Assimilation of sea surface temperature, sea ice concentration and sea ice drift in a model of the Southern Ocean

Alexander Barth[a], Martin Canter[a], Bert Van Schaeybroeck[b], Stéphane Vannitsem[b], François Massonnet[c], Violette Zunz[c], Pierre Mathiot[d], Aida Alvera-Azcárate[a], Jean-Marie Beckers[a]

[a] *GeoHydrodynamic and Environmental Research (GHER), University of Liège, Liège, Belgium*
[b] *Koninklijk Meteorologisch Instituut (KMI), Brussels, Belgium*
[c] *Georges Lemaitre Centre for Earth and Climate Research, Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium*
[d] *British Antarctic Survey, Natural Environment Research Council, Cambridge, UK*

Abstract

Current ocean models have relatively large errors and biases in the Southern Ocean. The aim of this study is to provide a reanalysis from 1985 to 2006 assimilating sea surface temperature, sea ice concentration and sea ice drift. In the following it is also shown how surface winds in the Southern Ocean can be improved using sea ice drift estimated from infrared radiometers. Such satellite observations are available since the late seventies and have the potential to improve the wind forcing before more direct measurements of winds over the ocean are available using scatterometry in the late nineties. The model results are compared to the assimilated data and to independent measurements (the World Ocean Database 2009 and the mean dynamic topography based on observations). The overall improvement of the assimilation is quantified, in particular the impact of the assimilation on the representation of the polar front is discussed. Finally a method to identify model errors in the Antarctic sea ice area is proposed based on Model Output Statistics techniques using a series of potential predictors. This approach provides new directions for model improvements.

Keywords: Ensemble Kalman Filter, data assimilation, sea ice drift, Model Output Statistics, Southern Ocean

Email address: a.barth@ulg.ac.be (Alexander Barth)
URL: http://modb.oce.ulg.ac.be/alex (Alexander Barth)

1. Introduction

Observations of the sea ice extent in the Southern Ocean derived from satellite data display a trend of 0.13 to 0.2 million km^2 per decade between November 1978 and December 2012 (Vaughan et al., 2013). Although the magnitude of this trend is subject to uncertainties (e.g., Eisenman et al., 2014), the behavior of the Antarctic sea ice cover is in sharp contrast with its Arctic counterpart which displays a decrease in sea ice extent over the last decades (e.g., Turner and Overland, 2009). Several explanations have been proposed to account for the slight increase in Antarctic sea ice extent but no consensus has been reached yet. Among the proposed mechanisms, a potential link with the stratospheric ozone depletion has been pointed out (Solomon, 1999) but this hypothesis is not compatible with recent analyses (e.g., Bitz and Polvani, 2012; Smith et al., 2012; Sigmond and Fyfe, 2013). Changes in the atmospheric circulation or in the ocean stratification may also have contributed to the observed expansion of the sea ice cover (e.g., Zhang, 2007; Stammerjohn et al., 2008; Goosse et al., 2009; Kirkman and Bitz, 2011; Landrum et al., 2012; Holland and Kwok, 2012; Bintanja et al., 2013; Goosse and Zunz, 2014; de Lavergne et al., 2014). The internal variability of the system, particularly strong in the Southern Ocean, may be responsible for the observed positive trend in Antarctic sea ice extent as well (e.g., Mahlstein et al., 2013; Zunz et al., 2013; Polvani and Smith, 2013; Swart and Fyfe, 2013).

Observations in the Southern Ocean are rather sparse in space and time. In particular, reliable observations of the sea ice concentration are available from the late 1970's only (e.g., Parkinson and Cavalieri, 2012). In this context, climate models constitute adequate tools to compensate for the lack of observations and investigate the processes that govern the behavior of the sea ice cover around Antarctica. Coupled climate models are particularly useful to analyze the interactions between the different components of the climate system. Present-day general circulation models involved in the 5th Coupled Model Intercomparison Project (Taylor et al., 2011) generally simulate a decrease in the Antarctic sea ice extent over the last 30 years but a positive trend such as the observed one remains compatible with the internal variability simulated by these models (e.g., Mahlstein et al., 2013; Zunz

et al., 2013; Polvani and Smith, 2013; Swart and Fyfe, 2013). Nevertheless, these models often display systematic biases in their representation of the seasonal cycle or of the internal variability (or both) of the Antarctic sea ice (e.g., Turner et al., 2013; Zunz et al., 2013). The reconstruction of the sea ice cover in the Southern Ocean provided by these models have thus to be considered cautiously.

