TECHNICAL MEMORANDUM

795

A generic ensemble generation scheme for data assimilation and ocean analysis

H. Zuo, M. A. Balmaseda,
E. de Boisseson, S. Hirahara,
M. Chrust, P. de Rosnay

Research Department

November 2017

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Abstract

A new generic perturbation scheme suitable for generation of an ensemble of ocean analysis is presented. The scheme consists of two distinct elements: perturbations to the assimilated observations, both profiles and surface observations, and perturbations to the surface forcing fields. The new scheme has been applied to the new Ocean ReAnalysis System-5 (ORAS5). The surface forcing perturbation has also been used to create oceanic surface forcing for ERA5, and in operational Ensemble Data Assimilation (EDA) from cycle 43R1.

The idea behind the observation perturbation scheme is to account for observation representativeness error. Instead of perturbing the value of the assimilated observations, the scheme perturbs the position of the observations. This is done by applying perturbations to the geographical location of the in-situ temperature and salinity profiles, and by random thinning, both in the horizontal for surface observations, and in the vertical for dense profiles. This method exploits the full observation data set and uses more observations (through ensemble approach) than the previous thinning method. The impact of the perturbation scheme in the ocean reanalysis is illustrated together with selected sensitivity experiments. It is shown that the observation perturbations have little impact in global or basin wide climate indices, but they have local effect. The ensemble spread shows large errors in regions with strong mesoscale eddy activities and in areas affected by the Mediterranean Outflow waters. These are regions where departures with respect to observations are also large. It is also shown that ensemble spread in the tropical upper-ocean is under-dispersive with only five ensemble members, but it improves by increasing the ensemble size.

The estimation of the diagonal elements of BackGround Error (BGE) covariances using the ensemble spread generated by observation perturbation has been compared with the specified BGE values and also with those diagnosed using Desroziers' method. Results show stronger agreement in spatial patterns and values between the ensemble and Desroziers' estimates than with the specified BGE values. However, it is discussed that the ensemble estimation is very sensitive to the way the ensemble is created, and will need to be corrected in regions where observations are scarce. A robust combination of parameterized and ensemble-derived BGE covariances is recommended for future developments.

A revised scheme for generating perturbations to surface forcing has also been developed. It is a generalization of the previous scheme and is still based on sampling past differences between different sources of information. The previous scheme, implemented as part of the seasonal forecasting system 2 (S2), created monthly perturbations for wind stress and Sea Surface Temperature (SST), based on sampled differences between atmospheric re-analysis products. The new scheme is more general in several aspects: i) it allows for representation of both analysis and structural uncertainty; ii) it permits different temporal de-correlation scales of the perturbations; iii) it encompasses a wider range of variables and iv) it preserves the multivariate relationships among the perturbed variables. The reference data sets for sampling the perturbations have also been updated. The analysis uncertainty is sampled using the ensemble information from ERA-20C. The structural uncertainty in SST is sampled using more up-to-date data sets of high resolution ESA-CCI and HadISSTv2.1. Sea Ice Concentration (SIC) structural uncertainty is sampled using differences between HadISSTv2.0 and v2.1. The scheme is not fully flow dependent yet as it represents only the seasonal variations of uncertainty. However, it has been designed to be compatible with the flow dependent perturbations such as those produced by the real-time EDA; in particular, the climatological analysis uncertainty perturbations can be replaced by those from the EDA when the latter becomes available. The new SST and sea-ice perturbation strategy developed is also used by ERA5 and by the operational EDA (albeit with different parameter choices). A version control number ($v3$) will allow further updates in the future.

Contents

1 Introduction **4**

2 Observation perturbations **5**

2.1 Representativeness errors and observation perturbations 5

2.2 Perturbation of in-situ observations 6

2.2.1 Perturbation of horizontal positions of in-situ observation 6

2.2.2 Perturbation of vertical positions of in-situ observation 7

2.3 Perturbation of surface observations 7

2.3.1 Perturbation of Sea Ice observations 7

2.3.2 Perturbation of Sea Level observations 10

3 Assessment of observation perturbations **13**

3.1 Experiments with observation perturbations 14

3.2 Geographical distribution of ensemble spread 14

3.3 Vertical profiles of ensemble spread 17

3.4 Temporal variability of ensemble statistics 21

3.5 Evaluation against independent data 23

3.6 Assessment of surface observation perturbations 25

4 Surface forcing perturbations **29**

4.1 The new forcing perturbations scheme 29

4.2 Temporal Decorrelation Scales and Temporal Interpolation 32

4.3 Multiple Variables and Sources of Uncertainty 33

5 Summary and discussion **39**

List of Figures

1 Illustration of horizontal perturbation of in-situ profiles 8

2 Illustration of vertical perturbation of in-situ profiles 9

3 A schematic plot of the surface observation perturbation 11

4 Illustration of SLA perturbation 12

5 Maps of $H\,pert200$ T/S ensemble spread at 100m 15

6 Vertical profiles of ensemble spread . . . 18
7 Vertical profiles of ensemble spread in N.Atlantic . . . 19
8 Maps of ensemble spread at 2000m . . . 20
9 Vertical profiles of ensemble spread in the Tropics . . . 21
10 Time series of ensemble mean OHC changes . . . 22
11 Time series of OHC ensemble spread . . . 23
12 Maps of ensemble mean differences to WOA13 . . . 24
13 Maps of SST ensemble variances . . . 26
14 Illustration of SIC perturbation . . . 27
15 Maps of SIC ensemble spread . . . 28
16 Maps of SIC and SIT ensemble spreads . . . 29
17 Maps of SLA ensemble variance . . . 30
18 Surface perturbation: maps of SST ensemble spread . . . 35
19 Surface perturbation: maps of wind stress ensemble spread . . . 36
20 Surface perturbation: maps of SIC ensemble spread . . . 37
21 Surface perturbation: maps of AE perturbation . . . 38

List of Tables

1 Summary of the system configurations for sensitivity experiments . . . 14
2 Summary of perturbation configurations for sensitivity experiments . . . 16
3 Summary of Forcing Perturbations *v*3 . . . 32

1 Introduction

The Ocean ReAnalysis System (ORAS in what follows) at ECMWF was from the beginning conceived as an integral part of the seasonal forecasting system, providing the ocean initial conditions that are required for skilful forecasts at seasonal time scales, in particular, for ENSO. A good estimation of the ocean initial condition should be accompanied by the corresponding uncertainty estimation of the ocean state. This is why an ensemble of five ocean reanalyses was created as part of the initialization procedure for the Seasonal Forecasting System S2 and S3 (Anderson et al. [2007]). The ensemble was generated by applying perturbations to the wind stress and to the SST. Vialard et al. [2005] describe the perturbation method and quantify the impact of these perturbations on seasonal forecasts of ENSO. The same technique was later employed in subsequent ocean reanalysis systems (ORAS3 and ORAS4, see Balmaseda et al. [2008, 2013]), and it was used in European projects such as DEMETER and ENSEMBLES [Weisheimer et al., 2007]. The technique was extended to include precipitation perturbations in Daget et al. [2009], who tried to use the the ensemble information in a prototype of EDA with the NEMOVAR ocean data assimilation system, in an attempt to assess the impact of introducing flow dependency in the NEMOVAR background error covariance matrix. The forcing perturbation code and the data base needed updating. We took the opportunity to generalize the perturbation scheme. The new scheme exploits the information from new data sets, such as ERA-20C [Poli et al., 2016] and ESA-CCI [Merchant et al., 2014], it allows different temporal decorrelation scales (for instance monthly, 5-days pentad), it is extended to more surface fluxes variables, such as solar radiation, precipitation and sea-ice, and, when possible, it preserves multivariate relationships. The same scheme is used to generate the SST and sea-ice perturbations in ERA-5 and in the EDA since Cycle 43r1 [Hirahara et al., 2016].

Daget et al. [2009] also implemented a scheme to perturb the values of the assimilated observations according to the specifications of the observation error covariance matrix $\mathbf{R}$, which is a common practice in ensemble methods [Shlyaeva et al., 2016]. However, the values of $\mathbf{R}$ are usually specified for convenience and practical reasons, rather than to reflect to the perceived observation errors. For instance, for practical reasons, $\mathbf{R}$ is usually diagonal and the diagonal variances are inflated as a way to mitigate the fact that the Representativeness Errors (REs) are correlated in space. In the case of the ocean, perturbing the observation values independently according to the values of $\mathbf{R}$ can lead to the assimilation of very unrealistic observation profiles, which are not hydro-statically stable. This is why here we try another approach in the observation perturbations, which aims at accounting only for REs and correlations in the observation errors. This is done by perturbing positions of the whole profiles within the scales that the model resolution permits to resolve (RE), and by applying random thinning algorithms where different ensemble members use different observations within a certain spatial criteria (correlation in observation error). The latter has the potential of exploiting the new high resolution observational data sets in ensemble-based data assimilation methods.

This memorandum is organized as follows. Section 2 gives a description of the observation perturbations applied to both in-situ (Section 2.2) and surface observations (Section 2.3). Evaluation of the perturbation methods is presented in Section 3, while ensemble estimates were assessed against the diagonal elements of BGE covariances and the independent observation data sets through a series of sensitivity experiments. Perturbation of surface forcing is described in Section 4. This surface perturbation scheme also allows for different temporal de-correlation scales (Section 4.2) and accounts for multivariate relationships among the perturbed variables (Section 4.3). A summary and conclusions are given in Section 5.

2 Observation perturbations

2.1 Representativeness errors and observation perturbations

The ECMWF's Ocean Data Assimilation system (ODA) uses a variational data assimilation scheme developed in collaboration with CERFACS, the UK Met Office and INRIA for the NEMO ocean model (NEMOVAR, see Weaver et al. [2005]). NEMOVAR is applied as an incremental three-dimensional variational assimilation (3D-VAR) using FGAT (the First-Guess at Appropriate Time) approach. Details about NEMOVAR data assimilation scheme in its 3D-VAR FGAT configuration can be found in Mogensen et al. [2012]. The increments δw can be obtained by approximately minimizing the quadratic cost function

$$J[\delta w] = \frac{1}{2}\delta w^T \mathbf{B}^{-1}\delta w + \frac{1}{2}(\mathbf{G}\delta w - \delta y^o)^T \mathbf{R}^{-1}(\mathbf{G}\delta w - \delta y^o) \tag{1}$$

where δy^o is innovation and can be written as

$$\delta y^o = y^o - \boldsymbol{G}(w^b) \tag{2}$$

Here w^b is the background state, $\boldsymbol{G}$ is the generalized non-linear observation operator that maps the control vector into the space of the observation vector. $\mathbf{G}$ is an approximation of the Tangent-Linear (TL) of $\boldsymbol{G}$. It includes a linearized balance operator ($\mathbf{K}$), which accounts for the multivariate relationship, and a linearized model propagator ($\mathbf{M}$), which propagates state vector to the observation time. $\mathbf{B}$ is a block diagonal univariate background error covariance matrix and $\mathbf{R}$ is a diagonal observation error covariance matrix. Here we will compare the specified diagonal elements of $\mathbf{B}$ (σ_b^s) with diagnosed value (σ_b^d) using Desroziers' method [Desroziers et al., 2005], and with estimation given by the ensemble spread (σ_b^e), as described below.

Correct estimations of both $\mathbf{B}$ and $\mathbf{R}$ are critical for optimal ocean data assimilation. BGE covariances comprise errors from initialization and boundary conditions as well as model errors. On the other hand, observation errors as quantified by $\mathbf{R}$ should include both instrument measurement errors and REs [Lorenc, 1986]. The later are introduced with degrading physical observations to model resolution and are associated with finite distance and time between two observations as well as finite resolution of the model fields. Estimation of REs of ocean observations, e.g. altimetry sea level, has been carried by Oke et al. [2008] with result suggesting that the values of REs are typically greater than or at least comparable to measurement errors, particularly in regions of strong mesoscale variability. Janssen et al. [2007] also proposed a method for estimation of the ocean observation errors using triple independent data sets.

Ideally, REs are resolution dependent and depend on the level of eddy activity in different regions [Schiller et al., 2008]. In practice, REs are normally assumed to be horizontally uniform and only depth dependent for temperature and salinity observations [Brasseur et al., 2005] with a non-uniform inflation to $\mathbf{R}$ at coastlines [Mogensen et al., 2012]. In the ECMWF ODA as well as in ORAS4 [Balmaseda et al., 2013], superobbing/thinning schemes have been used for averaging the observations toward the scales resolved by the model, as methods to reduce problems with representativeness. Here, for the development of the ORAS5 system, the same strategy of specifying total observation errors in $\mathbf{R}$ has been adapted from ORAS4. In addition, the superobbing/thinning schemes were exploited to include a perturbation scheme, and an ensemble approach has been adopted in order to account for the REs of both surface and sub-surface ocean observations. The ensemble spread in the background field introduced by observation perturbations can be used for estimation of the diagonal elements of $\mathbf{B}$. This is the first step

towards a flow-dependent estimation of BGE covariance matrix **B**. Implementation of this observation perturbation scheme is described in the following sections, which are organized by observation types that are assimilated in the ECMWF ODA system.

2.2 Perturbation of in-situ observations

An updated ocean in-situ observation data set from the Met Office Hadley Centre EN series (EN4, see Good et al. [2013]) has been assimilated in ORAS5. The EN4 data set includes temperature and salinity profiles obtained from Argo, Expendable BathyThermograph (XBT), Mechanical BathyThermograph (MBT), Conductivity-Temperature-Depth (CTD) instruments, moored buoys and other sources (i.e. seals), with depth correction [Gouretski and Reseghetti, 2010] applied to XBT and MBT data from 1975 onwards. All observation errors are assumed to be uncorrelated so only observation error standard deviations need to be specified. An empirical analytical function (see Mogensen et al. [2012]) is used for specification of total observation error standard deviations for both temperature and salinity. In addition, horizontal and vertical REs from temperature and salinity observations were taken into account in ORAS5 through an observation perturbation scheme that perturbs locations of in-situ observations and by applying a stratified vertical thinning in vertical profiles.

2.2.1 Perturbation of horizontal positions of in-situ observation

The location of observation matters when computing the innovation using observation and model-equivalent background field derived using observation operator. In practice not all ocean observations are point measurement, i.e., instrument location can drift during taking measurement. Take Argo float for example, it takes about 6 hours in ascending when taking measurement, before reaching surface to send report to satellite. As a result the recorded surface location includes an error spans from a few kilometres to approximately 20 km in regions with strong boundary currents. Assuming that profile observations in each single realization of perturbed member represent only a small scale which is not resolved by ocean model, the ensemble approach by perturbing profile's geographical location then spreads the same observation information into region with a representative value that can be measured in the model resolution, assuming that the perturbation distance and size of ensemble are large enough.

To do that, a random perturbation is added to the recorded geographic location of any given profile from the EN4 data set in order to account for its RE. It is equivalent to shifting the observation profile horizontally within a pre-defined distance (Perturbation Distance, PD thereafter). As the first attempt, the Probability Density Function (PDF) of RE is assumed to be unbiased with uniform distribution. Different perturbing strategies have been tested, including

- **P1**: perturb a profile with random distance (0 to PD) and random angle (0-360 degree)
- **P2**: perturb a profile with constant distance (as PD) and random angle (0-360 degree)
- **P3**: perturb latitude of a profile with random distance (0 to PD)
- **P4**: perturb longitude of a profile with random distance (0 to PD)
- **P5**: perturb both latitude and longitude of a profile with the same random distance (0 to PD)

As a results, for perturbing strategy **P1** the density function of perturbed profile location is uniformly distributed within an area defined by a circle of radius equals to PD. The implementation ensures that each

observation type can be perturbed independently, with different PD values. We chose to illustrate this profile horizontal location perturbation scheme in Fig. 1 with a large PD value (200 km) and ensemble size (200 members). The locations of in-situ observation profiles from EN4 for a given day (20130101) are shown in Fig. 1-(a) and (b) for un-perturbed and perturbed conditions, respectively. Readers, however, should note that types of perturbed observations and corresponding PD values as seen in Fig. 1 are chosen for illustration purpose only. Horizontal perturbation of in-situ observations in ORAS5 was not carried out in this configuration.

2.2.2 *Perturbation of vertical positions of in-situ observation*

A vertical thinning scheme is used in both ORAS4 [Balmaseda et al., 2013] and ORAP5 [Zuo et al., 2015b] for all in-situ observations, as a pragmatic way to reduce the impact of spatial observation error correlations. It is especially important for observation types with high vertical resolution (i.e. CTDs) as spatial correlation is unaccounted when specifying observation error covariance. In recent in-situ data sets such as EN4 (used for ORAS5), the vertical resolution of the profiles has increased considerably to 400 levels, compared to the 150 levels in the EN3 data set, which was assimilated in ORAS4. REs, however, are not parametrised in ORAS4, even though they are likely to be the main contribution to observation errors considering the low model resolution. In order to account for observation REs due to the finite vertical resolution of the model, a new vertical perturbation scheme has been implemented in ORAS5. It is based on the same thinning scheme as in ORAS4 together with a stratified sampling method that employs pre-defined depth ranges as sampling groups.

Te vertical perturbation scheme is characterised by a thinning factor N, which indicates the maximum number of observations that can be assimilated within each model level. In practice, a set of thinning levels is constructed using all model levels and $N-1$ additional levels that are equally distributed between each model level pair. In ORAS4 new thinned observation profiles are constructed using observations that are nearest in depth to this set of new thinning levels. In ORAS5 an additional step has been added to construct thinning layer spanning a depth range with each thinning level in the middle of the corresponding layer. These thinning layers are used as sampling groups to carry out stratified random sampling for in-situ observations. A schematic illustration of this new perturbation scheme can be seen in Fig. 2. Among all observations within a thinning layer only one is selected randomly. This allows to maintain the vertical structure of a given profile with no more than N observations within each model layer. The scheme guarantees that T/S observations are always selected from the same level. It also guarantees that the same observation number is used by each ensemble member. Compared with the vertical thinning strategy used in ORAS4, this method exploits the full vertical resolution of input observations and utilizes more observations through assimilating different observation sub-samples in different ensemble members.

