

On the other hand, the Gaussian fit procedure to calculate the length scales led to longer estimates, which in turn deteriorated the skill score statistics with respect to the approximate formulation of Eq. (4). We found very similar patterns but the fit gave too much weight to

the tails of the autocorrelation function. The Gaussian fit may be improved, for instance, by considering a prior localization within the regression of the autocorrelation function but in practice it may become very computationally expensive.

Finally, the use of overestimated uniform correlation length scales corrected only partly the background fields with respect to the non-assimilative experiment, and prevented a quick convergence of the 3DVAR minimization, suggesting that the accurate design of the correlation length scales is of crucial importance in the design of a global ocean data assimilation system.

*Acknowledgments.* This work has received funding from the Italian Ministry of Education, University and Research and the Italian Ministry for the Environment, Land and Sea under the GEMINA project and from the European Commission's Copernicus program, previously known as the GMES program, under the MyOcean and MyOcean2 projects. The EN3 subsurface ocean temperature and salinity data were collected, quality controlled, and distributed by the Met Office Hadley Centre. The altimeter products were produced by SSALTO/DUACS and distributed by AVISO, with support from CNES. The authors thank the NOAA/OSCAR group for providing satellite-derived current data. The authors are grateful to two anonymous reviewers for their fruitful comments.

## APPENDIX

### The Recursive Filter

The horizontal transformation operator is defined as

$$\mathbf{V}_h = \mathbf{W}_y \mathbf{G}_y \mathbf{W}_x \mathbf{G}_x, \quad (\text{A1})$$

where  $\mathbf{G}_x$  and  $\mathbf{G}_y$  represent the recursive filter operators in the  $x$  and  $y$  directions, respectively, while  $\mathbf{W}_x$  and  $\mathbf{W}_y$  are diagonal matrices with normalization coefficients that may account for variable grid resolutions and correlation scales in the  $x$  and  $y$  directions, respectively (Purser et al. 2003b). The normalization coefficients are needed in order to have correlations with unit amplitude (Mirouze and Weaver 2010; Yaremchuk and Carrier 2012). The recursive filter operator  $\mathbf{G}$ , along the direction  $x$ , is formulated as

$$\mathbf{G}_x = \begin{cases} B(x, y, z) = \alpha(x, y, z)B(x-1, y, z) + [1 - \alpha(x, y, z)]A(x, y, z) \\ C(x, y, z) = \alpha(x, y, z)C(x+1, y, z) + [1 - \alpha(x, y, z)]B(x, y, z) \end{cases} \quad (\text{A2})$$

and symmetrically along the  $y$  direction. In Eq. (A2)  $A$  is the original field value at location  $(x, y, z)$ ,  $B(x, y, z)$  is the value after the forward step of the filter, and  $C(x, y, z)$  is the value after one complete (forward and backward) iteration of the filter. The  $\alpha$  represents the spatially varying filter coefficients, which are formulated as in Lorenc (1992):

$$\alpha(x, y, z) = 1 + E(x, y, z) - \sqrt{E(x, y, z)(E(x, y, z) + 2)} \quad (\text{A3})$$

$$E(x, y, z) = 2N\Delta x(x, y)^2 / (4L_x(x, y, z)^2), \quad (\text{A4})$$

where  $\Delta x(x, y)$  is the zonal resolution,  $N$  is the number of filter iterations, and  $L_x(x, y, z)$  is the zonal correlation length scale, and likewise along the  $y$  direction. Appropriate boundary conditions for the solution of Eq. (A2) are discussed by Hayden and Purser (1995). The number of filter iterations is set to 4, which yields a good enough approximation of a Gaussian shape (Purser et al. 2003a). The filter coefficients, at each grid point, depend therefore on the ratio between the correlation length scale and the spatial resolution of the model grid. Figure A1 shows the filter coefficient  $\alpha$  as a function of the  $L_x/\Delta x$  ratio. The filter coefficient  $\alpha$  regulates the smoothness of the filter output field: the larger  $\alpha$  is, the smoother the output field will be. Thus, Fig. A1 suggests that the filter is very sensitive to small values of the ratio, while for values greater than 5 there is no significant change in the  $\alpha$  coefficient. Consequently, changes of the correlation length scales for length scales comparable to the grid resolution are expected to significantly impact the analyses; for large values of the  $L_x/\Delta x$  ratios, the tuning of the correlation length scales is expected to have a small impact.