One way to more closely constrain the simulation of the ocean and the sea ice is to prescribe the atmospheric conditions at the atmosphere/ocean–sea ice interface. These so-called "forced" simulations resort generally to atmospheric reanalyses as boundary conditions, and have been used extensively to study the past variability of the ocean and sea ice states (Fichefet and Morales Maqueda, 1999; Fichefet et al., 2003; Zhang, 2007; Holland et al., 2014). It is clear the quality of these forced simulations is strongly dependent on that of the atmospheric product utilized. Intercomparisons between different reanalysis products and assessments against in-situ measurements all suggest that the reanalyzed atmospheric data are subject to large uncertainties or systematic errors in the Southern Ocean (Bromwich et al., 2007; Hines et al., 2000; Vancoppenolle et al., 2011) translating inevitably to the ocean–sea ice system (Timmerman et al., 2004; Stössel et al., 2011).

An even tighter constraint on the oceanic and sea ice states can be realized if observations are used to update model estimates. Data assimilation has been an active area of research in climate science. A limited number of studies have, however, attempted to implement data assimilation in the Southern Ocean (Stammer et al., 2002; Stössel, 2008; Massonnet et al., 2013; Ferry et al., 2012; Balmaseda et al., 2008; Carton and Giese, 2008; Janjić et al., 2012) where pressing scientific questions remain, though.

Implementing a data assimilation method in a large-scale ocean–sea ice model presents a number of challenges as several methodological, statistical and physical questions are raised. In theory, the background error statistics should be perfectly known in order for the data assimilation to produce an optimal analysis. This is not feasible in practice, due to the very high dimensionality of the state vector. For this reason, the true covariance matrix of background errors is projected onto a space of much lower dimensionality and specified

either a priori (Ferry et al., 2012) or estimated on-the-fly (Sakov et al., 2012; Mathiot et al., 2012) using a finite-size ensemble. For computational reasons, it is also common to assume a diagonal structure for the observational error covariance matrix (i.e., uncorrelated errors) while this is not necessarily the case in reality.

Most data assimilation methods also rely on statistical hypotheses. The gaussianity of background and observational errors is often assumed, but rarely fulfilled. Not only can this lead to sub-optimal updates, this can also lead to physical inconsistencies. Resorting to the transformation of variables (e.g. Bertino et al., 2003; Simon and Bertino, 2009; Béal et al., 2010) can be a first step, but it only acts on the marginal, and not multivariate probability distribution functions. Likewise, since the background statistics are boiled down to the covariance matrix, the update of non-assimilated fields follows their linear relationship with the observable; this may result in an unphysical or imbalanced state after the update in regions where strong nonlinearities are present, e.g. between sea surface temperature and sea ice concentration (Lisæter et al., 2003).

Last but not least, a central and non-trivial issue concerns the decision on what should be estimated. While the state itself is commonly estimated for reanalysis purposes, the methods can be extended to the estimation of model bias to identify systematic errors (Sakov et al., 2012), to the estimation of model parameters to partly reduce such systematic errors (Massonnet et al., 2014) and ultimately to surface forcing estimation (Barth et al., 2011; Marmain et al., 2014; Ngodock and Carrier, 2014). The estimation of atmospheric forcing in the Southern Ocean has, to our knowledge, not been explored. Because Antarctic sea ice trends are largely controlled by the wind forcing (Kimura, 2004; Holland and Kwok, 2012), it seems natural to improve the representation of ice drift in the model. We propose to correct the wind forcing using satellite sea ice drift data, taking advantage of the strong relationship between sea ice drift and the wind field.