2.3 Perturbation of surface observations

2.3.1 *Perturbation of Sea Ice observations*

A prognostic thermodynamic-dynamic sea-ice model (LIM2, see Fichefet and Maqueda [1997]) and assimilation of SIC data were first introduced in the ORAP5 system [Zuo et al., 2015a]. Assimilation of SIC is treated as univariate meaning cross-correlations between SIC and other ocean state variables are not taken into account. Before assimilation in NEMOVAR, a thinning algorithm was applied to the gridded SIC data in order to reduce the spatial correlations of observation errors, which are not explicitly

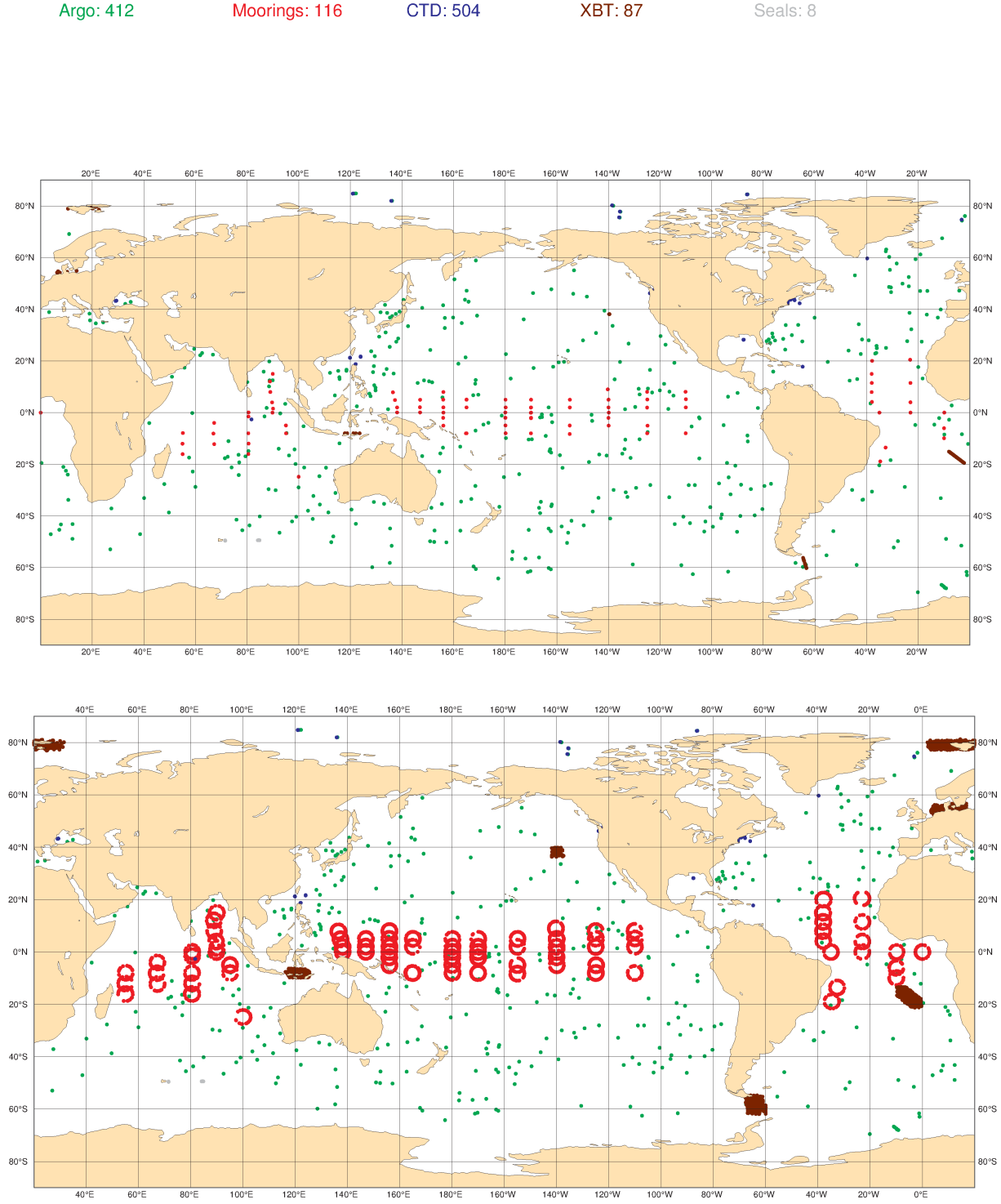

Figure 1: Geographic locations of (top) daily EN4 profiles in 20130101 and (bottom) perturbed EN4 profiles for the same day but with 200 ensemble members. For illustration purpose, only XBT and Moorings profiles are perturbed with a large PD value (200 km) and using perturbing strategy **P2** and **P5**, respectively. Here colour are used to denote different observation types.

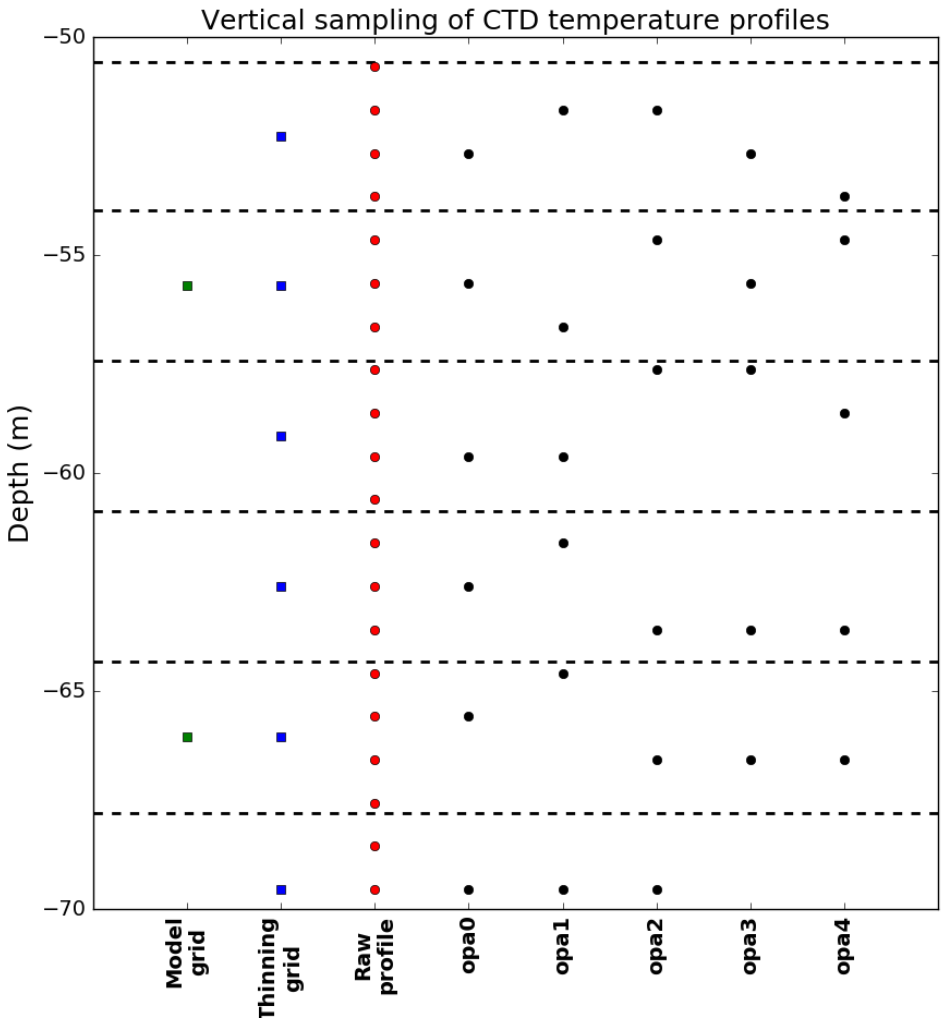

Figure 2: Example of the stratified sampling method used for vertical perturbation of in-situ observations, as demonstrated with a CTD profile from EN4 data set and thinning factor $N = 3$ with 5 ensemble members. The first ensemble member (opa0) is always un-perturbed and the same thinning method as used in ORAS4 is applied. Within each thinning layer only one observation, if exists, is selected using simple random sampling method for the remaining ensemble members (opa1-4).

accounted for in our ODA system, and to reduce the data density to a level commensurate with the resolution of the ORA system. This thinning of SIC data in ORAP5 was carried out using systematic sampling method with a fixed sampling interval (~0.5°) in both latitudinal and longitudinal directions. Here we go one step further and introduce a stratified random sampling method within this thinning scheme as a strategy for ensemble generation. It is implemented in ORAS5 as a generic perturbation scheme that can be applied to any type of gridded surface observation. Like in the vertical perturbation of in-situ temperature and salinity profiles, it is designed to account for the REs of surface observations arising due to finite NEMO model resolution. For simplicity, the PDF of observation REs is assumed to be unbiased with a uniform distribution.

First, a thinning box with reduced grid is constructed with its sides defined by a thinning length scale

(L_X, where X=SIC, or any other surface control variable). For gridded sea ice concentration data

$$L_{sic} = f_{sic} * L_o \tag{3}$$

where L_o is the length scale of the model grid; f_{sic} is a thinning factor for SIC, which is represented as a number of observation. For SIC data assimilation in ORCA1 configuration with $f_{sic} = 2$, L_{sic} is approximately 100 km in the Arctic Ocean. For any thinning box, we define

$$P_t(i,j) = \frac{M(i,j)}{f_{sic}^2} \tag{4}$$

where $P_t(i,j)$ is the sampling probability threshold; $M(i,j)$ is the number of observation samples contained in the thinning box (i,j), and f_{sic}^2 is the squared thinning factor for SIC. In most cases $M(i,j) = f_{sic}^2$, except when near boundaries of the gridded observation array (e.g. box(i, j-1) in Fig. 3). A random probability value ($P_s(i,j)$) between 0-1 is generated for each thinning box. A random sampling method is then performed in the thinning box (i,j) only when $P_s(i,j) \leq P_t(i,j)$, and at most n_{sic} of observations will be selected randomly for data assimilation from this thinning box. In the end we construct a logical 2D array ($S(i,j)$) for selection matrix. For simplicity n_{sic} is set to 1 in order to allow comparison of perturbed member with the control member for which a regular thinning scheme is always applied. An illustration of this perturbation scheme applied to gridded SIC observation is shown in Fig. 3.

In practice, the above perturbation scheme was implemented as part of the linearized observation operator (**G** in Eq. 1) for dealing with any gridded surface observation, after it was interpolated to the irregular NEMO ORCA grid. This scheme can be applied either to the global domain or to each sub-domain if NEMO is running with multiple processors. It guarantees that the same number of valid surface observations is ingested by the ECMWF ODA for both the control member, which uses a regular thinning scheme, and all perturbed members, which use the stratified random sampling scheme. In ORAS5 the actual number of assimilated surface observations varies in different ensemble members, due to additional QC process, e.g. check for land-sea mask and distance to coast. It is also worth noting that even though the sample selection matrix $S(i,j)$ is randomly generated specifically for each ensemble member, it is not evolving over time within the same assimilation window. Instead $S(i,j)$ is only updated at the beginning of each assimilation cycle. It can be considered as sub-optimal because daily averaged OSTIA SIC data is assimilated in the ECMWF ODA system. Revisiting this issue while considering the assimilation window length may be needed in the future.

2.3.2 Perturbation of Sea Level observations

Radar altimeter Sea-Level Anomaly (SLA) observations are assimilated in the ECMWF ODA system using NEMOVAR. The assimilated data set is along-track multi-mission altimeter SLA products from AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic data) and include observations from ERS-1, ERS-2, ENVISAT, TOPEX/Poseidon, Jason-1, Jason-2, Jason-3, GFO, CryoSat-2, SARAL and HY-2A. In order to assimilate the high spatial resolution along-track SLA data in the ECMWF ODA system and to avoid the problem of oversampling [Zuo et al., 2015b], a super-observation (superob) scheme as described by Mogensen et al. [2012] was implemented and was used in ORAS4 and ORAP5 productions. In this scheme, a reduced grid (L_{sla}) is constructed with a resolution typically comparable to or greater than that of the model. Altimeter observations are then binned in time and space to create super-observations. This way we effectively reduce the correlation of observation errors and alleviate the problems with representativeness due to finite model resolution.

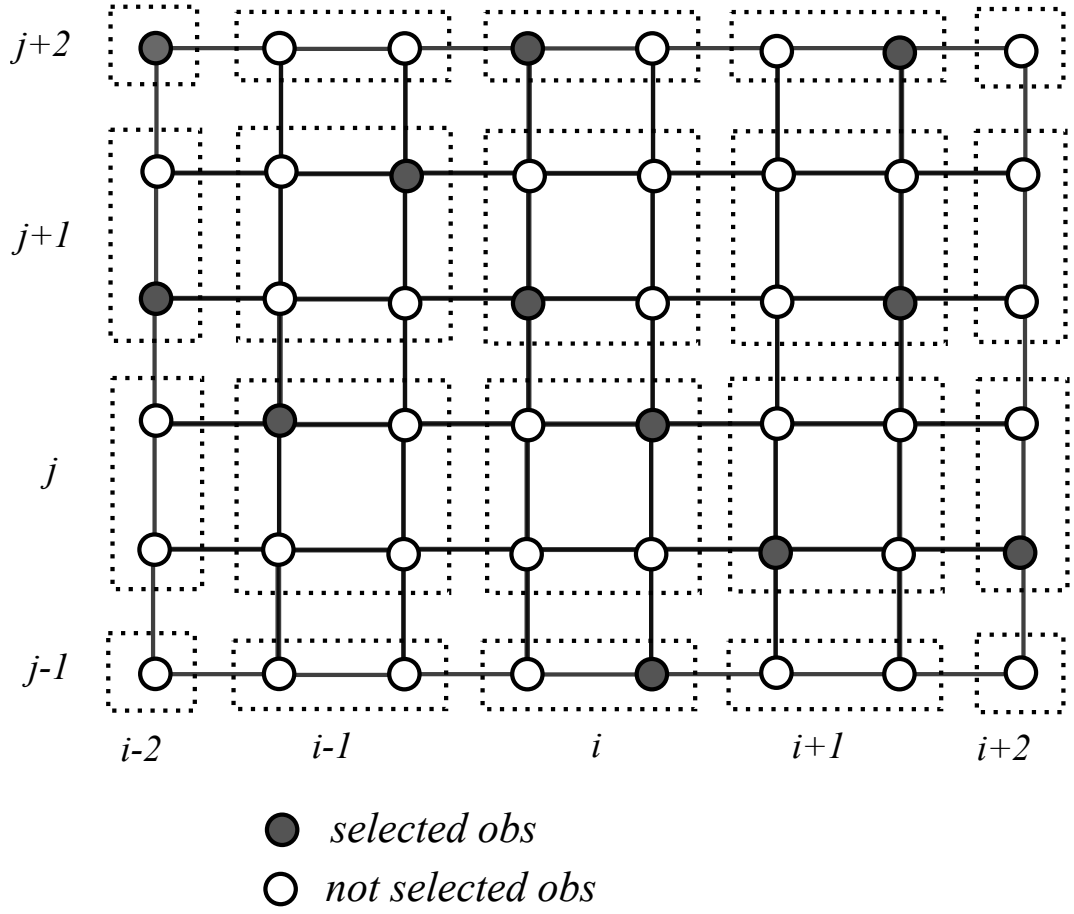

Figure 3: A schematic plot of the perturbation scheme for surface gridded observations, as demonstrated by a realization of perturbed SIC data with $n_{sic} = 1$ and $f_{sic} = 2$. Here circles represent input SIC records, with black-filled circles denoting selected observations after applying the stratified random sampling. i and j are indices for thinning boxes, with boundaries marked by dashed lines. For all thinning boxes not at the boundaries, $M(i, j) = 4$ and $P_t(i, j) = 1$, so one observation is always selected (see Eq.4); For $M(i, j-1) = 2$, $P_t(i, j-1) = 0.5$ so probability of selecting of one observation from the thinning box $(i, j-1)$ would be 50%.

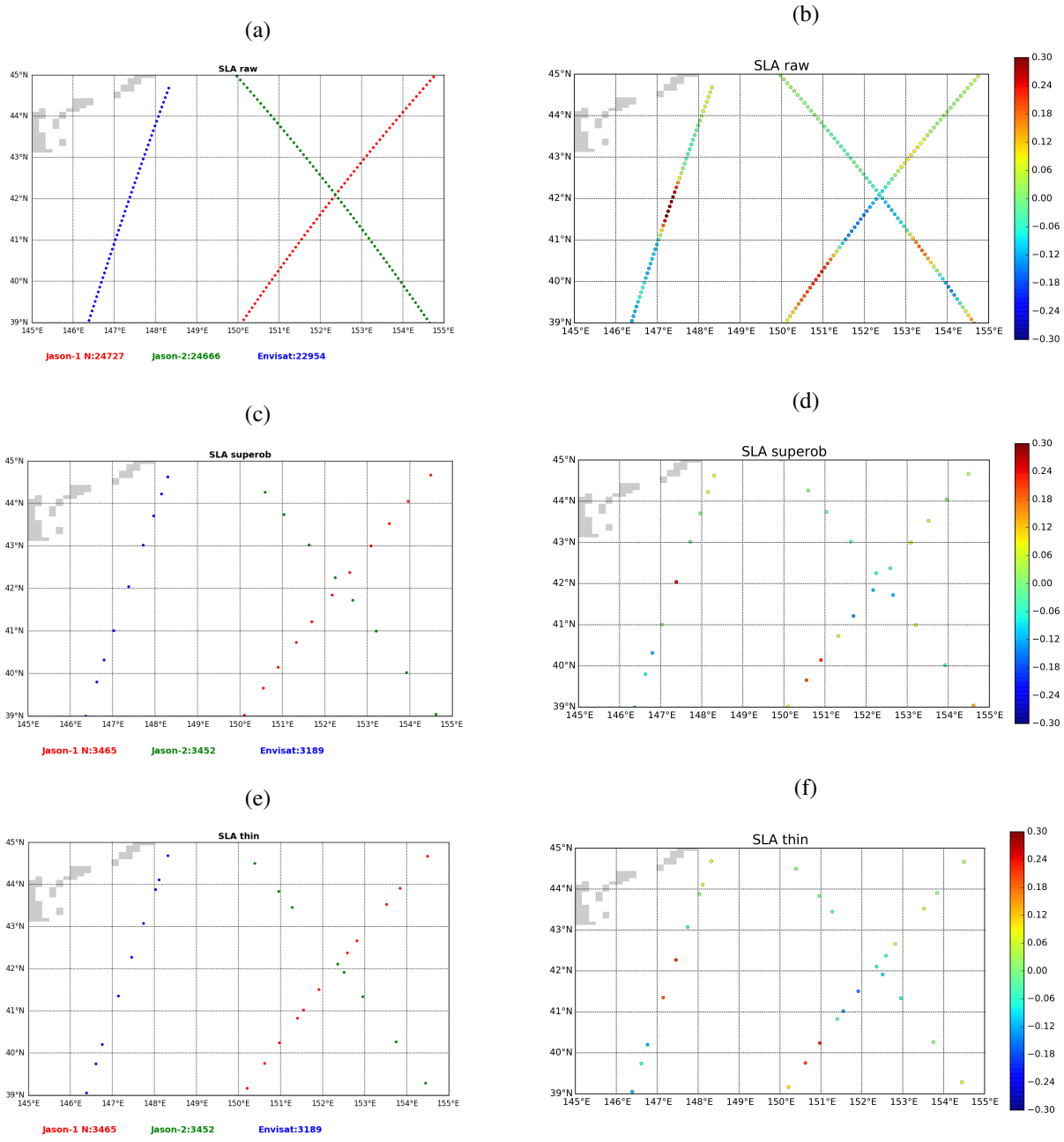

Figure 4: Satellite missions types (a,c,e) and values (in m) (b,d,f) of AVISO along-track SLA observations for an arbitrary chosen date (December 12, 2009) as: (a,b) the raw input data, (c,d) after superobbing and (e,f) thinning using stratified random sampling method. Three satellite missions are included here as Jason-1 new orbit (in blue), Jason-2 (in green) and ENVISAT (in red).