To use locally varying HCLSSs, the coefficients need to be calculated and normalized for each point of the three-dimensional grid. The normalization implies that a recursive filter operator must be run for each grid point of the three-dimensional state vector and separately for temperature and salinity in order to compute the attenuation of the filter from a unitary impulse. This step is quite costly for high-resolution global domains. To overcome this problem, a lookup table is built once offline and read at each assimilation step. The lookup table contains the normalization coefficients as a function of discrete values of the horizontal resolution and the correlation length scale. At each analysis step, the nonuniform correlation length scales (defined in section 3) are read and the normalization coefficients are assigned, accordingly, to every grid point by bilinearly interpolating the four closest discrete values present within the lookup table. Note that the lookup table

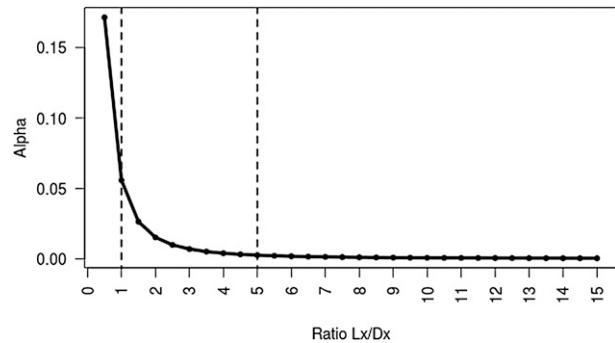


FIG. A1. The filter coefficient  $\alpha$  as a function of the ratio between the correlation length scale and the spatial resolution for realistic values of this ratio between 0.5–15. The left and right dashed lines correspond to values of the ratio equal to 1 and 5, respectively.

approach eases the possible extension of the recursive filter operator to the case of flow-dependent horizontal correlation length scales (e.g., Wang et al. 2008) or to multiscale applications, where several recursive filter operators may be used sequentially with different correlation scales.

As an additional remark, the formulation previously given implicitly provides an anisotropic structure of the HCLSSs, as correlations are set up independently along the  $x$  and  $y$  directions. This seems another important extension with respect to the uniform case, as tropical areas are known to be characterized by zonal correlations that are longer than the meridional ones (Meyers et al. 1991).

## REFERENCES

- Adani, M., S. Dobricic, and N. Pinardi, 2011: Quality assessment of a 1985–2007 Mediterranean Sea reanalysis. *J. Atmos. Oceanic Technol.*, **28**, 569–589, doi:10.1175/2010JTECHO798.1.
- Antonov, J. I., and Coauthors, 2010: *Salinity*. Vol. 2, *World Ocean Atlas 2009*, NOAA Atlas NESDIS 69, 184 pp.
- Barker, D. M., W. Huang, Y. R. Guo, and Q. Xiao, 2004: A three-dimensional data assimilation system for use with MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914, doi:10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2.
- Bellucci, A., S. Masina, P. Di Pietro, and A. Navarra, 2007: Using temperature salinity relations in a global ocean implementation of a multivariate data assimilation scheme. *Mon. Wea. Rev.*, **135**, 3785–3807, doi:10.1175/2007MWR1821.1.
- Belo-Pereira, M., and L. Berre, 2006: The use of an ensemble approach to study the background-error covariances in a global NWP model. *Mon. Wea. Rev.*, **134**, 2466–2489, doi:10.1175/MWR3189.1.
- Bernie, D. J., E. Guilyardi, G. Madec, J. M. Slingo, and S. J. Woolnough, 2007: Impact of resolving the diurnal cycle in an ocean–atmosphere GCM. Part 1: A diurnally forced OGCM. *Climate Dyn.*, **29**, 575–590, doi:10.1007/s00382-007-0249-6.
- Berre, L., S. Ștefănescu, and M. Belo, 2006: The representation of analysis effect in three error simulation techniques. *Tellus*, **58A**, 196–209, doi:10.1111/j.1600-0870.2006.00165.x.