A first set of preliminary experiments have shown the difficulty to assimilate ice drift in a coupled ocean-sea ice model. Sea ice drift is strongly related to the wind forcings (Kimura, 2004; Holland and Kwok, 2012) with a temporal scale of the order of days (about 4 days

based on the autocorrelation). The memory of the sea ice drift is thus relatively short. The corresponding time scale is in fact more similar to the temporal scale of the atmospheric variability than the temporal scale of ocean mesoscale circulation (order of weeks). This short scale would require in principle a very frequent assimilation of sea ice drift data to adequately resolve its underlying time-scale. However, a too frequent assimilation can deteriorate the model results (*e.g.* Malanotte-Rizzoli et al., 1989; Barth et al., 2007; Yan et al., 2014). To improve sea ice drift in the model, we therefore propose to correct the wind forcing. This is possible due to the strong relationship between wind field and sea ice drift (Holland and Kwok, 2012).

The objective of the study is to propose a methodology to use surface drift observations to constrain an ocean-sea ice large-scale circulation model. We also aim to test this approach in combination with sea surface temperature and sea ice concentration assimilation in a decadal simulation and to assess the quality of the results with independent data. This study also outlines an approach to evaluate the presence of model errors at the forecast step of the data assimilation and to identify their potential sources

The ocean model is introduced in section 2 and then the used observations along with their error covariance are discussed (section 3). The procedure adopted to correct the wind field is detailed and validated in section 4. The data assimilation implementation is discussed in section 5 and the results of the reanalysis are then presented and validated (section 6). In the last section, post-processing techniques are used to relate forecast errors in sea ice coverage with model errors associated with the dynamics of sea surface temperature.

2. Model

The primitive-equations model used in this study is NEMO (Nucleus for European Modelling of the Ocean, Madec (2008)), coupled to the LIM2 (Louvain-la-Neuve Sea Ice Model) sea ice model (Fichefet and Morales Maqueda, 1997; Timmermann et al., 2005; Bouillon et al., 2009). The global ORCA2 implementation is used, which is based on an orthogonal grid with a horizontal resolution of the order of 2° and 31 z-levels (Mathiot et al., 2011; Massonnet et al., 2013). The hydrodynamic model is configured to filter free surface gravity waves by

including a damping term. The leap-frog scheme uses a time step of 1.6 hours for dynamics and tracers. The model is forced using air temperature and wind from the NCEP/NCAR reanalysis (Kalnay et al., 1996). Relative humidity, cloud cover and precipitation are based on a monthly climatological mean. The sea surface salinity is relaxed towards climatology with a fresh water flux of -27.7 mm/day times the salinity difference in psu.

As in the following the link between sea ice drift and wind stress is studied, only the equation for sea ice drift is given here. The sea ice drift $\mathbf{u}$ is governed by the momentum equation where the advection of momentum is neglected by scale analysis (Fichefet and Morales Maqueda, 1997):

$$m\frac{\partial \mathbf{u}}{\partial t} = -mf\mathbf{e_z} \wedge \mathbf{u} + \boldsymbol{\tau}_{ai} + \boldsymbol{\tau}_{wi} - mg\nabla\zeta + \mathbf{F} \tag{1}$$

where m is the mass of the snow-ice system, f is the Coriolis parameter, $\mathbf{e}_z$ is a unit vector pointing upward, $\boldsymbol{\tau}_{ai}$ (resp. $\boldsymbol{\tau}_{wi}$) denotes the drag with air (resp. water), g is the acceleration due to gravity, ζ is the surface elevation and $\mathbf{F}$ the force due to the variation of internal stresses. For the complete model equations, the interested reader is referred to Madec (2008); Fichefet and Morales Maqueda (1997); Goosse and Fichefet (1999).

3. Observations

Global foundation sea surface temperature from OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis Roberts-Jones et al., 2012) at an original resolution of 0.05° was reduced to a resolution of 2° by averaging all temperature values within a 2° by 2° grid cell. This data set also provides an error standard deviation (varying in space and time). Since information about the length scales over which the SST errors are correlated are lacking, the error standard deviation is also reduced to a resolution of 2° by simply averaging the standard deviations of all values with a 2° by 2° grid cell (averaging the standard deviation instead of the variances corresponds to the case of perfectly correlated errors, which is more appropriate since the OSTIA data set is relatively smooth).