Here this superob scheme is adapted and we introduce a new thinning scheme allowing to account REs of SLA observation more explicitly using a stratified random sampling method. First, a sampling grid is constructed following the same strategy as for the superob grid [Mogensen et al., 2012]. Observations are then organized into sub-groups according to satellite mission types as well as their spatial and temporal locations. Instead of creating a superob based on sample mean of SLA observation values in the sub-group, a random sampling is performed in each sampling grid box and one SLA observation from each satellite mission type is selected as representative of this observation sub-group. The selected SLA observation will be assimilated with its value and space/time position unchanged. As a first attempt, the OBservation Error (OBE) standard deviation is specified in this thinning scheme in the same way as in the superob scheme. It is defined as the standard deviation of the sample, bounded by a minimum value σ_η (here σ_η=0.05 m) to compensate for cases when the observation sample size is too small (see Eq.38 in Mogensen et al. [2012]). For illustration purpose, we show a daily example of the DUACS2014 AVISO (see Pujol et al. [2016]) along-track SLA as the raw input data (Fig. 4a, 4b), after being processed by the superob scheme as in ORAS4 (Fig. 4c, 4d), and by the thinning scheme as implemented in ORAS5 (Fig. 4e, 4f). The spatial resolution of along-track AVISO SLA is approximately 14 km for raw input data, as demonstrated in Fig. 4a with three satellite missions (Jason-1 New Orbit, Jason-2 and ENVISAT). Compared to the raw data, SLA superob-observations or observations after thinning have been significantly reduced (by a factor of ~7) in observation number, to a spatial resolution of approximately 100 km. The same number of valid SLA observations are used in each realization of the thinning perturbation. Unlike the superob scheme, which was designed to reduce the REs by averaging the SLA observations to a scale resolved by the model, the new thinning perturbation scheme tries to take into account the REs of SLA observations explicitly using a stratified random sampling method. Therefore SLA values with larger spatial variability are expected in the thinning SLA samples (Fig. 4f) compared to that in the superobs (Fig. 4d).

3 Assessment of observation perturbations

Here we assess the ensemble estimates from different perturbation methods (See Table. 2) against the diagonal elements of **B** and different independent observation data sets. For consistency, the ensemble estimates were computed using model forecast (background) fields. Assessment of ensemble spread against the diagonal elements of **B** was carried for all in-situ perturbation experiments using observation space diagnostics.

We define the ensemble spread (σ_b^e) as the square root of the ensemble variance with respect to all ensemble members. The ensemble spread arising from applying the above observation perturbation is evaluated against specified BGE standard deviation (σ_b^s) in the current ECMWF ORAS [Zuo et al., 2015a]. In addition, Desroziers et al. [2005] proposed a way to diagnose a-posteriori background and observation error covariances by using innovations and analysis increments and assuming that the observation and background errors are mutually uncorrelated. The BGE variances were diagnosed using the extended Desroziers' method (σ_b^d) as developed by Weaver [2013], which can be applied to a biased system, by removing the spatial mean error. As a result, it can be used as a good reference for checking the statistical consistency of BGE covariance specifications.

Readers are reminded here that the specified BGE variances $(\sigma^s_{b,X})^2$ at observation points correspond to the diagonal elements of univariate block-diagonal **B**. As a result, $(\sigma^s_{b,X})^2$ only accounts for the unbalanced component of the BGE covariances, e.g. $X = T, S_U$ or η_U, where S_U and η_U are unbalanced salinity and sea-level, respectively. In practice $(\sigma^s_{b,X})^2$ at observation points can be estimated using

a randomization algorithm as described by Andersson et al. [2000]. The diagnosed BGE variances ($(\sigma^d_{b,X})^2$) using Desroziers' method, however, account for both balanced and unbalanced components of BGE covariances.

3.1 Experiments with observation perturbations

Assessment of observation perturbations as described above has been carried out using an ORAS5-equivalent configuration but with a reduced model resolution (see Table 1 for common configurations). Perturbation of sea-level observations is an exception as it has been carried out between 2000 and 2004 in ORCA025.L75 configuration [Bernard et al., 2006]. Because an eddy-permitting model resolution is essential for accounting for the influence of sea-level perturbation, the results for the low resolution model runs show that sea-level increment is dominated by balanced contribution from assimilation of in-situ observations. Here perturbations of in-situ observations and surface observations were carried out independently in order to assess contribution from different components in this ensemble generation scheme. Table 2 gives a summary of all sensitivity experiments that were carried out with different perturbation methods. Horizontal perturbations were always applied with **P1** strategy, but with different PD values. Three PD values of 200km ($Hpert200$), 100km ($Hpert100$) and 50km ($Hpert50$) were selected, which represent approximately the super-grid, grid, and sub-grid levels of the error scales with respect to the horizontal resolution of the NEMO model. A thinning factor of $N = 3$ was used in all experiments with vertical perturbations of in-situ observations.

Table 1: Summary of the system configurations for sensitivity experiments

NEMO model	1×1 degree with 42 vertical levels (ORCA1-Z42)
Observation Assim.	temperature and salinity profiles from EN4, SIC from OSTIA [Donlon et al., 2012]. No SST constrains.
Ensemble	5 ensemble members, including one control member that assimilates unperturbed in-situ observations
Period	Experiments cover 2004-2011

3.2 Geographical distribution of ensemble spread

Geographical distributions of ensemble spread (σ^e_b) of temperature and salinity are compared with specified BGE standard deviations (σ^s_b) and diagnosed BGE standard deviations (σ^d_b) using Desroziers' method. These three different estimates of σ_b are compared at observation locations, after being grouped into regular 5°×5° grid, with their representative values derived by averaging all observation point values within the box. For consistency, all three estimates have been evaluated by first computing the variances $(\sigma_b)^2$ using their representative values of the same grid, averaging the variances in space and time, and then taking the square root to obtain the standard deviations. The spatial distribution of σ^e_b from three Hpert experiments are rather similar, except that $Hpert100$ and $Hpert50$ show reduce magnitude compared to $Hpert200$. For simplicity, only σ^e_b of $Hpert200$ together with σ^s_b and σ^d_b at depth 100m is shown here in Fig. 5.

In ORAS4 as well as in ORAS5, BGE standard deviations for temperature are specified in terms of the

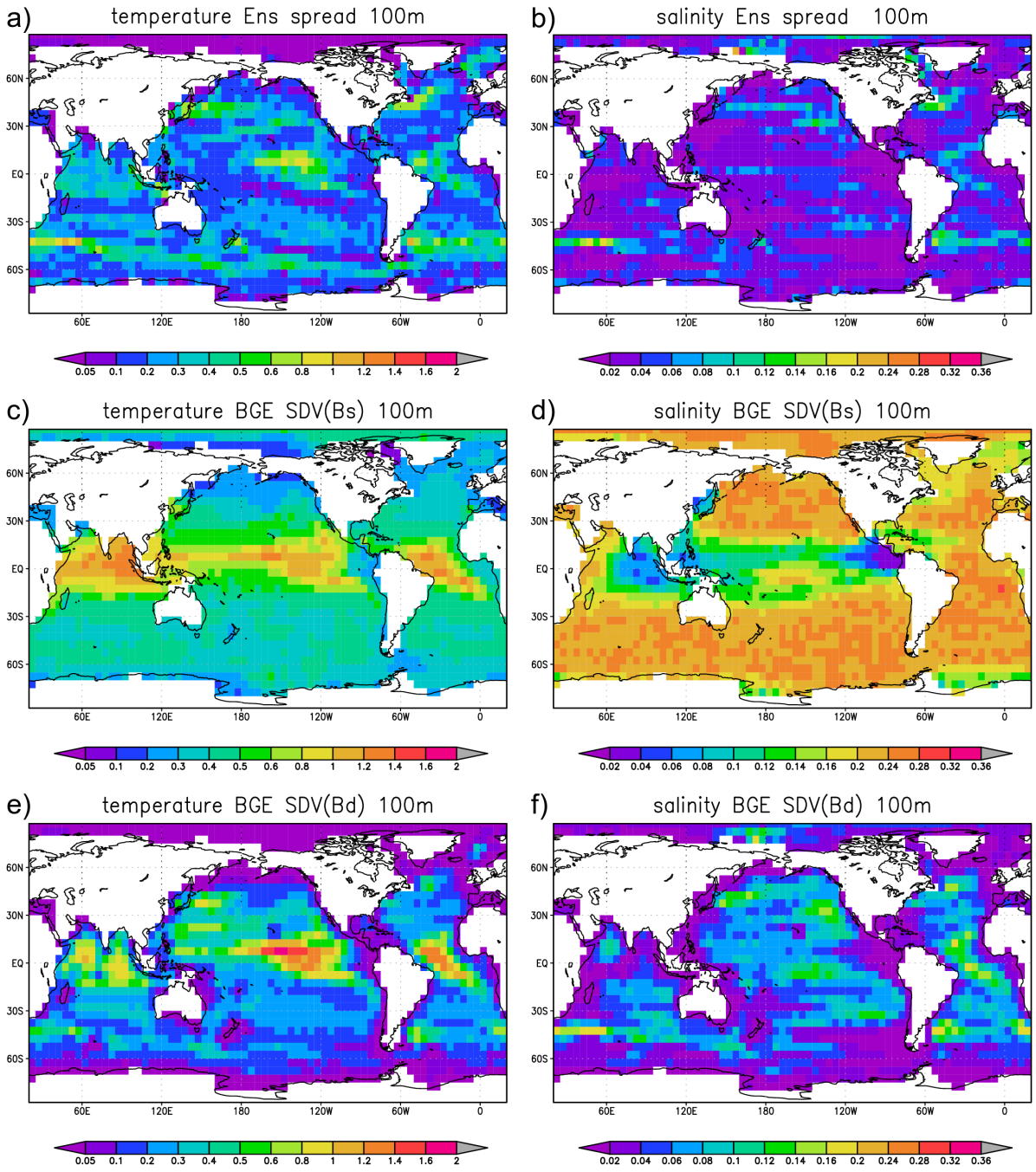

Figure 5: Diagonal elements of B as estimated using the background ensemble spread (σ_b^e) (a,b), specified BGE standard deviations (σ_b^s) (c,d) and diagnosed BGE standard deviations (σ_b^d) (e,f). Shown are the estimates for temperature (in K, left panels) and salinity (in PSU, right panels) at 100m depth. Ensemble spreads were computed as the standard deviations of temperature and salinity background fields, temporally averaged over the 2004-2011 period, and spatially averaged into 5°×5° grid boxes.

Table 2: Summary of perturbation configurations for sensitivity experiments

	experiment	H. pert.	V. pert.	SIC pert.	SLA pert.
In-situ pert.	$Hpert50$	PD=50 km	-	-	-
	$Hpert100$	PD=100 km	-	-	-
	$Hpert200$	PD=200 km	-	-	-
	$HpertS15^{*}$	PD=200 km	-	-	-
	$Vpert$	PD=0 km	$N=3$	-	-
	$HVpert100$	PD=100 km	$N=3$	-	-
	$HVpert200$	PD=200 km	$N=3$	-	-
surf. obs. pert.	$SICpert$	-	-	L_{sic}=~100km	-
	$SLApert^{**}$	-	-	-	L_{sla}=~1°

N is the vertical thinning factor

L_{sic} is the SIC thinning length scale (See Eq. 4) and is ~100km in the Arctic Ocean

L_{sla} is the sea-level thinning length scale and is ~1° in latitude/longitude at the Equator

* $HpertS15$ is the same as $Hpert200$ but with 15 ensemble members

** $SLApert$ is in ORCA025.L75 configuration and integrated over the 2000-2004 period

vertical gradient of model background temperature using an analytical function, and with fixed lower bounds in the mixed layer and in the deep ocean [Mogensen et al., 2012]. This way large $\sigma^s_{b,T}$ occurs in the thermocline where $\partial T^b/\partial z$ is large, e.g. in the Tropics where the background temperature profile is strongly stratified and the thermocline is relatively shallow (Fig. 5-(c)). This parametrization value is sensible as verified against Desroziers' diagnosed $\sigma^d_{b,T}$ (Fig. 5-(e)). However, there are other regions also with large $\sigma^d_{b,T}$ values that were not captured by $\sigma^s_{b,T}$ in our current ODA system, e.g. in the Kuroshio Current extension, Agulhas Current and Falkland Current Loop. The map of temperature ensemble spread (Fig. 5-(a)), on the other hand, shows a similar amplitude and spatial pattern with $\sigma^d_{b,T}$, except for the Tropics where the amplitude of temperature ensemble spread is underestimated compared both to the $\sigma^d_{b,T}$ and $\sigma^s_{b,T}$.

Prescribed BGE standard deviations for unbalanced salinity (σ^s_{b,S_U}) are parameterized in such a way so that its largest value appears between the surface and the level of maximum $\partial S^b/\partial T^b$, and decreases monotonically below this level. It was considered important to specify large values of σ^s_{b,S_U} in the mixed layer [Ricci et al., 2005]. As a result, minimum values of σ^s_{b,S_U} can be found in the Tropical Indian and Pacific as shown in Fig. 5-(d), because mixed layer depths are normally less than 100m in these regions. In contrast, the maps of $\sigma^e_{b,S}$ and $\sigma^d_{b,S}$ (Fig. 5-(b,f)), show a different spatial structure, more similar to that of $\sigma_{b,T}$ in the left panels, as if dominated by the balance component of the error in salinity. This is not entirely surprising in the case of the ensemble estimate, since these experiments lack perturbations to the surface fresh-water fluxes, a large source of uncertainty for the unbalanced salinity, but it does

not explain the structure in the Desroziers' estimate. The magnitudes are also very different among the three estimates. The lower values of $\sigma^d_{b,S}$ and $\sigma^e_{b,S}$ would suggest that σ^s_{b,S_U} as specified in our current ODA system is overestimated for unbalanced salinity. This is, however, subject to the specification of observation errors since Desroziers' diagnostics need to be discussed in the context of the ratio between observation and background errors. Nevertheless, the large discrepancies between the specified σ^s_{b,S_U} and the other two estimates ($\sigma^e_{b,S}$ and $\sigma^d_{b,S}$) is a reason for concern, indicating that the specified σ^s_{b,S_U} values should be revised. In the end, using an ensemble-based estimation of BGE covariances as $\sigma^e_{b,S}$ may lead to a better statistical consistency.

3.3 Vertical profiles of ensemble spread

In order to gain better understanding about the system uncertainty associated with different perturbation strategies, we computed the vertical profiles of σ^e_b for temperature and salinity, together with σ^d_b and σ^s_b. Because in-situ observations assimilated in each ensemble member are different with perturbed positions, this observation-space diagnostics were also carried out by grouping variables in observation locations into regular 5°×5° grid. The representative values for the same grid cell are then used to compute ensemble variances and $(\sigma_b)^2$, before averaging in space and time, and finally taking the square root to obtain the standard deviations. Here, the spatial averaging is performed over different ocean sub-domains (global, northern extratropics: 30°N to 70°N, southern extratropics: 70°S to 30°S and tropics: 30°S to 30°N) within the vertical model grid cells. The time averaging is performed over the 2004-2010 period. The results can be found in Fig. 6 and 7.

Fig. 6 shows vertical profiles of temperature ($\sigma^e_{b,T}$) and salinity ($\sigma^e_{b,S}$) ensemble spread from three $Hpert$ runs and for $Vpert$ (see Table.2 for detailed configurations), as well as comparisons with σ^s_b and σ^d_b for different ocean domains. The ensemble $\sigma^e_{b,T}$ increases over depth in the Tropics (Fig. 6-(e)) until reaching maximum (~0.24°C for $Hpert200$) at thermoclines around 200m. The maximum value of $\sigma^d_{b,T}$, however, is almost twice as large as $\sigma^e_{b,T}$ of $Hpert200$ in the Tropics. Below 300m, $\sigma^d_{b,T}$ and $\sigma^e_{b,T}$ in $Hpert200$ are very close to each other, with only small discrepancies at around 1000m. In both northern and southern extratropics (Fig. 6-(a,c)), $\sigma^e_{b,T}$ from $Hpert100$ shows very close correlation with the diagnosed BGE standard deviation, except for the top 100m where $Hpert100$ spread is smaller than $\sigma^d_{b,T}$.

Salinity ensemble spread decreases monotonically with depth on all the ocean domains. In the Tropics (Fig. 6-(f)) $\sigma^e_{b,S}$ from in-situ horizontal perturbations varies between 0.03 psu ($Hpert50$) and 0.06 psu ($Hpert200$) at the ocean surface. This is smaller than $\sigma^d_{b,S}$ for the upper 300m, where salinity is described predominately by its unbalanced component [Ricci et al., 2005]. Both temperature and salinity ensemble spread drop quickly below 200m, with their magnitudes reduced significantly below 1000m. This is partially because ocean observations are non-uniform and are often sparse and in-complete below this depth range. The other contributing factor is that the reduced vertical gradients, so the RE is reduced. As a result, a limited number of perturbed temperature and salinity profiles cannot effectively maintain an adequate spread in the deep ocean. Spatial variations exist though, e.g. relatively larger temperature ensemble spread exists up to to 2000m in the North Atlantic Ocean (Fig. 7-(left)). As expected, increasing the PD values in $Hpert$ experiments results in an larger ensemble spread for both temperature (Fig. 6-(a,c,e)) and salinity (Fig. 6-(b,d,f)). The vertical perturbation alone ($Vpert$ denoted by black solid lines) introduces a noticeable system uncertainty but smaller than any horizontal perturbation experiments ($Hpert$). Ensemble spread in $Vpert$ is very weak for both temperature (<0.03°C) and salinity (<0.005 psu) at all depths.

Compared with σ^d_b and σ^e_b, the vertical profiles of prescribed BGE standard deviation (σ^s_b) for both tem-

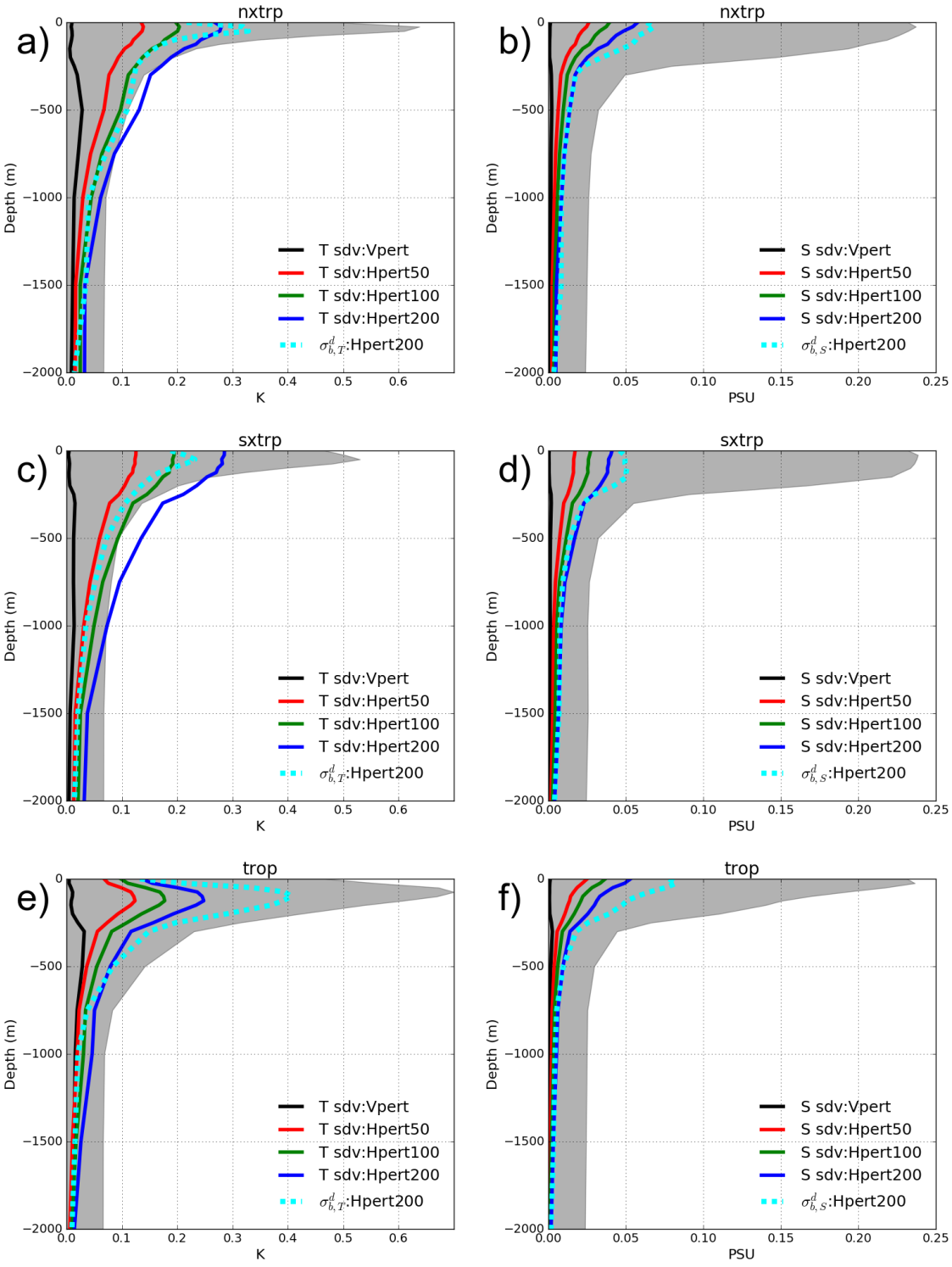

Figure 6: Vertical profiles of σ_b^e for (left panels) temperature (in K) and (right panels) salinity (in PSU) in different perturbation experiments and σ_b^d from $Hpert200$ (Cyan dashed lines), temporally averaged over the 2004-2010 period, and spatially averaged over (a,b) northern extratropics (nxtrp: 30°N to 70°N); (c,d) southern extratropics (sxtrp: 70°S to 30°S) and (e,f) tropics (trop: 30°S to 30°N). σ_b^s from $Hpert200$ (grey shaded areas) is also included here as a reference.