- Bonjean, F., and G. Lagerloef, 2002: Diagnostic model and analysis of the surface currents in the tropical Pacific Ocean. *J. Phys. Oceanogr.*, **32**, 2938–2954, doi:10.1175/1520-0485(2002)032<2938:DMAAOT>2.0.CO;2.
- Carton, J. A., B. S. Giese, X. Cao, and L. Miller, 1996: Impact of altimeter, thermistor, and expendable bathythermograph data on retrospective analyses of the tropical Pacific Ocean. *J. Geophys. Res.*, **101**, 14 147–14 159, doi:10.1029/96JC00631.
- , G. Chepurin, X. Cao, and B. S. Giese, 2000: A Simple Ocean Data Assimilation analysis of the global upper ocean 1950–95. Part I: Methodology. *J. Phys. Oceanogr.*, **30**, 294–309, doi:10.1175/1520-0485(2000)030<0294:ASODAA>2.0.CO;2.
- Cavaliere, D. J., C. L. Parkinson, P. Gloersen, J. C. Comiso, and H. J. Zwally, 1999: Deriving long-term time series of sea ice cover from satellite passive-microwave multisensor data sets. *J. Geophys. Res.*, **104**, 15 803–15 814, doi:10.1029/1999JC900081.
- Chelton, D., R. Deszoeke, M. Schlax, K. El Naggar, and N. Siwertz, 1998: Geographical variability of the first baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.*, **28**, 433–460, doi:10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2.
- Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387, doi:10.1002/qj.49712051912.
- Cummings, J. A., 2005: Operational multivariate ocean data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3583–3604, doi:10.1256/qj.05.105.
- Dai, A., and K. E. Trenberth, 2002: Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *J. Hydrometeorol.*, **3**, 660–687, doi:10.1175/1525-7541(2002)003<0660:EOFDFC>2.0.CO;2.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Dee, D., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Derber, J., and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333–1347, doi:10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2.
- Dobricic, S., and N. Pinardi, 2008: An oceanographic three-dimensional assimilation scheme. *Ocean Modell.*, **22**, 89–105, doi:10.1016/j.ocemod.2008.01.004.
- Fichefet, T., and M. A. Morales Maqueda, 1997: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *J. Geophys. Res.*, **102**, 12 609–12 646, doi:10.1029/97JC00480.
- Gaspari, G., S. Cohn, J. Guo, and S. Pawson, 2006: Construction and application of correlation functions with variable length-fields. *Quart. J. Roy. Meteor. Soc.*, **132**, 1815–1838, doi:10.1256/qj.05.08.
- Hayden, C., and R. Purser, 1995: Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. *J. Appl. Meteor.*, **34**, 3–15, doi:10.1175/1520-0450-34.1.3.
- Ingleby, B., and M. Huddleston, 2007: Quality control of ocean temperature and salinity profiles—Historical and real-time data. *J. Mar. Syst.*, **65**, 158–175, doi:10.1016/j.jmarsys.2005.11.019.
- Isaksen, L., M. Fisher, and J. Berner, 2007: Use of analysis ensembles in estimating flow-dependent background error variance. *Proc. ECMWF Workshop on Flow-Dependent Aspects of Data Assimilation*, ECMWF, Reading, United Kingdom, 37 pp. [Available online at [http://old.ecmwf.int/newsevents/meetings/workshops/2007/data\\_assimilation/presentations/Isaksen.pdf](http://old.ecmwf.int/newsevents/meetings/workshops/2007/data_assimilation/presentations/Isaksen.pdf)].
- Large, W. G., and S. G. Yeager, 2004: Diurnal to decadal global forcing for ocean and sea-ice models: The data sets and flux climatologies. NCAR Tech. Note NCAR/TN-460+STR, 105 pp., doi:10.5065/D6KK98Q6.
- Le Traon, P. Y., F. Nadal, and N. Ducet, 1998: An improved mapping method of multisatellite altimeter data. *J. Atmos. Oceanic Technol.*, **15**, 522–534, doi:10.1175/1520-0426(1998)015<0522:AIMMOM>2.0.CO;2.
- Locarnini, R. A., A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, and D. R. Johnson, 2010: *Temperature*. Vol. 1, *World Ocean Atlas 2009*, NOAA Atlas NESDIS 68, 184 pp.
- Lorenc, A., 1992: Iterative analysis using covariance functions and filters. *Quart. J. Roy. Meteor. Soc.*, **118**, 569–591, doi:10.1002/qj.49711850509.
- Madec, G., and M. Imbard, 1996: A global ocean mesh to overcome the north pole singularity. *Climate Dyn.*, **12**, 381–388, doi:10.1007/BF00211684.
- , P. Delecluse, M. Imbard, and C. Lévy, 1998: OPA 8.1 Ocean General Circulation Model reference manual. IPSL Note du Pôle de Modélisation 11, 91 pp.
- Meyers, G., H. Phillips, N. Smith, and J. Sprintall, 1991: Space and time scales for optimal interpolation—Tropical Pacific Ocean. *Prog. Oceanogr.*, **28**, 189–218, doi:10.1016/0079-6611(91)90008-A.
- Mirouze, I., and A. Weaver, 2010: Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quart. J. Roy. Meteor. Soc.*, **136**, 1421–1443, doi:10.1002/qj.643.
- Pannekoucke, O., L. Berre, and G. Desroziers, 2008: Background-error correlation length-scale estimates and their sampling statistics. *Quart. J. Roy. Meteor. Soc.*, **134**, 497–508, doi:10.1002/qj.212.
- Parrish, D., and J. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763, doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.
- Pujol, M.-I., S. Dobricic, N. Pinardi, and M. Adani, 2010: Impact of multialtimeter sea level assimilation in the Mediterranean Forecasting Model. *J. Atmos. Oceanic Technol.*, **27**, 2065–2082, doi:10.1175/2010JTECHO715.1.
- Purser, R., W.-S. Wu, D. Parrish, and N. Roberts, 2003a: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1524–1535, doi:10.1175/1520-0493(2003)131<1524:NAOTAO>2.0.CO;2.
- , —, —, and —, 2003b: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.*, **131**, 1536–1548, doi:10.1175/2543.1.
- Raynaud, L., L. Berre, and G. Desroziers, 2011: An extended specification of flow-dependent background error variances in the Météo-France global 4D-Var system. *Quart. J. Roy. Meteor. Soc.*, **137**, 607–619, doi:10.1002/qj.795.
- Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution blended analyses for sea surface temperature. *J. Climate*, **20**, 5473–5496, doi:10.1175/2007JCLI1824.1.
- Simmons, A., S. Uppala, D. Dee, and S. Kobayashi, 2007: ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter*, No. 110, ECMWF, Reading, United Kingdom, 25–35.
- Storto, A., and R. Randriamampianina, 2010: Ensemble variational assimilation for the representation of background error