Global sea ice fraction from the EUMETSAT Ocean and Sea Ice Satellite application Facility (OSI-SAF Roberts-Jones et al., 2012) was also reduced to a resolution of 2° and assimilated with an error standard deviation of 0.1. The OSTIA sea surface temperature and the OSI-SAF sea ice fraction are distributed by MyOcean. Daily sea ice drift from NSIDC (National Snow and Ice Data Center) is also assimilated in the ocean model. The sea ice drift is based on data from the Advanced Very High Resolution Radiometer (AVHRR), Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave/Imager (SSM/I), and International Arctic Buoy Programme (IABP) data (Fowler, 2003). The ice drift is derived from the satellite data by maximizing the correlation of groups of pixels between image pairs. We use version 1 of this data set which does not include derived ice drift based on NCEP wind reanalysis (unlike the version 2 of the gridded and combined ice drift data set). As the focus of this study is the Antarctic Ocean, only data from the southern hemisphere is used. The error standard deviation for the assimilation is assumed to be 0.1 m/s. The value of this parameter was determined by a series preliminary experiments to find the right balance between correcting as much as possible the sea ice drift without degrading unobserved variables. The sea ice drift is used at two stages in our study: first it is used to adjust the wind fields and, at a later stage, is used during the analysis to correct the model state vector. This approach has been adopted as errors in the wind field and errors in the ocean circulation manifest themselves in the model sea ice drift (and not only error in the wind field). Experiments with different values of the standard deviation of the global sea ice fraction error and sea ice drift error were also conducted as detailed in section 5.5.

4. Wind field adjustment

Errors in the sea ice drift can be attributed either to errors in the winds fields or to error in the ocean currents. As winds and currents have two very different time scale, a two-step approach was adopted. First, the wind fields are adjusted using sea ice drift as described in this section. In a second step the sea ice drift is assimilated into the ocean-sea ice model in order to reduce errors due to the ocean currents (section 5).

4.1. Relation between sea ice drift and wind

The model sea ice drift is strongly related to the used wind forcing. To quantify the relationship between sea ice drift and wind fields, the complex correlation coefficient (Kundu and Allen, 1976) between the daily NEMO-LIM sea ice drift $(u_{\text{ice}}, v_{\text{ice}})$ and daily NCEP winds $(u_{\text{wind}}, v_{\text{wind}})$ has been computed by introducing the following complex variables ($i^2 = -1$):

$$\begin{aligned} w_{\text{ice}} &= u_{\text{ice}} + i\, v_{\text{ice}} \\ w_{\text{wind}} &= u_{\text{wind}} + i\, v_{\text{wind}} \end{aligned}$$

In order to maximize the correlation, we correlate the sea ice drift with different transformations of the wind field. More specifically, we use different combinations of shifts and filters in time of the wind field. We use a time filter because we anticipate the sea ice drift to have a certain inertia and thus a memory of previous winds. The time shift and the temporal scale of the filter will be determined later. The time filter is implemented using an iterative diffusion scheme using a forward Euler step and a 2nd-order center diffusion operator (Alvera-Azcárate et al., 2009). The complex correlation coefficient between sea ice drift and filtered and shifted wind fields is given by:

$$\gamma = \frac{\langle \overline{w_{\text{ice}}} w_{\text{wind}} \rangle}{\sqrt{\langle \overline{w_{\text{ice}}} w_{\text{ice}} \rangle \, \langle \overline{w_{\text{wind}}} w_{\text{wind}} \rangle}}$$

The over-line denotes the complex conjugate and the angular brackets an average over time. The absolute value of the complex correlation coefficient is maximized by changing the timeshift and time-filter. The complex regression coefficient r derives an empirical relationship between the sea ice drift and the wind field. This relationship will be used later for wind field adjustment.

$$r = \frac{\langle \overline{w_{\text{ice}}} w_{\text{wind}} \rangle}{\langle \overline{w_{\text{ice}}} w_{\text{ice}} \rangle}$$

The complex correlation and regression coefficients are used instead of the (real) correlation/regression coefficient derived on the zonal and meridional component individually because the complex coefficients can represent a rotation by a constant angle between the two

vectors (as a result for the Coriolis force) and is thus commonly used to analyze horizontal velocities (*e.g* Kundu and Allen, 1976; Barth et al., 2008).