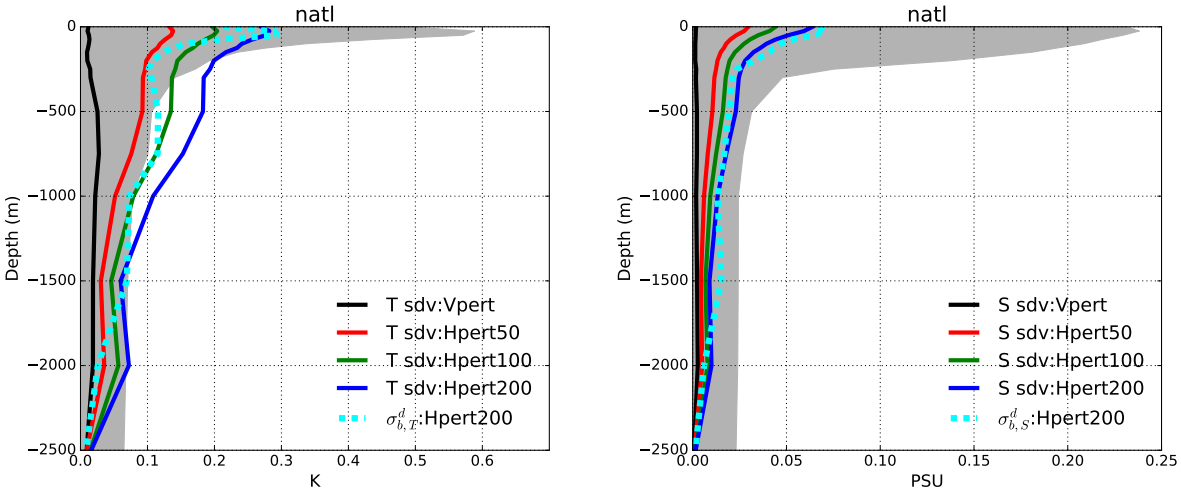

Figure 7: As in Figure 6, but showing the ensemble spread of temperature (left panel, in K) and salinity (right panel, in PSU) vertical profiles for the Northern Atlantic Ocean (between 40^oW to 15^oE and 30^oN to 70^oN).

perature and salinity are more homogeneous across different ocean basins. Similar to the temperature ensemble spread profiles, maximum $\sigma^s_{b,T}$ appears in the thermoclines and varies between 0.55°C-0.75°C, depending on the ocean domains. A constant minimum value of $\sigma^s_{b,T}$ was set to 0.07°C for the deep ocean. Vertical profiles of σ^s_{b,S_U} have similar features, with a maximum value of 0.25 psu and a minimum value of 0.02 psu. In the upper 200m, amplitudes of both $\sigma^e_{b,T}$ and $\sigma^e_{b,S}$ from all four sensitivity experiments are smaller than those of σ^s_b. The ensemble spread is also smaller than the specified background error in the deep/abysal ocean. Since the ensemble generation only relies on observation perturbations to create spread, it is likely that the ensemble spread results in an underestimation of the background error. Inflating the ensemble spread for the upper 200m and applying a minimum value for the deep ocean should be considered in the future when implementing an ensemble based specification of **B** covariances for the NEMOVAR ODA system. A hybrid strategy combing the empirical parametrization with an ensemble spread information for σ_b specification should be exploited as well.

North Atlantic Ocean is quite a unique region when diagnosing ensemble spread generated by perturbing temperature and salinity profiles. Unlike in the tropics and southern extratropics, $Hpert100$ maintained reasonable temperature spread below 100m (See Fig. 7-(left)) with a magnitude similar to $\sigma^d_{b,T}$. It is also worth noting that a second maximum in the temperature spread appears at 2000m, with a value of 0.07°C and is larger than $\sigma^d_{b,T}$. Unlike other regions, temperature ensemble spread from $Vpert$ remains non-negligible up to 2000m, with a mean value of ~0.03°C below 500m. It suggests that relatively large uncertainty is associated with the reduced model vertical resolution in the deep ocean. Our ODA system is less constrained and is very sensitive to the vertical structure of temperature and salinity observations in the North Atlantic Ocean. It could be related to the misrepresentation of the Mediterranean Outflow waters and their propagation over the North Atlantic Ocean, as discussed in Zuo et al. [2015b]. Fig. 8 shows the mean ensemble spread of the model temperature at 2000m for $Hpert50$ and $Vpert$, together with $\sigma^s_{b,T}$ and $\sigma^d_{b,T}$ as derived from $Hpert50$ in 5°×5° grid cells. Here the magnitude of ensemble spread in $Vpert$ (Fig. 7-(b)) is comparable with that in $Hpert50$ (Fig. 7-(a)), both with pronounced temperature spread (up to 0.25K) located in regions strongly affected by the Mediterranean outflow waters. These regions with large model BGEs are also confirmed in the map of $\sigma^d_{b,T}$ (Fig. 7-(d)), where large values

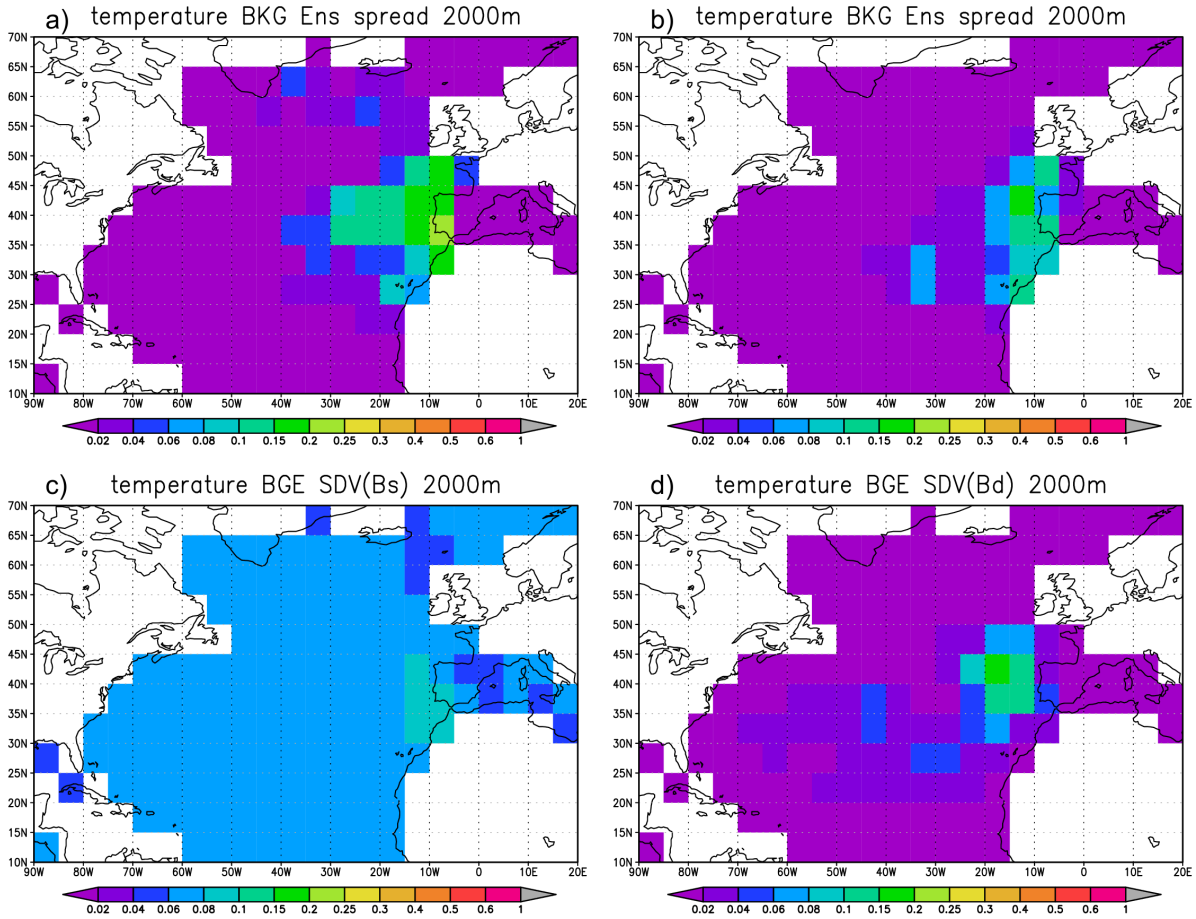

Figure 8: Geographical distribution of temperature ensemble spread (in K) at 2000m in (a) $Hpert50$ and (b) $Vpert$. Specified (c) and diagnosed (d) BGE standard deviations were derived from $Hpert50$. Model ensemble spread was computed as standard deviations of background temperature with respect to all ensemble members, and averaged over the 2004-2011 period.

of $\sigma^s_{b,T}$ (Fig. 7-(c)) are only found in regions very close to the Strait of Gibraltar. Additional attempts have been made in order to reduce model errors in the North Atlantic Ocean, including implementation of a salinity capping scheme and additional Quality Control (QC) check for rejecting spurious salinity observations. Discussion of these other newly developed schemes is beyond the scope of this study, which aims to focus on the perturbation scheme in ORAS5.

Additional experiments have been carried out in order to assess the additive effect of different perturbation strategies when accounting for both horizontal and vertical REs, and to examine sensitivity to ensemble size. These include sensitivity experiments $HVpert100$ and $HVpert200$ in which both horizontal and vertical perturbations are activated (See Table 2), and a rerun of $Hpert200$ with 15 ensemble members ($HpertS15$). All other configurations are exactly the same as in Table 1. Vertical profiles of ensemble spread from these three experiments are shown in Fig. 9. For simplicity we only show vertical profiles in the Tropics for these experiments. Adding vertical perturbations to horizontal perturbations has little impact on the temperature/salinity ensemble spread, which was demonstrated by comparing HVpert200 with Hpert200. However, a small increase in temperature spread was found below 500m with additional vertical perturbation. Increasing ensemble size to 15 in $HpertS15$ increases the temperature ensemble spread at the thermocline, with its magnitude increased by ~20%. Growth of salinity

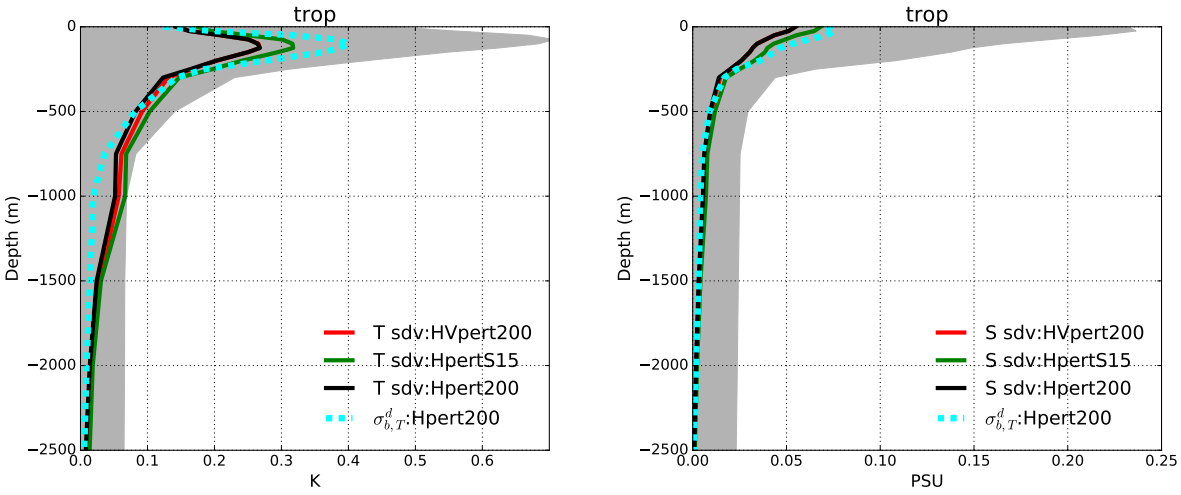

Figure 9: As in Figure 6, but showing the tropical mean ensemble spread of the temperature (in K, left panel) and salinity (in PSU, right panel) profiles from $Hpert200$, $HVpert200$ and $HpertS15$, averaged over the 2004-2006 period. σ^d_b (dashed lines in Cyan) and σ^s_b (grey shaded areas) derived from $Hpert200$ are also shown here.

ensemble spread was also noticed (Fig. 9-(right)), but with lower impact for the upper 300m. Therefore, increasing the ensemble size is desirable considering that both temperature and salinity ensemble spreads are sensitive to the ensemble size and becoming closer to σ^d_b in the Tropical thermocline with larger number of perturbed members.

3.4 Temporal variability of ensemble statistics

The time-varying aspect of ensemble statistics associated with perturbation of in-situ observations has been evaluated with the focus on Ocean Heat Content (OHC). OHC is a common Climate Change Indicator (CCI) which represents the overall temporal evolution of the mean ocean state. Temporal variations of global mean OHCs from $Vpert$ and $Hpert$ experiments at different depth ranges are shown in Fig. 10 as ensemble mean. In spite of different PD values, the ensemble mean OHCs from all three $Hpert$ runs show similar trends both in the upper 700m (dashed curves) and for the whole water column (solid curves). Compared to $Hpert50$ and $Hpert100$, in which PDs are less or equal to the model's horizontal resolution, $Hpert200$ shows enhanced inter-annual variabilities of total OHC and slightly reduced OHCs for the upper 700m.

In order to minimize spurious temporal signals introduced by changing observing system [Zuo et al., 2015a], here the ensemble spread of OHC was computed using model analyses in model grids. This is different from the observation space diagnostics in Section 3.2 and 3.3. Fig. 11 shows time-series of the upper 700m OHC ensemble spread (the square root of the ensemble variance) from four in-situ perturbation experiments. The ensemble spread gives a measure of the uncertainty introduced by REs from temperature and salinity profiles. By construction, larger OHC ensemble spread is associated with larger PD value when in-situ observations are horizontally perturbed. The OHC ensemble spread is approximately proportional to the PD value, with three OHC ensemble spreads of 0.16, 0.35 and 0.81 ($\times 10^7 Jm^{-2}$), for $Hpert50$, $Hpert100$ and $Hpert200$, respectively. There is no obvious temporal trend in the time series of OHC ensemble spread. However strong inter-annual variations occur in the $Hpert200$

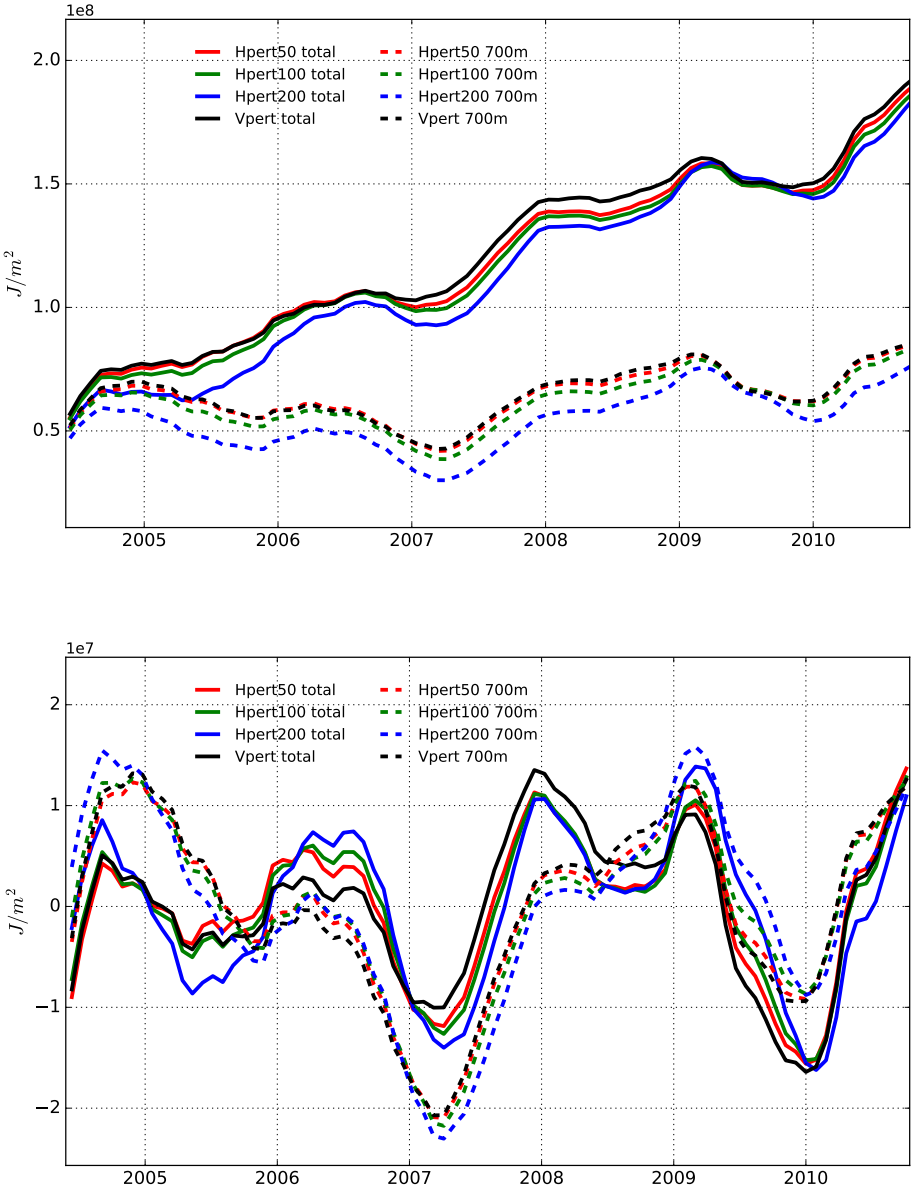

Figure 10: Time series of ensemble mean global Ocean Heat Content changes (in J/m^2) as (top) OHC anomalies with trends and (bottom) de-trended OHC anomalies. Shown are anomalies from $Hpert50$ (red), $Hpert100$ (green), $Hpert200$ (blue) and $Vpert$ (black) after removal of the seasonal signals; integrated for the whole water column (solid curves) and for the upper 700m (dashed curves). OHCs are computed using monthly mean model analyses and are normalized by the global sea surface area. A 12-month running mean smooth has also been applied to time series.

ensemble spread, suggesting some flow-dependent features which are still subject to the changing observing system. The amplitude of the ensemble spread produced by perturbations in the observation spatial distribution in $Hpert200$, although small in comparison with the total OHC trend since 2004 (less than 10% of the total OHC trend), is about 85% of the mean interannual variability ($0.95 \times 10^7 Jm^{-2}$, black dashed line in Fig. 11) of the global OHC during the analyzed period.