- covariances in a high-latitude regional model. *J. Geophys. Res.*, **115**, D17204, doi:[10.1029/2009JD013111](https://doi.org/10.1029/2009JD013111).
- , S. Dobricic, S. Masina, and P. Di Pietro, 2011: Assimilating along-track altimetric observations through local hydrostatic adjustments in a global ocean reanalysis system. *Mon. Wea. Rev.*, **139**, 738–754, doi:[10.1175/2010MWR3350.1](https://doi.org/10.1175/2010MWR3350.1).
- , I. Russo, and S. Masina, 2012: Interannual response of global ocean hindcasts to a satellite-based correction of precipitation fluxes. *Ocean Sci. Discuss.*, **9**, 611–648, doi:[10.5194/osd-9-611-2012](https://doi.org/10.5194/osd-9-611-2012).
- , S. Masina, and S. Dobricic, 2013: Ensemble spread-based assessment of observation impact: Application to a global ocean analysis system. *Quart. J. Roy. Meteor. Soc.*, **139**, 1842–1862, doi:[10.1002/qj.2071](https://doi.org/10.1002/qj.2071).
- Wang, X., D. M. Barker, C. Snyder, and T. Hamill, 2008: A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116–5131, doi:[10.1175/2008MWR2444.1](https://doi.org/10.1175/2008MWR2444.1).
- Weaver, A. T., and P. Courtier, 2001: Correlation modelling on the sphere using a generalized diffusion equation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1815–1846, doi:[10.1002/qj.49712757518](https://doi.org/10.1002/qj.49712757518).
- , and I. Mirouze, 2013: On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **139**, 242–260, doi:[10.1002/qj.1955](https://doi.org/10.1002/qj.1955).
- , C. Deltel, E. Machu, S. Ricci, and N. Daget, 2005: A multivariate balance operator for variational ocean data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3605–3625, doi:[10.1256/qj.05.119](https://doi.org/10.1256/qj.05.119).
- Yaremchuk, M., and M. Carrier, 2012: On the renormalization of the covariance operators. *Mon. Wea. Rev.*, **140**, 637–649, doi:[10.1175/MWR-D-11-00139.1](https://doi.org/10.1175/MWR-D-11-00139.1).
- Zhou, G., W. Fu, J. Zhu, and H. Wang, 2004: The impact of location-dependent correlation scales in ocean data assimilation. *Geophys. Res. Lett.*, **31**, L21306, doi:[10.1029/2004GL020579](https://doi.org/10.1029/2004GL020579).