The correlation analysis showed a strong correlation with magnitude of 0.9363 and a phase of -19.52° between sea ice drift and 3-day average wind fields (panel (a) of Figure 1). This phase (which is also the phase of the complex regression coefficient) represents the angle between the sea ice drift vector and the wind vector. The maximum value was obtained with no time lag. These results did not confirm the initial expectation of a time lag between wind and sea ice drift as one could assume that the wind (the cause) precedes sea ice drift (the effect). The maximum of the correlation as a function of the time lag is very well defined while the correlation as a function of the filtering time scale is a bit flatter (panel (b) and (c) of Figure 1). A scatter plot of wind versus sea ice drift using the optimal parameters (filtering time scale and time lag) shows a good correspondence (Figure 2).

This strong relationship has been used to correct the surface winds. The general approach is to use the regression coefficient to transform the observed sea ice drift as pseudo wind observations and to attempt to improve zonal and meridional wind fields components. In particular the following procedure has been adopted to compute the adjusted wind field:

- the first guess wind field is the NCEP reanalysis
- the model is run with this wind field (here for the year 2000)
- the sea ice drift error is calculated by comparing model with observed sea ice drift
- the sea ice drift error is transformed to "wind increment" using the regression coefficient r
- "wind increment" is analyzed with the tool divand (detailed in the next section) on the ORCA grid and the first guess (the NCEP reanalysis) is added

While other calibration experiment are carried out for the year 1985, the wind field adjustments are first tuned for the year 2000 due to the availability of the Cross-Calibrated Multi-Platform (CCMP) Ocean Surface Wind Vector Analyses (Atlas et al., 2011) which will be used to independently validate the results.

4.2. Spatial analysis with divand

The sea ice drift provides only information about the wind field where the model has ice. However, if the sea ice drift indicates that the model should be corrected at a particular location, one can expect that it should also be corrected in a similar way at neighboring grid cells (even if they are not covered by ice). The tool divand (Barth et al., 2014) (Data Interpolating Variational Analysis in n-dimensions) is used to spatially interpolate the "wind increment" derived from the sea ice drift on the full ORCA2 grid. This tool is similar to the variational analysis DIVA (Brasseur et al., 1996) but this latter works on a triangular mesh. The DIVA tool cannot represent a periodic domain as is the case with the NEMO grid. Therefore the new tool divand was adapted to operate directly on a structured model grid with periodic boundary conditions.

For variational analyses, one requires that the analyzed field φ (here the wind increment) must be close to the N_d observations d_j ($j = 1, \ldots, N_d$) and "smooth". This is quantified using a cost function J:

$$J[\varphi] = \sum_{j=1}^{N_d} \mu_j [d_j - \varphi(\mathbf{x}_j)]^2 + \|\varphi\|^2 \tag{2}$$

Each observation has a weight μ_j which is directly related to its error variance. Abrupt variations of the analysis field are penalized using a regularization constraint (as in norm spline interpolation):

$$\|\varphi\|^2 = \frac{1}{c} \int_D \phi^2 + 2 \left(\tilde{\nabla}\phi\right) \cdot \left(\tilde{\nabla}\phi\right) + \left(\tilde{\nabla}^2\phi\right)^2 dx \tag{3}$$

The normalization coefficient c is set to 4π to ensure that the background covariance matrix has a unit variance (Brasseur et al., 1996). The data weight μ_j represents the radio of the error variance of the background NCEP field and the error variance of the observation.

The correlation length L is introduced by scaling the differential operator for gradient and Laplacian:

$$\tilde{\nabla} = L\nabla$$
$$\tilde{\nabla}^2 = \nabla \cdot \left(L^2 \nabla\right)$$

Here the correlation length is a scalar, but it can also be a diagonal matrix in the previous equation if the length scale is different for the zonal and meridional dimensions. The regularization constraint guarantees that the interpolated field and its first order derivative (such as the wind curl) is continuous.