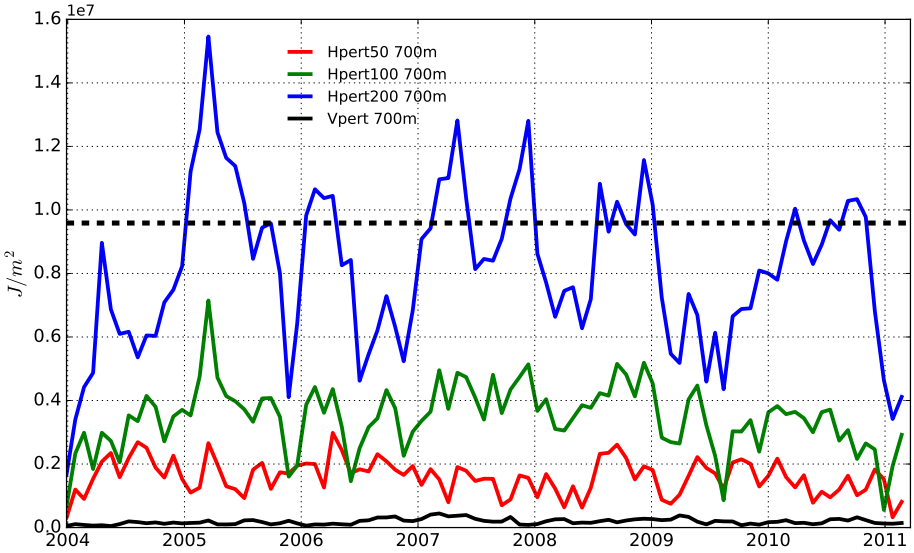

Figure 11: Same as in Fig. 10 but for the ensemble spread of upper 700m OHC. The ensemble spread was computed as the square root of the ensemble variance from 5 ensemble members. Here the mean value of 700m OHC inter-annual variability is also shown here in black dashed line as a reference.

In $Vpert$ where only the vertical perturbation was activated, the inter-annual variability of the ensemble mean OHC is very similar to that in $Hpert50$ (Fig. 10), but with much reduced ensemble spread ($0.2\times10^6 Jm^{-2}$, see Fig. 11) when integrated over upper 700m. As expected, most system variance associated with vertical perturbation resides in the deep ocean below 700m (not shown), due to reduced vertical resolution in both observation profile and model in this region.

3.5 Evaluation against independent data

The reliability of the ensemble has been evaluated using independent observation data sets, with focus on mean ocean state and ensemble variance. First, the mean ocean state was compared to World Ocean Atlas 2013 data set (WOA13, see Locarnini et al. [2013]). WOA13 is an objective analysis data set containing the mean state of global ocean as derived using broad ocean observations. Only results from $Hpert$ experiments are discussed here for simplicity, since these are the major contributor to the ensemble spread, as demonstrated in previous sections. Fig. 12-(top) shows the ensemble mean temperature biases of $Hpert50$ with respect to WOA13. Some of the largest biases are found in the Western Boundary Currents (WBC) and their extensions. With the increase of PD value, the warm bias in the pathway of Gulf Stream was reduced, as demonstrated by the averaged temperature differences between $Hpert100$ and $Hpert50$ in Fig. 12-(middle), and between $Hpert200$ and $Hpert50$ in Fig. 12-(bottom). However it is not always the case and there is no obvious bias reduction in other regions with significant mesoscale eddy activities. It is likely due to the fact that in-situ observations are unevenly distributed over space and time. E.g. in-situ observations are much more densely populated in the Northern Atlantic Ocean than in the Southern Ocean. It is also worth noting that the ensemble mean from any $Hpert$ experiments shows reduced temperature bias with respect to the control member (not shown).

Reliability of the ensemble variance has been verified against independent observation data set, following

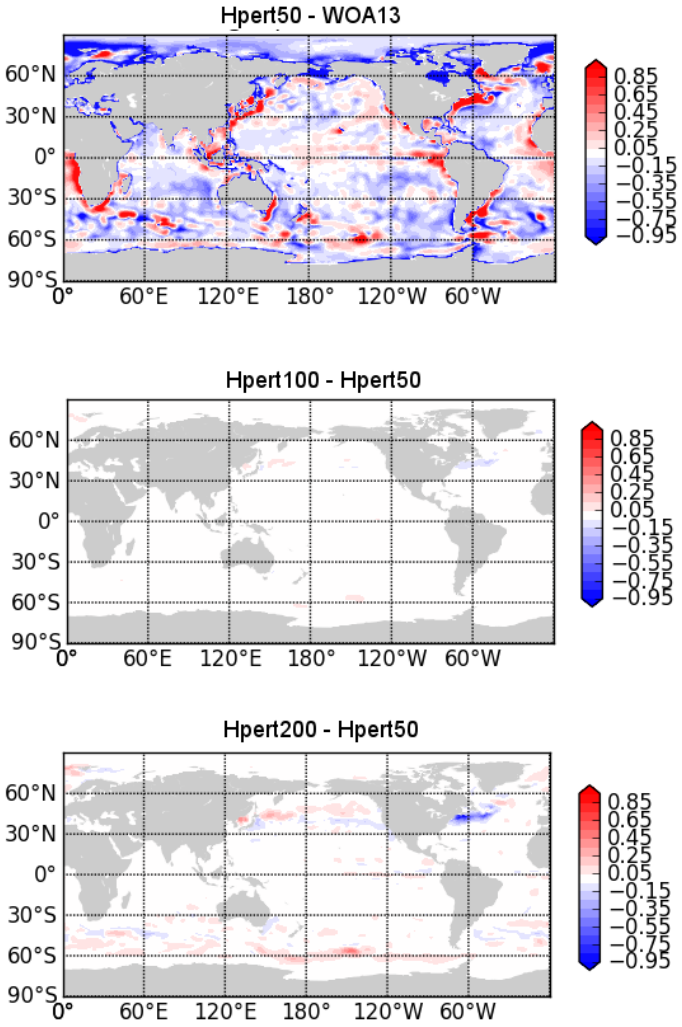

Figure 12: Mean differences in temperature at 800m depth between: (top) $Hpert50$ - WOA13; (middle) $Hpert100$ - $Hpert50$ and (bottom) $Hpert200$ - $Hpert50$. Differences have been computed using ensemble mean model data after averaging over the 2004-2010 period.

a similar approach as that described in Yamaguchi et al. [2016].

$$\frac{1}{N}\sum_{i=1}^{N}(\varepsilon_i^2 - \sigma_{o,i}^2 - \frac{M+1}{M-1}(\sigma_{b,i}^e)^2) \rightarrow 0,\ for\ N \rightarrow \infty \tag{5}$$

Where ε_i^2 is the squared ensemble mean departure against observations, $\sigma_{o,i}^2$ is squared observation error, $(\sigma_{b,i}^e)^2$ is ensemble variance. Here M is the actual ensemble size and the subscribe i denotes the i_{th} sample within the total N independent sample cases. In a perfect ensemble with a M large enough, $\frac{M+1}{M-1} \rightarrow 1$ and $(\sigma_b^e)^2$ should match with $\varepsilon^2 - \sigma_o^2$ in an average sense.

Here the squared ensemble mean errors (ε^2 in Eq. 5) are estimated in observation space using the ensemble mean departure against OSTIA SST, after removal of the mean bias and seasonal cycle signal. The OSTIA SST data was not assimilated in any sensitivity experiments and therefore can be treated as an

independent observation data set. Both OSTIA SST and the model ensemble mean SST are interpolated to a regular 1°×1° grid for computing the ensemble variance and the ensemble mean error. Readers are reminded to keep in mind that both ensemble variance ($(\sigma_b^e)^2$) and ensemble mean error are computed using model analyses instead of model forecasts here, because model forecasts were not archived in our system. Nevertheless, we expect that the analysis variance is a good approximation of the forecast ensemble variance in both spatial and temporal distribution, except for slightly reduced magnitude [Daget et al., 2009]. As a results, comparison as shown below still provides a good estimation of perturbation performance when verified against observation data set.

Verification results from the three *Hpert* experiments are shown in Fig. 13. The observation error of OSTIA SST was estimated following a similar approach to Good et al. [2013], with an averaged value of σ_o^2 of 0.05 K^2 for the Tropics (-20°S - 20°N) and 0.11 K^2 for the Northern Hemisphere (20°N - 70°N), respectively. Fig. 13-(b,d,f) show that $\varepsilon^2 - \sigma_o^2$, when verified against OSTIA SST data, remains almost unchanged regardless of the different PD values used in these experiments. All three experiments show similar geographical distribution pattern in SST ensemble variances. The amplitude of $(\sigma_b^e)^2$ increases from *Hpert*50 (Fig. 13-(a)) to *Hpert*200 (Fig. 13-(e)), following the increases of PD values. There is also similarity between $(\sigma_b^e)^2$ pattern and $\varepsilon^2 - \sigma_o^2$, both with large values at WBCs, the Antarctic Circumpolar Current (ACC) and the Falkland Current Loop. The ensemble variance, however, is in general under-dispersive compared to the $\varepsilon^2 - \sigma_o^2$, even for *Hpert*200 with super-grid PD value.

3.6 Assessment of surface observation perturbations

The generic perturbation scheme for surface SIC observations has been evaluated in experiment *SICpert* (See Table. 2), in which only gridded SIC data from OSTIA was perturbed with 5 ensemble members, including a control member using regular thinning of SIC data. Maps of gridded SIC data valid for assimilation after thinning in the control member and in one of the perturbed members from *SICpert* are shown in Fig. 14.

At the stage of writing this report there is no suitable independent SIC observation data set for assessment of the ensemble variance of SIC. Therefore, we choose to use the same OSTIA SIC product that was assimilated in *SICpert* for verification. One should keep in mind that it is only to provide a guideline for the ensemble variance reliability check and is subject to lack of knowledge about the observation uncertainty from OSTIA SIC data. Here we compare the ensemble spread of SIC in *SICpert* to ensemble mean RMS errors against OSTIA SIC, with the results shown in Fig. 15. Ensemble spread was computed as SIC standard deviation using model analyses, then averaged over the period 2004-2011. As expected, a large ensemble spread of SIC is found mostly in the ice edge regions and near the coast, with a standard deviation normally less than 2 percent. In the Arctic Ocean, maximum SIC ensemble spread can be found near the Canadian Archipelago, in the northern Baffin Bay and along the west coast of Greenland (Fig. 15-(c)), with a maximum value of 5 percent. There are, however, large seasonal variations and SIC spread can reach as much as ~10 percent in the boreal summer, e.g. in the northern edges of the Barents Sea and Kara Sea when the ice is retreating (Fig. 16-(left)). It is equivalent to 50% of SIC uncertainty if normalized against local SIC. There is less SIC ensemble spread in the Antarctic ((Fig. 15-(d)) than in the Arctic Ocean in *SICpert*. If only SIC observations are perturbed, SIC ensemble spread is considered under-dispersive almost everywhere when compared to the SIC ensemble mean RMS errors (Fig. 15-(a,b)).

Additional SIC ensemble spread is expected from surface perturbations that account for uncertainties from both the structural errors and analysis errors of SIC observations, and will be discussed in details in Section 4. It is also worth noting that regardless of the fact that SIC assimilation is treated as univariate in

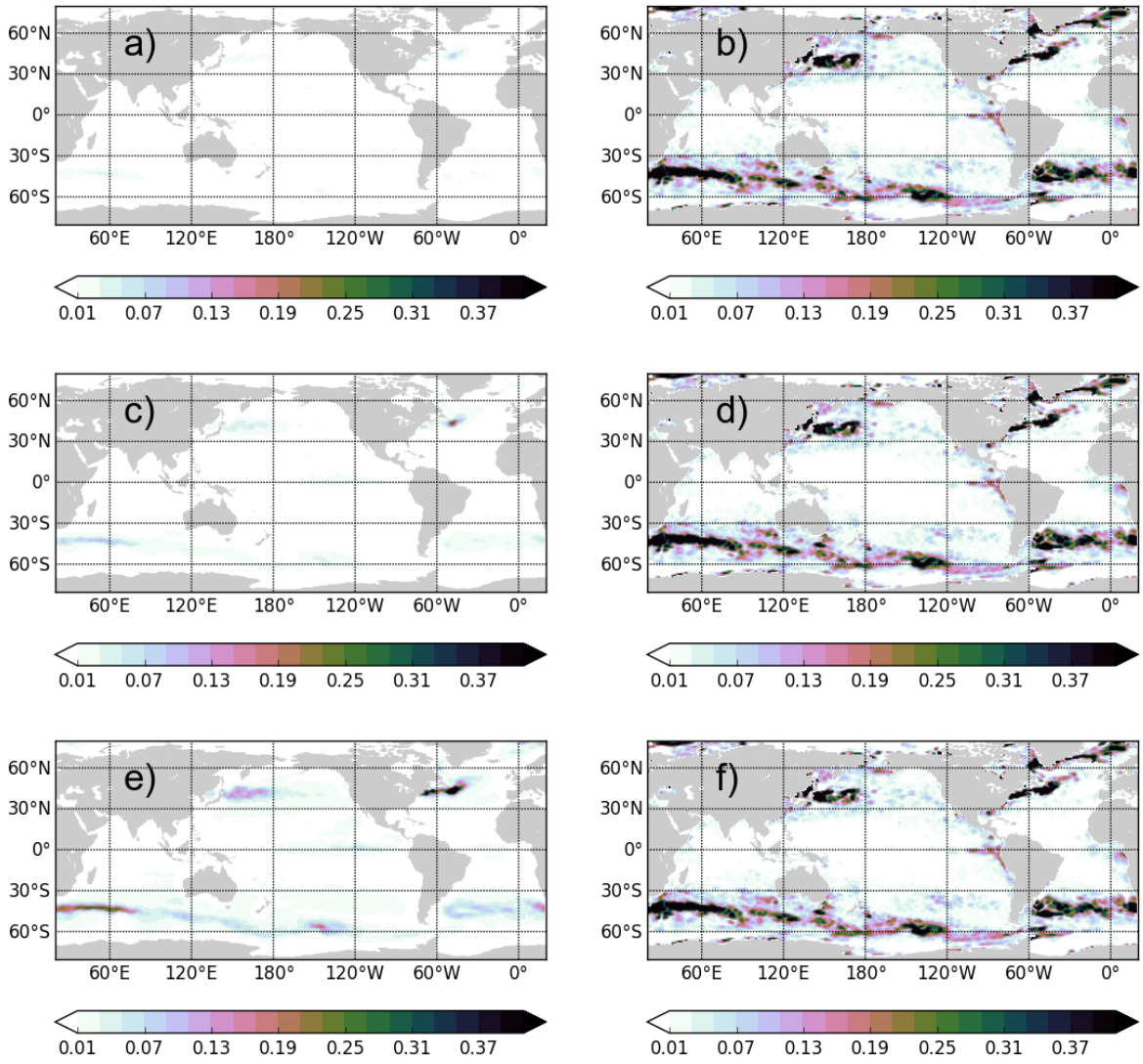

Figure 13: Geographical distribution of (a,c,e) $(\sigma_b^e)^2$ (in K^2) and (b,d,f) $\varepsilon^2 - \sigma_o^2$ when verified against OSTIA SST, for (a,b) $Hpert50$, (c,d) $Hpert100$ and (e,f) $Hpert200$. Both $(\sigma_b^e)^2$ and ε^2 have been computed using monthly mean SST from three perturbation experiments, and averaged over the 2004-2011 period.

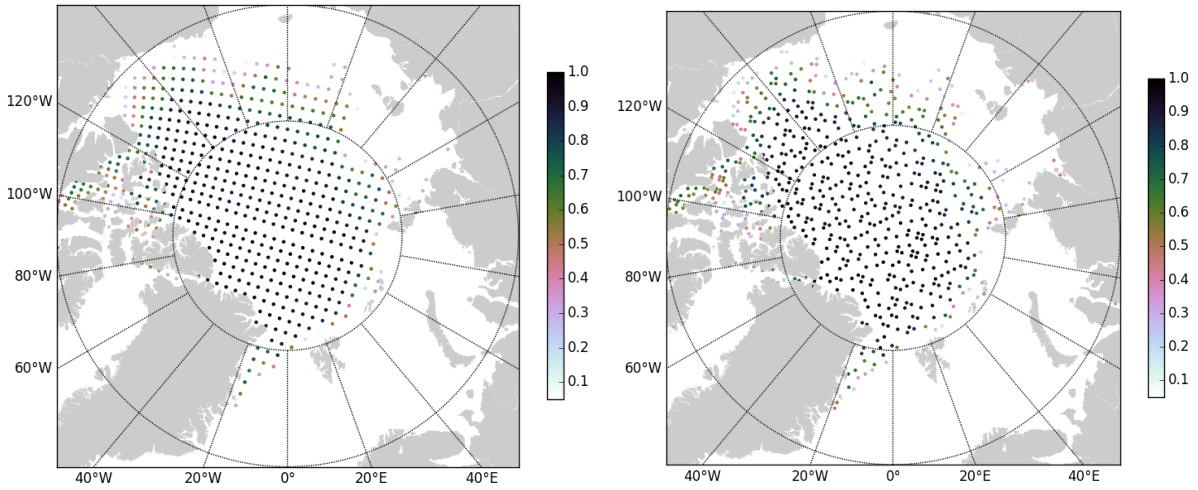

Figure 14: Daily averaged gridded SIC data from OSTIA as assimilated in *SICpert* in (left) the control member with regular thinning; and in (right) a perturbed member using stratified random sampling method with $f_{sic} = 2$ and $n_{sic} = 1$. Here thinning box length-scale is approximately 100 km in the Arctic Ocean.

our current ODA system, perturbation of input SIC data has an indirect impact on the sea ice thickness, especially for areas covered by thick Multi-Year sea ice (Fig. 16-(right)). Furthermore, introducing REs for SIC data affects other ocean state variables, e.g. leads to growth of heat transport uncertainty across the Denmark Strait. It is because the changed sea ice concentration will have an influence on the model first guess for the next assimilation window.