An estimation of the background NCEP error standard deviation is necessary to define the data weight μ_i. In the present case, the pseudo observation (the wind information derived from the sea ice drift observation) error standard deviation is derived using the error standard deviation of the sea ice drift. The remaining parameter of the analysis, namely the correlation length scale and background NCEP error standard deviation, will be determined in the following. The correlation length controls the distance over which the information from the pseudo-wind observation is extrapolated spatially and the NCEP error standard deviation determines how close the analyzed field has to come to these pseudo-observations.

4.3. Calibration

The correlation length is varied from 300 km to 5000 km and the standard deviation error is varied from 0.1 m/s to 10 m/s. For each parameter 10 values are tested. These values are uniformly distributed in logarithmic space. For each of these parameters, the divand analysis is performed and the NEMO-LIM2 model is run simulating the year 2000 with the adjusted wind fields. Besides the momentum equation, the norm of the adjusted wind field is also used in the heat flux computation via the bulk formulas (which is a separate input file for NEMO).

Figure 3 shows the RMS error between the model sea ice drift obtained using the adjusted wind and the observed sea ice drift. This is not an independent validation since the observed sea ice drift is used to adjust the wind fields. This comparison is rather a confirmation that the adjustment works as expected. The RMS error between the model sea ice drift

and observations is 0.1235 m/s with original (i.e. non-adjusted) NCEP forcing. The RMS error is indeed reduced thanks to the adjustment (Figure 3) and, as expected, the lowest RMS error is obtained when using a large value of the standard deviation of the NCEP wind error. In this case, the adjusted wind will thus be forced to come closer to the pseudo-wind observations (based on sea ice drift).

As an independent comparison the CCMP Ocean Surface Wind Vector Analyses (Atlas et al., 2011) based on ERA-40 and observations such as QuickScat for the year 2000 south of 60°S is used. Sea ice drift is not used in the CCMP product. For every tested correlation length scale and NCEP error standard deviation, the RMS error between the adjusted wind and the CCMP wind field is computed (Figure 3, panel b). This comparison shows that the wind field is indeed improved using the sea ice drift measurements. As before, the general tendency is that the RMS difference between the adjusted winds and CCMP winds decreases as the error standard deviation increases. This comparison shows that the optimal range of the correlation length scale is between 500 km and 1500 km.

The model sea ice concentration obtained by the adjusted wind is also compared to the OSTIA/OSI-SAF observations (Figure 3, panel c). In general the sea ice concentration varies only weakly by changing the parameter of the analysis. This suggests that only a small part of the RMS error in sea ice concentration can be attributed to the wind forcing and that sea ice concentration is mostly driven by thermodynamic forcings. Contrary to the previous comparison the error slightly increases for large values of the NCEP error standard deviation.

Overall a large value of the background error improves sea ice drift and reduces the RMS error in comparison with CCMP winds but it degrades the sea ice concentration (if error standard deviation is larger than 5 m/s). Correlation lengths between 500 km and 1000 km give acceptable results. By combining the results from the different comparisons, the wind fields have been adjusted using a correlation length of 700 km and a background error standard deviation of 2 m/s.

The adjustment procedure has been applied to the wind field time series from 1985 to 2006.

The average wind vector over this period has been computed (Figure 4). The most significant change occurs near the coast where the adjustment generally increase the offshore wind and decreases (the generally negative) zonal wind component. Those changes are consistent with the changes from low to high resolution atmospheric models (Mathiot et al., 2010) and are attributed to katabatic winds which are important to the formation of polynyas (Massom et al., 1998; Morales Maqueda et al., 2004). As the wind curl is an important forcing for Ekman pumping it has been computed for the original and adjusted wind in order to determine the effects on the wind curl. The most important changes in the wind curl (in term of the RMS error) have been applied in the Weddell and Ross Seas (panel (b) of Figure 5). Even in these areas the RMS difference is still smaller by a factor of 2 to 3 than the standard deviation of the wind curl (panel (a) of Figure 5).

For the year 2006, a problem in version 1 of the NSIDC ice drift product has been identified for the Arctic sea ice drift (Sumata et al., 2014). However the authors of this study did not analyse the ice-drift in the Southern Hemisphere. The comparison of the free-running model with the Antarctic ice drift did not reveal a sudden change in the RMS error for the year 2006. However, wind field corrections based on version 2 of the NSIDC ice drift data (using only Advanced Very High Resolution Radiometer (AVHRR), Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave/Imager (SSM/I) but not NCEP wind field data), could reduce the RMS error of the wind field compared to CCMP could by 7% compared to version 1 of the NSIDC ice drift.

5. Data assimilation

The implemented data assimilation scheme is the Ensemble Transform Kalman Filter (Bishop et al., 2001). In ensemble-based assimilation schemes, the error statistics of the model state vector is estimated by perturbing uncertain aspects of the model. In the present configuration we perturb surface winds (10 m) and surface air temperature (2 m). The adjusted wind from the previous section are used. Atmospheric parameters coming from climatology are not perturbed. The data assimilation scheme employs an ensemble with

50 members. Observations are assimilated every 5 days which is a compromise between available computer resources and maximizing the usage of the observations.

5.1. Perturbed forcings

The perturbation scheme is based on a Fourier decomposition. Formally all perturbed variables are grouped into a time-dependent vector:

$$\mathbf{x} = (u_{\text{air}}\ v_{\text{air}}\ T_{\text{air}})^T \quad \text{at all grid points} \tag{4}$$

A Fourier decomposition of the NCEP wind vectors and air temperature over the time domain is performed (Barth et al., 2011; Marmain et al., 2014):

$$\mathbf{x}(t) = \sum_k \mathbf{a}_k \exp(i\omega_k t), \tag{5}$$

where ω_k is the k-th angular frequency (positive or negative) and $\mathbf{a}_k$ are complex spatial vector fields corresponding to the Fourier coefficients of the angular frequency ω_k (Δt is here 24 hours):

$$\omega_k = \frac{2\pi k}{\Delta t} \qquad k = -\frac{k_{max}}{2}, \ldots, \frac{k_{max}}{2} - 1 \tag{6}$$

where k_{max} is the number of Fourier modes. Using the spatial and multivariate structure of the Fourier modes $\mathbf{a}_k$, perturbations $\mathbf{x}'(t)$ are constructed by:

$$\mathbf{x}'(t) = \alpha\ \text{Re}\left(\sum_k \mathbf{a}_k z_k(t)\right), \tag{7}$$

where z_k is a complex random time series with a temporal correlation scale of $T_k = 2\pi/|\omega_k|$, zero mean and unit variance. The value of α is determined by ensemble simulations with perturbed forcings to ensure that the resulting ensemble spread is comparable to the expected error of the model.

Only Fourier modes with a time period between 20 and 70 days are used for the perturbations in order to exclude the seasonal variations (which have a large variance and whose amplitude is not representative for its expected error) and short-scale variations which are not the

primary focus of this study. Since the perturbation scheme is multivariate, the same range of time scales is used to perturb wind and air temperature. The real and imaginary parts of the random time series have the following covariance $C_T(t, t')$:

$$C_T(t, t') = e^{-\frac{(t-t')^2}{T_k^2}} \tag{8}$$

These perturbations have been added to the first guess estimate:

$$\mathbf{x}^{(l)} = \mathbf{x} + \mathbf{x}'^{(l)} \tag{9}$$

where l is the index of the ensemble member. The perturbation scheme is similar to perturbations generated by Empirical Orthogonal Functions (EOF; *e.g.* Vandenbulcke et al., 2008; Béal et al., 2010). The advantage of perturbations proportional to a Fourier mode is that it is easier to associate a time scale to a Fourier mode than to an EOF. The wind perturbations have a magnitude similar to the wind correction derived in the previous section. However, the wind perturbations have a zero mean so that the wind corrections are still relevant for the whole ensemble (and in particular its mean state).