The perturbation of SLA observations using the thinning scheme was evaluated in the experiment *SLApert* (See Table 2). In order to take into account mesoscale eddy activities, *SLApert* is conducted with an eddy-permitting resolution ORCA configuration (ORCA025.L75 with a grid of 0.25° at the equator), and AVISO along-track SLA data is assimilated in additional to other observations (See Table 1 for details). A sampling grid greater than that of the model is constructed, with L_{sla}=~1° in latitude/longitude at the equator. Following the same approach as in Section 3.5, the SLA ensemble variance ($(\sigma_b^e)^2$) in *SLApert* was verified against the gridded Maps of SLA (MSLA) from AVISO (see Pujol et al. [2016]). Here the squared observation error (σ_o^2) in L4 AVISO MSLA was estimated using the same method as in Ducet et al. [2000]. The squared ensemble mean departure against observation (ε^2) was computed using monthly mean SLA fields. By design, this L4 gridded MSLA is correlated with the L3 along-track SLA data assimilated in *SLApert*, and therefore it cannot be considered totally independent. Nevertheless, these diagnostics still provide useful information about regions with large model errors in sea level and can be used as a guideline for checking the SLA ensemble variances reliability. Geographical distributions of $(\sigma_b^e)^2$ and $\varepsilon^2 - \sigma_o^2$ from *SLApert* can be found in Fig. 17 as climatological mean. Regions with large SLA ensemble variance (>0.04 m^2) include the WBCs (e.g. in Gulf Stream and Kuroshio extensions, Agulhas and Malvinas Current regions) and in the ACC pathways. These are the regions with strong mesoscale eddy activity which are not fully resolved in our model (even in ORCA025 resolution), thus corresponding to large REs. $(\sigma_b^e)^2$ and $\varepsilon^2 - \sigma_o^2$ show very similar spatial patterns in general. However, $(\sigma_b^e)^2$ is slightly under-dispersive, e.g. for the East Australian current. Additional sensitivity experiment has been carried out to test SLA perturbation in ORCA1-Z42 configuration, with a sampling grid that is comparable to the model grid. This is effectively reducing specified REs compared to *SLApert*. Magnitude of the SLA ensemble variance in this experiment is almost everywhere much

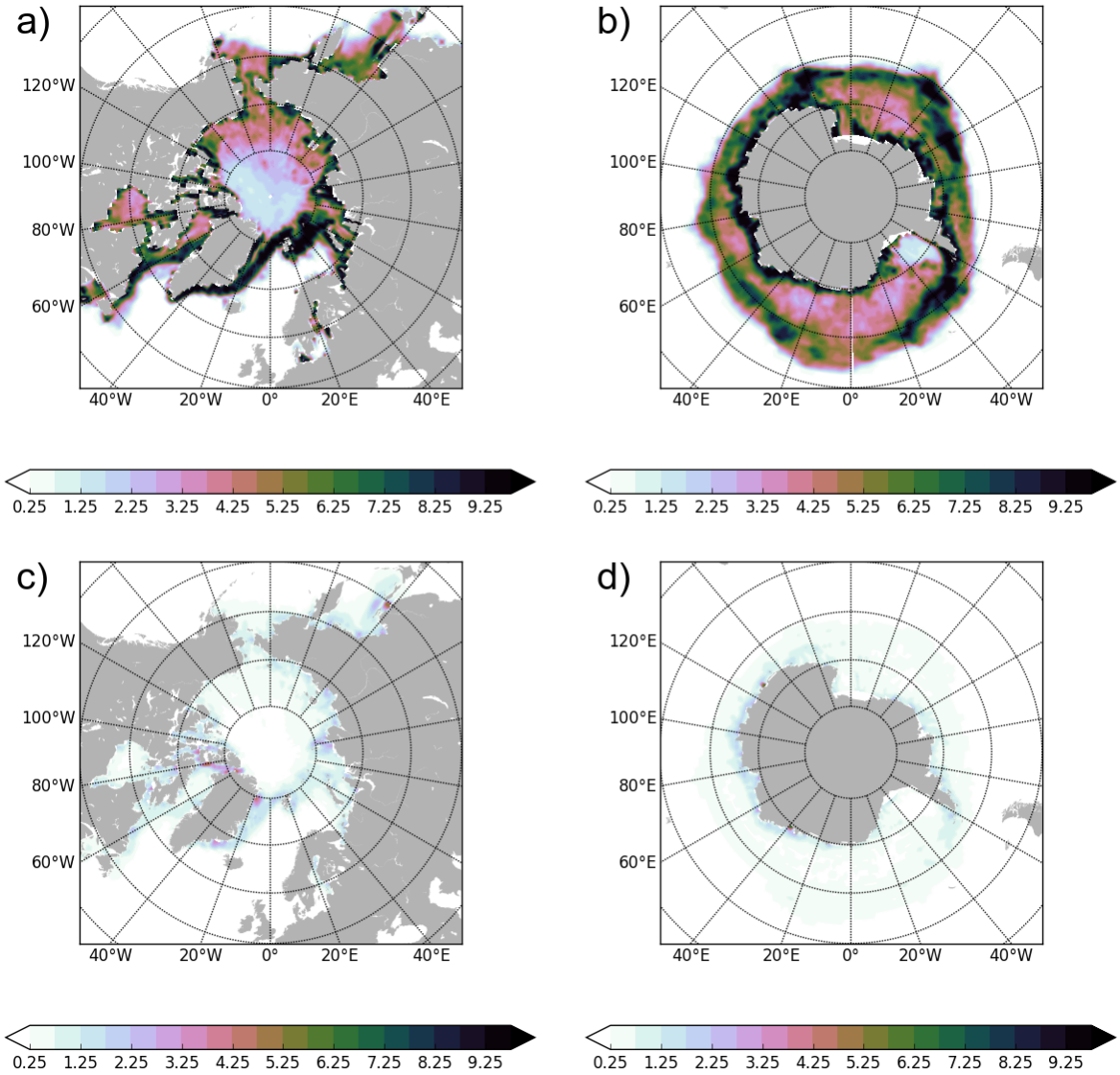

Figure 15: Geographical distributions of (a,b) SIC ensemble mean RMS error (in percentage) as verified against OSTIA gridded SIC, and (c,d) SIC ensemble spread from *SICpert* for (a,c) the Arctic and (b,d) the Antarctic. Both ensemble mean RMS error and ensemble spread were computed using monthly mean SIC analyses from *SICpert* and averaged over the 2004-2011 period.

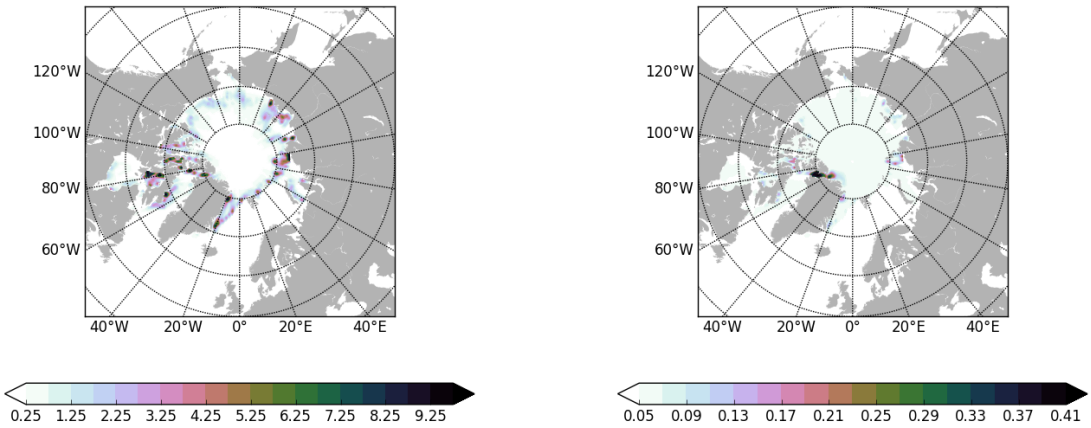

Figure 16: Geographical distributions of ensemble spread from (left) SIC (in percentage) and (right) sea ice thickness (in m) in *SICpert*. Ensemble spread was computed using monthly mean fields in July 2007.

smaller (<0.001 m^2) than that in *SLApert* (not shown). This is mainly because that mesoscale activity is not represented at all in the low-resolution model set-up and therefore SLA assimilation is dominated by contributions from the balance operator (e.g. increments from temperature and salinity). As a result, the sea level variability is strongly constrained by in-situ observations in ORCA1 configuration.

4 Surface forcing perturbations

4.1 The new forcing perturbations scheme

The use of forcing or surface perturbations in the production of ocean reanalysis was first introduced at ECMWF in the context of the seasonal forecasting S2 [Vialard et al., 2005]. The perturbations were applied to sample the uncertainty in the ocean initial conditions. This first version of forcing perturbations ($v1$ in what follows) consisted of SST and wind stress (TAU, hereafter) perturbations. Wind stress perturbations were added to the forcing fields during the production of the ocean reanalysis, resulting in spread in the ocean subsurface, especially along the thermocline. SST perturbations were applied within the ocean mixed layer at the beginning of the coupled forecasts. The SST perturbations were simply differences of monthly anomalies (i.e., the differences in the mean seasonal cycle were removed) between two data sets. Thus, the SST perturbations were differences between the SST OIv2 [Reynolds et al., 2002] and the SST used in ERA40 [Uppala et al., 2005], which came from an early version of the NCEP SST product. For any given calendar month, the perturbation pattern was randomly selected from a sample repository constructed from monthly SST anomalies differences that covered the period 1982-2002 and stratified by calendar month. For each ensemble member and date, a random number was generated and was used for picking the perturbation pattern. By repeating the procedure, this scheme was able to produce an arbitrary number of ensemble realisations. Similar approach was taken in order to construct the repository for wind stress perturbations, using monthly differences in the wind stress anomalies from two historical reconstructions. In this first version of the perturbations, monthly mean wind stress anomalies differences between the estimates from Southampton Oceanographic Center (SOC) and the

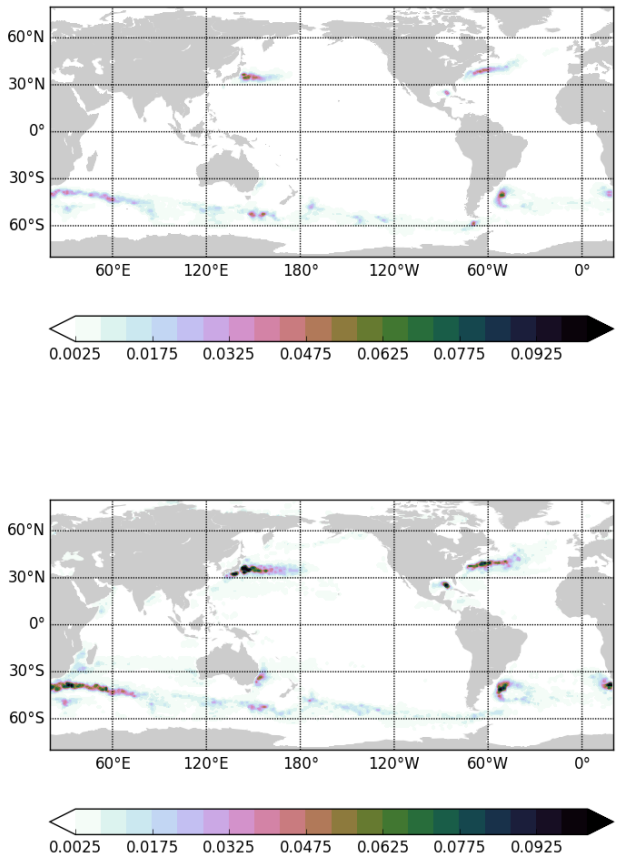

Figure 17: Geographical distributions of (top) $(\sigma_b^e)^2$ of SLA (m^2) and (bottom) $\varepsilon^2 - \sigma_o^2$ when verified against AVISO DUACS2014 MSLA. Both SLA ensemble variances and ε^2 have been computed using monthly mean SLA analysis from *SLApert*, and averaged over the 2000-2004 period.

ERA-15 reanalyses were used. The differences sampled the 1979-1993 time span.

Later on, with the introduction of the seasonal forecasting system S3 [Anderson et al., 2007], the perturbation data set was updated to version $v2$, which was also used for the EU funded ENSEMBLES project. In this revised version, the wind stress perturbation repository was updated to represent the differences between the ERA40 and NCEP-CORE atmospheric reanalysis, which allowed longer records (1958-2002) and therefore larger number of independent realizations. The SST perturbation repository was also increased to include differences between the historical data sets HadISSTv1 and ERSST (NOAA Extended Reconstruction of monthly SST). These SST perturbations and methodology were later used in the atmospheric EDA (see document in http://www.ecmwf.int/sites/default/files/docu_perturbations.pdf).

Although the $v2$ perturbation method is generic enough, the error patterns are now considered to be obsolete since they probably over-estimate uncertainties for recent SST and wind stress products. The $v2$ perturbations are based on monthly fields. Representing uncertainties in shorter time scales is considered important, especially for the medium range forecasting and for the imminent atmospheric reanalysis ERA5, which also uses an EDA approach. There is also a need to extend perturbations to the sea-ice concentration and to other surface fluxes, like the surface fresh water flux from Precipitation and Evaporation (P-E) and Solar Radiation (SR). It is also considered important to preserve the multivariate relationships between the perturbed variables.

In view of these shortcomings, we decided to create a new perturbation scheme ($v3$) based on more recent observational data sets, such as the 10 ensemble members ERA-20C atmospheric reanalysis [Poli et al., 2016], the high resolution SST and SIC analysis (ESA-CCI), and the 10 ensemble members HadISSTv2 [Titchner and Rayner, 2014]. A summary of this new $v3$ perturbation scheme can be found in Table 3. Designed for ensemble generation of ocean reanalyses, this perturbation scheme $v3$ has been applied in the production of ORAS5. It has also been used to provide SST/SIC perturbations in ERA5 production, as described in Hirahara et al. [2016]. Since Cycle 43R1, this new perturbation scheme has been introduced to generate the operational EDA.

In general, a perturbation is a k-dimensional vector $\varepsilon = (\varepsilon_1, \varepsilon_2, ...\varepsilon_k)$, representing the uncertainty of k different variables. Both surface state variables (e.g. SST and SIC) and surface fluxes (e.g. TAU, P-E, and SR) are included in the $v3$ perturbation. Another major update in $v3$ is that we distinguish between two fundamental kind of uncertainties: Structural Error (SE), which is obtained by comparing data sets produced by different methods; and the Analysis Error (AE), which is the error estimated by a single analysis method such as the EDA. Sometimes, the uncertainty from one variable can propagate to others (Multivariate Error, hereafter ME) when considering multivariate relationship among variables. For instance, the uncertainty in SIC can propagate to SST (this does not appear explicitly in Table 3).

Each kind of uncertainty is characterized by the data sets and the temporal records used to build the perturbation repository. The size of the temporal record and number of data sets (or ensemble members) will determine the sample size of the perturbation repository, as discussed below. A given data set can sometimes allow for different temporal sampling (or temporal decorrelation scale δ). We denote $\varepsilon_{\delta,i}(d)$ a single realisation of the estimated error valid for a date d, with a given temporal decorrelation scale δ, arising from a given source of uncertainty ($i = SE, AE, ME$). Here both the temporal correlation and source of uncertainty are considered independent, therefore the total perturbation $\varepsilon(d)$ at any given time d can be expressed as the sum of the individual perturbations as:

$$\varepsilon(d) = \sum_{i=1}^{N} \sum_{\delta=1}^{n} \varepsilon_{\delta,i}(d) \tag{6}$$

Table 3: Summary of Forcing Perturbations *v*3

Variable	**Source**	**Temporal Correlation**	**Time Record and Size** ($N^j_{\delta,i}$)	**Sample Size** ($N_{\delta,i}$)
		Structural Error (SE)		
TAU univariate	Bulk 1: NWPatm - $< Bulk >$	monthly	2006-2013 (8x2)	144 (8x2x9)
	Bulk 2: EIatm - $< Bulk >$	monthly	2006-2013 (8x2)	
	Bulk 3: EIwaves - $< Bulk >$	monthly	2006-2013 (8x2)	
	Bulk 4: EIcore - $< Bulk >$	monthly	2006-2013 (8x2)	
	NCEP - ERA40: time period 1	monthly	1993-2000 (8x2)	
	NCEP - ERA40: time period 2	monthly	1985-1992 (8x2)	
	NCEP - ERA40: time period 3	monthly	1977-1984 (8x2)	
	NCEP - ERA40: time period 4	monthly	1969-1976 (8x2)	
	NCEP - ERA40: time period 5	monthly	1961-1968 (8x2)	
SST univariate	$< HadISSTv2.1 >$ - ESACCI	monthly	1997-2006 (10x2)	20
SIC univariate	$< HadISSTv2.1 >$ - $< HadISSTv2.0 >$	monthly	1997-2006 (10x2)	20
		Analysis Error (AE)		
SST*/P-E /SR/TAU multivariate	ERA20C-i** - $< ERA20C >$ i=1, ..10	monthly	1979-2006 (28x2) i=1, ..10	560 (28x2x10)
	HadISSTv2-i - $< HadISSTv2 >$ i=1, ..10	pentad	1997-2006 (10x2) i=1, ..10	200 (10x2x10)

* SST *AE* repository from HadISSTv2

** HadISSTv2 SST is used in production of ERA20C

where $\varepsilon_{\delta,i}(d)$ is an error realisation taken randomly from the i-th category of uncertainty, with temporal decorrelation scale δ and at a date d. Here $n \geq 1$ and δ can be monthly and/or pentad (see Table 3); $N = 2$ and $i = SE, AE$. We will discuss the different uncertainty categories below, but before it is pertinent to describe the temporal aspects of the perturbations.

4.2 Temporal Decorrelation Scales and Temporal Interpolation

Error patterns can have different characteristic temporal resolution δ (monthly or pentad in v3) depending on the temporal resolution of original products, or depending on the application. Monthly perturbations are obtained by computing monthly means and removing the 12-month climatological seasonal cycle. The pentad temporal resolution is obtained in a similar manner. Perturbations with different temporal decorrelation scales are considered additive. On the other hand, the temporal decorrelation scale of a perturbation is not the same as the temporal discretization of the application, which often requires daily fields. To generate daily perturbation values from a pentad or monthly repository, a variance preserving time interpolation is applied. For a given date d, random patterns representative of the two closest

adjacent dates, d^- and d^+, are drawn and then interpolated to the date d as follows:

$$\varepsilon_{\delta,i}(d) = w^- \varepsilon_{\delta,i}(d^-) + w^+ \varepsilon_{\delta,i}(d^+) \tag{7}$$

The values d^- and d^+ represent the centre of the temporal interval (day 15th for monthly and day 3rd for pentad). In the case of monthly perturbations, they represent indices corresponding to previous (d^-) and current (d^+) calendar months, if d is before day 15th; or to current (d^-) and next calendar (d^+) months, if d is after day 15th. The temporal weights w^- and w^+ are proportional to the distance between dates and are normalized in the quadratic sense, such as ${w^-}^2 + {w^+}^2 = 1$:

$$w = \frac{d - d^-}{d^+ - d^-} \tag{8}$$

$$w^- = \frac{1 - w}{\sqrt{w^2 + (1-w)^2}} \tag{9}$$

$$w^+ = \frac{w}{\sqrt{w^2 + (1-w)^2}} \tag{10}$$

The quadratic normalization is chosen in order to homogenise the error spread over time. A standard bilinear interpolation would result in an artificially reduced spread when $w = 0$, e.g., whenever $d = d_+$ or $d = d_-$. This procedure also allows preserving the temporal correlation of the original sampling in our daily perturbation record. In the previous version *v*2 the temporal interpolation was standard bilinear.

4.3 Multiple Variables and Sources of Uncertainty

As for previous perturbation schemes, the perturbations in *v*3 are sampled from different data sets, which provides a natural way of specifying the spatial and temporal correlation of the perturbations. In *v*3 the range of perturbations has been substantially increased. The source of *v*3 perturbations for SST and SIC are the same as those described in Hirahara et al. [2016] for the pentad perturbations used in ERA5. The temporal resolution has been extended to include monthly time scales in *v*3, which is more relevant for ocean reanalysis applications. For consistency, all the perturbation repositories in *v*3 are interpolated into a common 1°lon/lat regular grid.

Table 3 gives details about the data sets and sample sizes of the perturbation repositories. An individual perturbation can be applied either with a + or a − sign, and therefore the effective sample size of an individual data record j as $N^j_{\delta,i}$ in Table 3 is twice the number of years in the record. When multiple data sets ($j \geq 1$) are available for a given source of uncertainty i and temporal resolution δ, the total sample size $N_{\delta,i}$ is the sum of the sample size of the individual data records, as if we were dealing with different ensemble members

$$N_{\delta,i} = \sum_j N^j_{\delta,i} \tag{11}$$

When additive perturbations are considered, the total sample size is the product of the different additions

$$N = \prod_{\delta,i} N_{\delta,i} \tag{12}$$

As mentioned before, *v*3 considers two fundamental types of uncertainty (*AE*, *SE*). *SE* is provided for three variables: SST, SIC and TAU. These are univariate, although when applying SIC perturbations the total SST field would be adjusted to impose consistency between SST and SIC (the total SST should be at freezing point in the presence of a given threshold of SIC). The univariate SST perturbations have been created by sampling differences between HadISSTv2.1 ensemble mean and ESACCI spanning the period 1997-2006, with pentad and monthly temporal discretization. Hirahara et al. [2016] discuss that these perturbations are commensurable with the uncertainty in the OSTIA SST product, which is the main SST data set from 2008 onwards. The SIC perturbations are monthly, obtained as the differences between HadISSTv2.1 and HadISSTv2.0. The main difference between these two products is in the SIC component, being Rayner et al. [2003] and Titchner and Rayner [2014] (hereafter RA03 and TR14) in v2.0 and v2.1, respectively. RA03 is based on a microwave-derived sea ice from the Goddard Space Flight Center (GSFC) and the National Centers for Environmental Prediction (NCEP). In TR14 the reprocessed SIC from Ocean and Sea Ice Satellite Application Facilities (OSI-SAFr) is bias-adjusted toward ice charts from the US National Snow and Ice Data Center (NSIDC), and connected to chart analyses in the past. These two products are inter-calibrated and applied a melt-pond bias correction before being combined. The *SE* SST and SIC perturbations span the period 1997-2006, and therefore each of them contains 10×2 independent samples per calendar month (or pentad), when considering the arbitrary $+/-$ sign.

The *SE* TAU perturbations are univariate vector fields, with coherent values in the zonal and meridional components of the wind stress error, but they do not have correlations with any other variable. The temporal resolution is monthly. The existing *v*2 repository (differences between ERA40 and NCEP) is still used - in a different manner as in *v*2, see below - but more recent data sets are exploited to create the repository (See Fig. 19 for differences between *v*2 and *v*3). In addition, the SE TAU perturbation repository in *v*3 contain records of wind stress monthly fields spanning the period 2006-2013, estimated in four different ways:

- Daily fluxes from ERA-Interim reanalysis (EIatm, 0-24h forecasts)
- Daily fluxes from Operational Numerical Weather Prediction (NWPatm, also the 0-24h forecast)
- Daily wind stress using a bulk formulation that takes into account the ocean currents and the drag coefficient from the atmospheric waves (EIwaves, following Breivik et al. [2015])
- Daily wind stress using a modification of the CORE bulk formulation (EIcore, see Janssen [2008]), also using ERA-I 6-h analysis variables.

Departures of these different estimates with respect to their ensemble mean $< Bulk >$ are referred as "Bulk 1,2,3,4" in Table 3, and are treated as four different ensembles of 8-year length each. The old perturbations *v*2 are divided in 5 periods of 8-year length each, each period being treated as a separate ensemble in the repository, with the most recent period having the lowest time period number in Table 3. In total there are 9 ensemble members in the *SE TAU* repository, each containing 16 (8x2) realizations. By separating the period into different ensemble members it is possible to include some low frequency variability in the perturbations when sampling the repository. In practice, perturbations in ORAS5 before 1990 sample the 9 ensemble members of *SE TAU* repository, which implies using differences between ERA40 and NCEP from the whole 1960-2000 period. For the 1990-2000 period, ORAS5 perturbations use only the first 7 ensemble members (excluding time period 4 and 5), implying that only NCEP-ERA40 differences since 1977 are utilized. After 2000, ORAS5 does not use any ERA40-NECP differences. Choice of the ensemble members as above has been made based on the assumption that these low frequency variabilities in model error are only important for the period before 2000 in ORAS5.

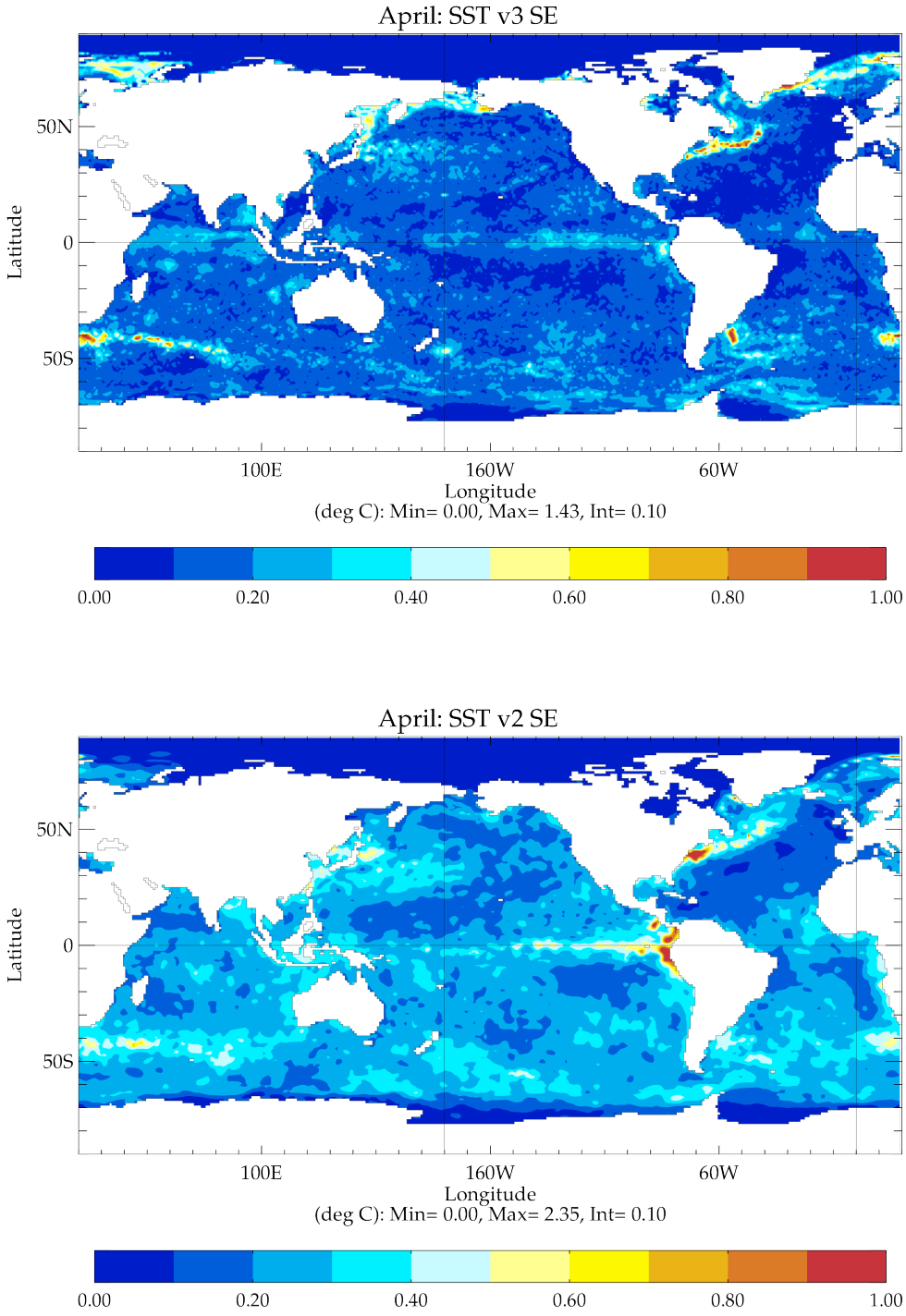

Figure 18: Standard deviation (sdv) of the SST *SE* perturbations in *v*3 (top) and *v*2 (bottom). The perturbations in *v*3 have smaller spatial scale than in *v*2, and are largest along the eddy active regions in the ocean. In contrast, the SST uncertainty in *v*2 is larger along the Eastern Equatorial Pacific.

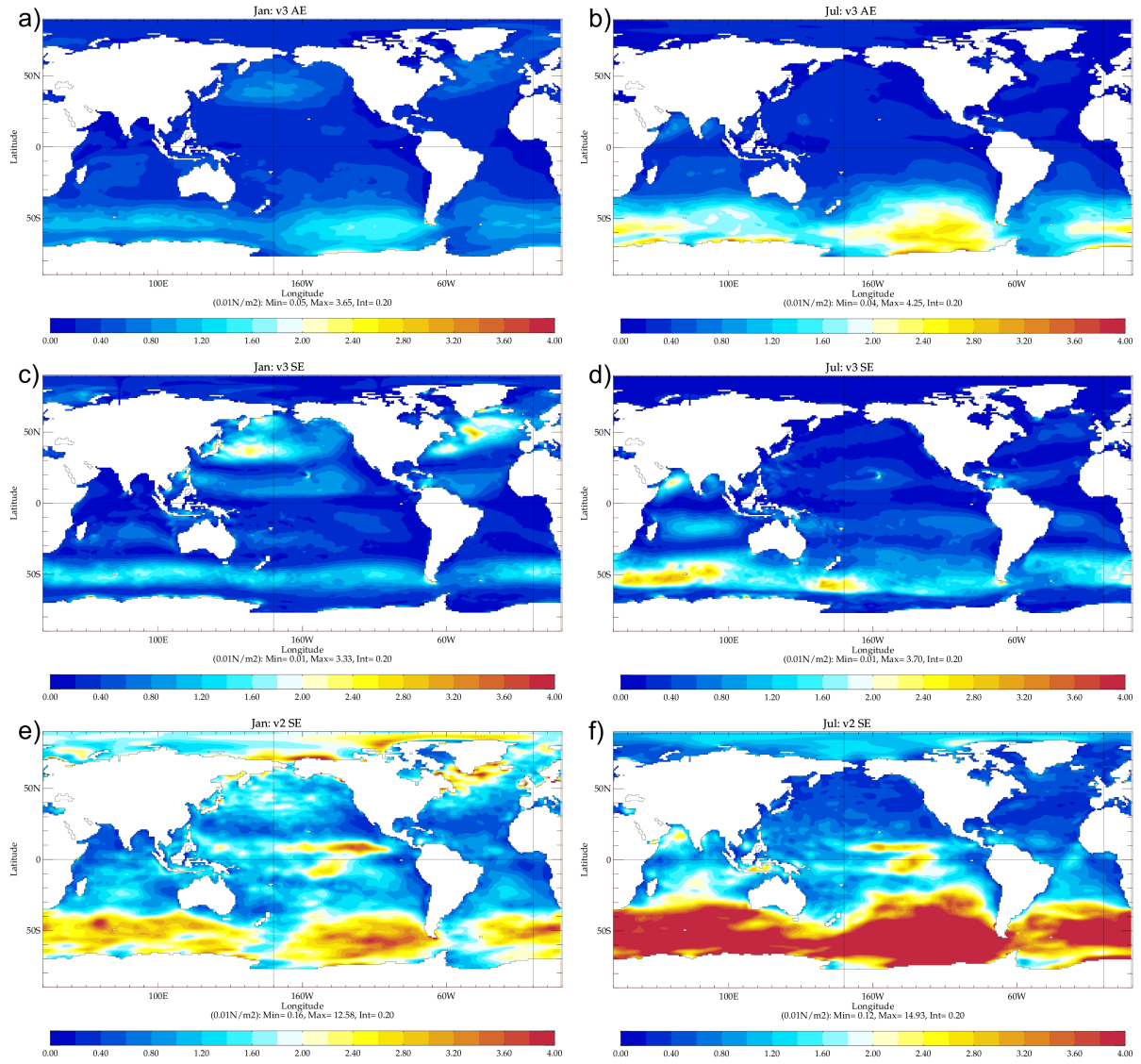

Figure 19: Standard deviation of the zonal wind stress perturbations for January (a,c,e) and July (b,d,f). The analysis and structural uncertainty in *v*3 appears in (a,b) and (c,d), respectively. (e,f) show the standard deviation of perturbations in *v*2 as *SE*. The analysis uncertainty (*AE*) in *v*3 occurs mainly in the Southern Ocean. The *v*3 structural uncertainty (*SE*) affects mainly the gyres, and shows a clear seasonal dependence. The wind stress perturbations in *v*2 are much larger than these in *v*3, e.g. in the Eastern Tropical Pacific and almost everywhere in the southern extratropics. It is commensurable with the zonal wind stress interannual variability.

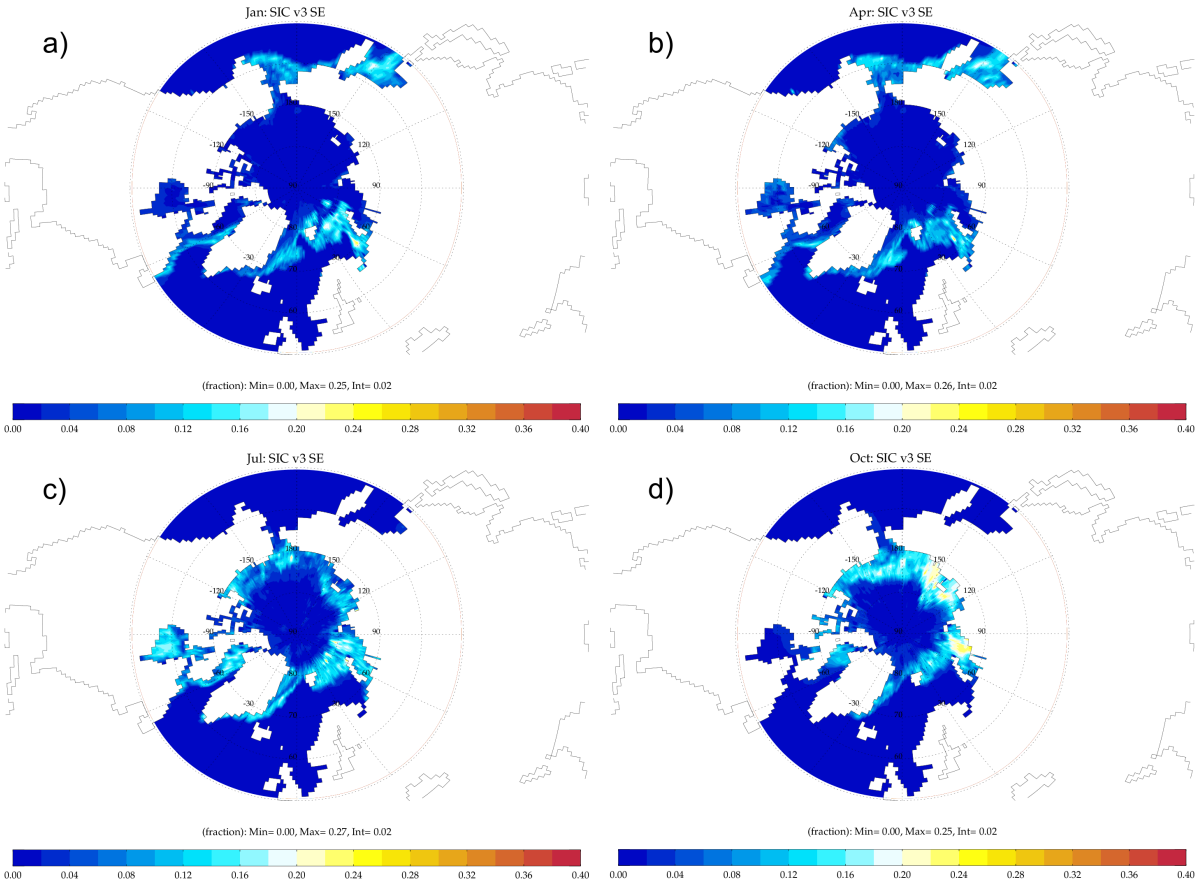

Figure 20: Standard deviation of the SIC *SE* perturbations in *v*3 for January (a), April (b), July (c) and October (d). There are clear seasonal variations in the SIC *SE* perturbation, in both the spatial distribution and magnitude.

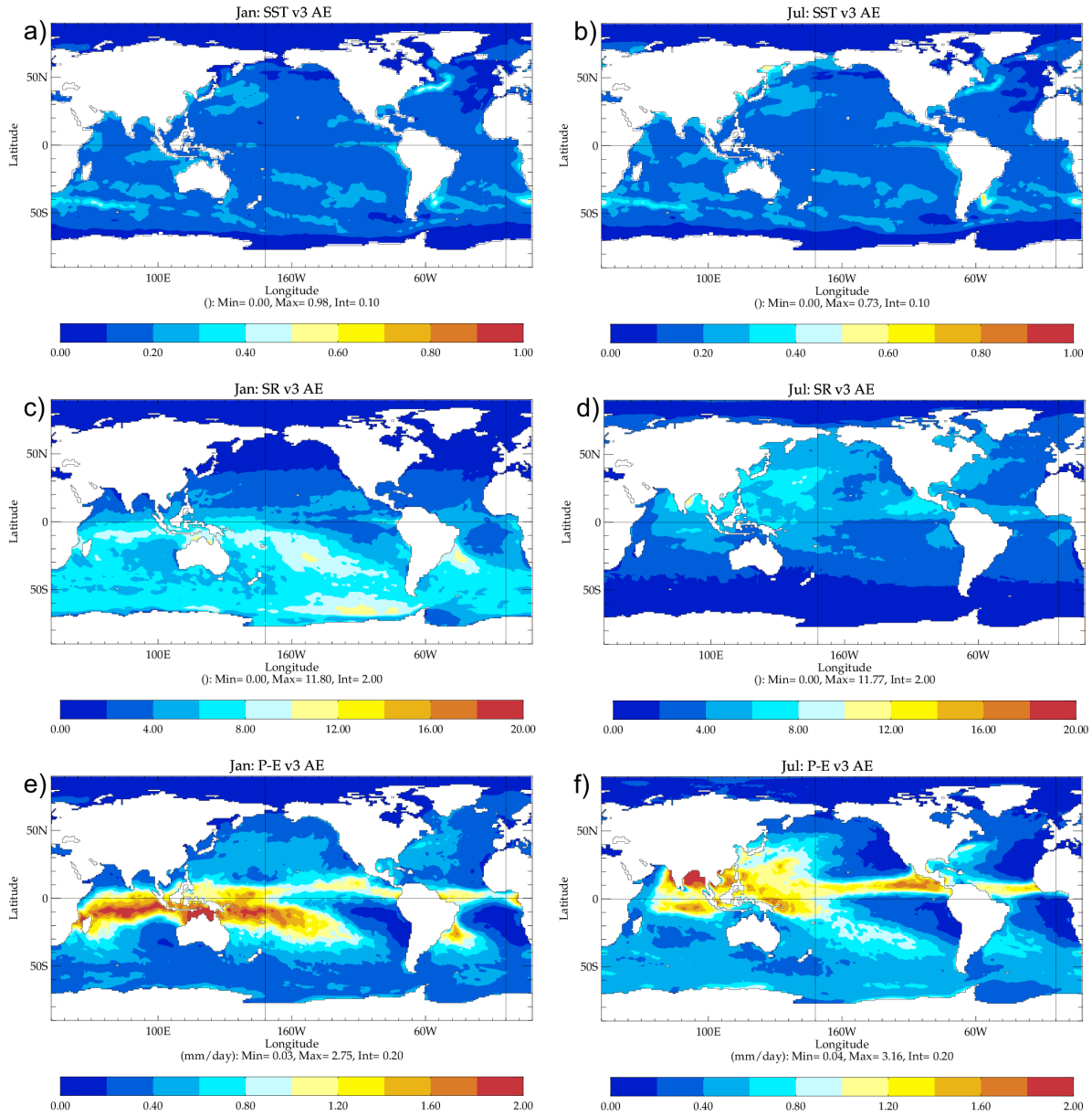

Figure 21: Standard deviation of the *v3 AE* perturbations in January (a,c,e) and July (b,d,f), for SST (a,b), solar radiation (SR) (c,d) and precipitation minus evaporation (P-E) (e,f).

To sample the *AE*, the information from the 10 ensemble members of HadISSTv2 and ERA20C is used. Monthly perturbations have been created for the period 1979-2006, forming a total sample size of 560. We also computed pentad perturbations for the period 1997-2006, with a sample size of 200. Multivariate relationship between variables are preserved in the *AE* perturbations, by using the the same ERA20C ensemble member for any given set of SST (SST is constrained by HadISSTv2 in ERA20C), TAU, P-E and SR perturbations. As shown in Fig. 21, there is a clear spatial correlation between the SR and P-E *AE* perturbations in both January and July. There is also connection in spatial pattern between P-E and SST perturbations, e.g. both show large values in the pathway of Gulf Stream.

Tests comparing monthly versus pentad perturbations indicated that the monthly perturbations were more effective in creating spread in the ocean fields (not shown). Therefore, the five ensemble members of the ocean reanalysis ORAS5 only use the monthly perturbations for either *SE* and *AE*. The pentad perturbations of *SST* can be later added at the time of initializing the coupled forecasts with the purpose of expanding the ensemble of ocean initial conditions (from 5 ensemble ocean reanalyses to N ensemble members, where N is typically 50 or 51). The main features of the new *v*3 perturbations are illustrated in figures Fig. 18, 19, 20, 21, in terms of their standard deviation: how they compare with *v*2 perturbation (Fig. 18), the magnitude of the *SE* and *AE* errors (Fig. 19), the seasonal dependence (Fig. 20) and the multivariate aspects (Fig. 21).

5 Summary and discussion

A new generic ensemble generation scheme has been developed at ECMWF for ODA and ORAS, including perturbations of both assimilated observations and surface forcing fields. This new perturbation scheme has been applied in production of ORAS5. The surface forcing perturbation has also been applied in ERA5 production and in operational EDA system since cycle 43R1. The observation perturbation is designed to account for observation REs. It includes perturbation of geographical location of the in-situ observation and horizontal thinning of the surface observations. It also includes vertical thinning of the in-situ observation profiles. In general, system variance generated by observation vertical perturbation is small compared to that generated by horizontal perturbation, subject to the way of constructing the thinning levels and to the vertical resolution of model. For vertical perturbation, contribution to model uncertainty predominately comes from depths below 700m where the vertical resolution of both the model and observation profiles is significantly reduced.

The diagonal elements of **B** estimated using ensemble spread (σ_b^e) generated by the observation perturbation have been evaluated against prescribed (σ_b^s) and diagnosed (σ_b^d) BGE standard deviation, as well as independent observation data sets, in a series of sensitivity experiments. Large discrepancies are found in the spatial structure and values of different estimations of σ_b, with the σ^s_{b,S_U} being notably different from those given by $\sigma^e_{b,S}$ and $\sigma^d_{b,S}$. Therefore, a revision of the specied BGE and ensemble generation is needed. Improvement in assimilation statistics are expected by combining parameterized and ensemble based flow-dependent estimation of BGE covariances in the ECMWF ODA system.

When verified against observations, large $(\sigma_b^e)^2$ and squared ensemble mean departures appear mostly in eddy active regions, e.g. in the WBCs and extensions, and the ACC, although the ensemble variance appears underestimated. The SST ensemble variance from all sensitivity experiments with in-situ observation perturbations appears also under-dispersive when compared to $\varepsilon^2 - \sigma_o^2$ as verified against OSTIA SST. In the deep ocean large errors can be found in the North Atlantic Ocean around 2000m, which is related to the misrepresentation of the Mediterranean Outflow waters. The errors in these areas are well captured by the ensemble spread generated by the observation perturbation.

Results from sensitivity experiments suggest that super-grid PD value in horizontal perturbation is normally needed in order to maintain a reasonable ensemble spread allowing to account for observation REs. It is also noticed that a reasonably large size of the ensemble member is needed at the Tropical thermocline, where the ensemble spread with 5 ensemble members is considered under-dispersive for both temperature and salinity when compared to σ_b^d. More tests with different ensemble sizes may be needed in order to justify any potential improvement for estimation of the BGE covariances against the extra computational cost of producing more ensemble members.

For ORAS5 production, the geographic locations of all ocean in-situ profiles are horizontally perturbed with $PD = 50km$ using strategy P1. This perturbation distance is chosen while considering the REs of ocean in-situ observations at ORAS5 model resolution. Arguably this is a choice that may need revisiting since different observation types may have different features regarding their REs, e.g. observations from tropical moored buoys are stationary and less biased, therefore can be assigned with a reduced PD value. The in-situ observation vertical perturbation was also applied in ORAS5 with a thinning factor of $N = 2$. So no more than 2 observations within each model level are randomly sampled for data assimilation. Additional impact studies of using REs in the NEMOVAR system is also needed in the future, e.g. impact on the shallow water region due to observation rejection and on regions with strong eddy activities.

Additional perturbations were introduced in ORAS5 for surface observation types such as SIC and SLA. The SIC from OSTIA were perturbed using a stratified random sampling method with a thinning factor of $f_{sic} = 2$, which can be applied to any surface gridded observations. In ORAS5, additional SIC perturbation was also generated from surface perturbation including both SE and AE uncertainties. The SLA observations from AVISO were perturbed through a similar random sampling method, which was developed based on the superobbing scheme as implemented in ORAS4. Assessment of ensemble spread from SLA perturbation was carried out against AVISO MSLA gridded maps, with the results showing very consistent geographical distribution between ensemble spread and ensemble mean errors.

A new version of the surface perturbation ($v3$) scheme has been developed. It perturbs surface variables including SST, SIC, wind stress, net precipitation and solar radiation, by taking into accounts both SE from different analyses data sets (e.g. ESACCI and HadISSTv2), and AE from the same analysis method with multiple ensemble members (e.g. ERA20C). This new surface perturbation scheme also preserves the multivariate relationship between different variables, by using the same ERA20C ensemble member for a given set of SST, TAU, P-E and SR perturbations. Daily perturbation is created using a variance preserving time interpolation method, and can be derived from perturbation repositories constructed with different temporal de-correlation scales. Uncertainties from different sources (AE, SE) or with different temporal de-correlation scales are considered to be independent and therefore can be added together. As with previous versions, the surface forcing perturbation scheme does not sample systematic biases by construction. This new surface perturbation scheme has been applied in ORAS5 production, and also used to create ocean surface perturbation for ERA5 and for operational EDA since Cycle 43R1. At the moment only monthly perturbation has been used in ORAS5 ensemble generation. In the future, both monthly and pentad perturbation should be considered, and multivariate relationship between SIC and SST perturbation should be taken into account in a more explicit way.

Acknowledgements

This work has been partly funded by the ESA CCI Sea Level project. The authors would like to thank Simon T.K. Lang, Kristian Mogensen and Martin Leutbecher for fruitful discussions. Anthony Weaver (CERFACS) also helped in interpreting the results.

References

D. Anderson, T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, F. Molteni, F. Doblas-Reyes, K. Mogensen, and A. Vidard. Development of the ECMWF seasonal forecast System 3. *ECMWF Tech Memo*, (Technical Report 503), 2007.

E. Andersson, M. Fisher, R. Munro, and A. McNally. Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation scheme, and the explanation of a case of poor convergence. *Quarterly Journal of the Royal Meteorological Society*, 126(565):1455–1472, 2000. ISSN 1477870X. doi: 10.1256/smsqj.56511. URL http://www.ingentaselect.com/rpsv/cgi-bin/cgi?ini=xref&body=linker&reqdoi=10.1256/smsqj.56511.

M. A. Balmaseda, A. Vidard, and D. L. T. Anderson. The ECMWF Ocean Analysis System: ORA-S3. *Monthly Weather Review*, 136(8):3018–3034, 2008. ISSN 0027-0644. doi: 10.1175/2008MWR2433.1. URL http://journals.ametsoc.org/doi/abs/10.1175/2008MWR2433.1.

M. A. Balmaseda, K. Mogensen, and A. T. Weaver. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1132–1161, 2013.

B. Bernard, 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, S. Gulev, E. Remy, C. Talandier, S. Theetten, M. Maltrud, J. McClean, and B. De Cuevas. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dynamics*, 56(5-6):543–567, 12 2006. ISSN 1616-7341. doi: 10.1007/s10236-006-0082-1. URL http://link.springer.com/10.1007/s10236-006-0082-1.

P. Brasseur, P. Bahurel, L. Bertino, F. Birol, J.-M. Brankart, N. Ferry, S. Losa, E. Remy, J. Schröter, S. Skachko, C.-E. Testut, B. Tranchant, P. J. van Leeuwen, and J. Verron. Data assimilation for marine monitoring and prediction: The MERCATOR operational assimilation systems and the MERSEA developments. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3561–3582, 10 2005. ISSN 00359009. doi: 10.1256/qj.05.142. URL http://doi.wiley.com/10.1256/qj.05.142.

O. Breivik, K. Mogensen, J. R. Bidlot, M. A. Balmaseda, and P. A. E. M. Janssen. Surface wave effects in the NEMO ocean model: Forced and coupled experiments. *Journal of Geophysical Research C: Oceans*, 120(4):2973–2992, 2015. ISSN 21699291. doi: 10.1002/2014JC010565.

N. Daget, A. T. Weaver, and M. A. Balmaseda. Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Quarterly Journal of the Royal Meteorological Society*, 135(641):1071–1094, 2009. ISSN 00359009. doi: 10.1002/qj.412.

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613): 3385–3396, 2005.

C. J. Donlon, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer. The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*, 116: 140–158, 2012. ISSN 00344257. doi: 10.1016/j.rse.2010.10.017.

N. Ducet, P. Y. Le Traon, and G. Reverdin. Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *Journal of Geophysical Research: Oceans*, 105(C8):19477–19498, 8 2000. ISSN 01480227. doi: 10.1029/2000JC900063. URL http://doi.wiley.com/10.1029/2000JC900063.

T. Fichefet and M. A. Maqueda. Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *Journal of Geophysical Research: Oceans (1978–2012)*, 102(C6):12609–12646, 1997.

S. A. Good, M. J. Martin, and N. A. Rayner. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12):6704–6716, 2013. ISSN 21699291. doi: 10.1002/2013JC009067.

V. Gouretski and F. Reseghetti. On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep-Sea Research Part I: Oceanographic Research Papers*, 57(6):812–833, 2010. ISSN 09670637. doi: 10.1016/j.dsr.2010.03.011.

S. Hirahara, M. A. Balmaseda, and H. Hersbach. Sea Surface Temperature and Sea Ice Concentration for ERA5, 2016.

P. A. E. M. Janssen. Air-sea interaction through waves. In *ECMWF Workshop on Atmosphere-Ocean Interaction*, pages 47–60. ECMWF, 2008.

P. A. E. M. Janssen, S. Abdalla, H. Hersbach, J.-R. Bidlot, P. A. E. M. Janssen, S. Abdalla, H. Hersbach, and J.-R. Bidlot. Error Estimation of Buoy, Satellite, and Model Wave Height Data. *Journal of Atmospheric and Oceanic Technology*, 24(9):1665–1677, 9 2007. ISSN 0739-0572. doi: 10.1175/JTECH2069.1. URL http://journals.ametsoc.org/doi/abs/10.1175/JTECH2069.1.

R. A. Locarnini, A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, C. R. Paver, J. R. Reagan, D. R. Johnson, M. Hamilton, and D. Seidov. World Ocean Atlas 2013. Vol. 1: Temperature. *S. Levitus, Ed.; A. Mishonov, Technical Ed.; NOAA Atlas NESDIS*, 73:40, 2013.

A. C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 10 1986. ISSN 00359009. doi: 10.1002/qj.49711247414. URL http://doi.wiley.com/10.1002/qj.49711247414.

C. J. Merchant, O. Embury, J. Roberts-Jones, E. Fiedler, C. E. Bulgin, G. K. Corlett, S. Good, A. McLaren, N. Rayner, S. Morak-Bozzo, and C. Donlon. Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geoscience Data Journal*, 1(2):179–191, 11 2014. ISSN 20496060. doi: 10.1002/gdj3.20. URL http://doi.wiley.com/10.1002/gdj3.20.

K. Mogensen, M. Alonso Balmaseda, and A. Weaver. The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. Technical Report Technical Report 668 (internal), European Centre for Medium-Range Weather Forecasts, 2012.

P. R. Oke, P. Sakov, P. R. Oke, and P. Sakov. Representation Error of Oceanic Observations for Data Assimilation. *Journal of Atmospheric and Oceanic Technology*, 25(6):1004–1017, 6 2008. ISSN 0739-0572. doi: 10.1175/2007JTECHO558.1. URL http://journals.ametsoc.org/doi/abs/10.1175/2007JTECHO558.1.

P. Poli, H. Hersbach, D. P. Dee, P. Berrisford, A. J. Simmons, F. Vitart, P. Laloyaux, D. G. H. Tan, C. Peubey, J.-N. Thépaut, Y. Trémolet, E. V. Hólm, M. Bonavita, L. Isaksen, M. Fisher, P. Poli, H. Hersbach, D. P. Dee, P. Berrisford, A. J. Simmons, F. Vitart, P. Laloyaux, D. G. H. Tan, C. Peubey, J.-N. Thépaut, Y. Trémolet, E. V. Hólm, M. Bonavita, L. Isaksen, and M. Fisher. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, 29(11):4083–4097, 6 2016. ISSN 0894-8755. doi: 10.1175/JCLI-D-15-0556.1. URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-15-0556.1.

M.-I. Pujol, Y. Faugère, G. Taburet, S. Dupuy, C. Pelloquin, M. Ablain, and N. Picot. DUACS DT2014: the new multi-mission altimeter data set reprocessed over 20 years. *Ocean Science*, 12(5):1067–1090, 9 2016. ISSN 1812-0792. doi: 10.5194/os-12-1067-2016. URL http://www.ocean-sci.net/12/1067/2016/.

N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108(D14):4407, 2003. ISSN 0148-0227. doi: 10.1029/2002JD002670. URL http://doi.wiley.com/10.1029/2002JD002670.

R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang. An improved in situ and satellite SST analysis for climate. *Journal of Climate*, 15(13):1609–1625, 2002. ISSN 08948755. doi: 10.1175/1520-0442(2002)015⟨1609:AIISAS⟩2.0.CO;2.

S. Ricci, a. T. Weaver, J. Vialard, and P. Rogel. Incorporating State-Dependent TemperatureSalinity Constraints in the Background Error Covariance of Variational Ocean Data Assimilation. *Monthly Weather Review*, 133:317–338, 2005. ISSN 0027-0644. doi: 10.1175/MWR2872.1.

A. Schiller, P. Oke, G. Brassington, M. Entel, R. Fiedler, D. Griffin, and J. Mansbridge. Eddy-resolving ocean circulation in the AsianAustralian region inferred from an ocean reanalysis effort. *Progress in Oceanography*, 76(3):334–365, 2008. ISSN 00796611. doi: 10.1016/j.pocean.2008.01.003.

A. Shlyaeva, M. Buehner, A. Caya, J.-F. Lemieux, G. C. Smith, F. Roy, F. Dupont, and T. Carrieres. Towards ensemble data assimilation for the Environment Canada Regional Ice Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 142(695):1090–1099, 1 2016. ISSN 00359009. doi: 10.1002/qj.2712. URL http://doi.wiley.com/10.1002/qj.2712.

H. a. Titchner and N. a. Rayner. The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, 119: 2864–2889, 2014. ISSN 2169897X. doi: 10.1002/2013JD020316. URL http://doi.wiley.com/10.1002/2013JD020316.

S. M. Uppala, P. W. KÅllberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. V. D. Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. a. E. M. Janssen, R. Jenne, A. P. Mcnally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131 (612):29613012, 2005. ISSN 1477-870X. doi: 10.1256/qj.04.176.

J. Vialard, F. Vitart, M. a. Balmaseda, T. N. Stockdale, and D. L. T. Anderson. An Ensemble Generation Method for Seasonal Forecasting with an OceanAtmosphere Coupled Model. *Monthly*

Weather Review, 133(2):441–453, 2 2005. ISSN 0027-0644. doi: 10.1175/MWR-2863.1. URL http://journals.ametsoc.org/doi/abs/10.1175/MWR-2863.1.

A. T. Weaver, C. Deltel, . Machu, S. Ricci, and N. Daget. A multivariate balance operator for variational ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3605–3625, 2005.

C. L. B. M. A. Weaver, Anthony T. Evaluation of the ECMWF Ensemble of Ocean Reanalyses using Assimilation Diagnostics. *Poster for GODAE OceanView Symposium 2013*, 2013.

A. Weisheimer, F. Doblas-Reyes, P. Rogel, E. D. Costa, N. Keenlyside, M. A. Balmaseda, J. M. Murphy, D. M. Smith, M. Collins, B. Bhaskaran, and T. N. Palmer. Initialisation strategies for decadal hindcasts for the 1960-2005 period within the ENSEMBLES project, 3 2007.

M. Yamaguchi, S. T. K. Lang, M. Leutbecher, M. J. Rodwell, G. Radnoti, and N. Bormann. Observation-based evaluation of ensemble reliability. *Quarterly Journal of the Royal Meteorological Society*, 142 (694):506–514, 1 2016. ISSN 00359009. doi: 10.1002/qj.2675. URL http://doi.wiley.com/10.1002/qj.2675.

H. Zuo, M. A. Balmaseda, and K. Mogensen. The ECMWF-MyOcean2 eddy-permitting ocean and sea-ice reanalysis ORAP5. Part 1: Implementation. Technical Report Technical Report 736 (internal), European Centre for Medium-Range Weather Forecasts, 2015a. URL http://www.ecmwf.int/publications/.

H. Zuo, M. A. Balmaseda, and K. Mogensen. The new eddy-permitting ORAP5 ocean reanalysis: description, evaluation and uncertainties in climate signals. *Climate Dynamics*, 6 2015b. ISSN 0930-7575. doi: 10.1007/s00382-015-2675-1. URL http://link.springer.com/10.1007/s00382-015-2675-1.