As the wind speed is an independent forcing field for NEMO, the perturbed wind speed is computed for each ensemble member:

$$\|\mathbf{u}_{\text{air}}{}^{(l)}\| = \sqrt{u_{\text{air}}^{(l)\,2} + v_{\text{air}}^{(l)\,2}} \tag{10}$$

The perturbation scheme (without assimilation) is illustrated for the year 2007. Figure 6 shows the ensemble standard deviation of the wind speed for a particular day (2007-02-21). The ensemble standard deviation is high near the polar regions where time variability (between 20 and 70 days) is relatively large. The same behaviour was also seen in the air temperature (not shown). Figure 7 shows the globally-averaged ensemble spread for one-year of spin-up. The spread in SST stabilized relatively fast after 2 months of simulation. Globally the uncertainty of the ensemble SST is about 0.8 °C which is of the order of magnitude of the model error in SST. A meaningful correction of the model state can only occur at locations where the model develops a sufficient ensemble spread. The spread of sea ice concentration

was computed for the period of minimum and maximum sea ice extent (Figure 8). For the minimum sea ice extent period, a spread of 0.3 or larger was generated except in the eastern part where nearly all sea ice has melted. During the maximum sea ice extent, areas with significant ensemble spread form a ring structure. All ensemble members have no sea ice outside this ring and are essentially completely sea ice covered inside this ring. The width of this ring represents the uncertainty of the sea ice edge.

Ensemble simulations have also been carried out with only air temperature perturbations and only wind field perturbations in order to determine the impact of those error sources individually. Figure 9 shows the resulting ensemble variance after a one-year ensemble spin-up for ice concentration and ice drift variance. The air temperature perturbations generate a relatively uniform ensemble spread while the wind perturbations enhance the ensemble spread mostly near the coastline. As expected, the ice drift variance (computed only over model grid points with nonzero ice concentration) does not respond to the air temperature perturbations, but it responds quite significantly to the wind field perturbations. The impact is strongest in the open ocean where the ice movement is not constrained by the coast. The combined effect of air temperature and wind perturbations is relatively close to the sum of both perturbations individually which indicates that the wind field and air temperature perturbations as well as their response in the model are relatively independent from each other even after a one-year ensemble simulation.

5.2. State vector

In data assimilation, all model variables to be corrected by the observations are gathered in the state vector which is here composed of various hydrodynamic and sea ice variables. It includes the horizontal velocity components, temperature, salinity, surface elevation, curl of horizontal velocity components, divergence of horizontal velocity components, turbulent kinetic energy and barotropic stream function trends. Those variables are necessary to restart NEMO. Some of these variables are interdependent as the horizontal velocity components are directly related to their divergence and curl. However since the link is linear, the analysis will preserve their relationship. As the model uses a leap-frog time step, two time instances

of these variables are included in the state vector.

For the sea ice model, sea ice concentration, sea ice thickness, the horizontal sea ice-velocity components, snow thickness and temperature inside the ice/snow layer (at three layers) are included in the state vector. The sea ice concentration was transformed with a Gaussian anamorphosis (see section 5.3). In total, the state vector contains 28 different variables and about 6 million elements (all variables combined).

A set of experiments was conducted with a reduced state vector, where the snow thickness and temperature inside the ice/snow layer was not corrected by the assimilation in order to determine if the assimilation has a beneficial impact on these variables.

5.3. Gaussian anamorphosis

The Kalman filter analysis provides the most likely state if errors are Gaussian-distributed. However some variables are clearly not Gaussian-distributed, in particular sea ice concentration which is bound between 0 and 1. A linear analysis scheme can produce unrealistic values outside of this range. Gaussian anamorphosis (Bertino et al., 2003; Lenartz et al., 2007; Simon and Bertino, 2009; Béal et al., 2010) consists of applying a non-linear transformation onto the model variable which should make the pdf of the state vector more similar to a Gaussian pdf. In practice such transformations are applied to individual elements of the state vector and thus operating only on the marginal distribution and not on the full multidimensional pdf. Such transformations can be based on an analytic transformation (e.g. logarithm, for lognormal distributions) or empirically based on the distribution of the observations.

The initial distribution of sea ice concentration is estimated from a 1-year free-running ensemble simulation. The derived transformation function is here independent in time and space (Figure 10). The anamorphosis transform was applied to all ensemble members so that the ensemble members include the transformed sea-ice concentration. Observed sea ice concentration was not transformed, therefore the observation operator includes the inverse anamorphosis transform and is non-linear. This approach allows to define the error standard deviation of the observations in the original units. The analysis scheme is implemented with

the non-linear observation operator as described in Chen and Snyder (2007) and Barth et al. (2011).

Other model variables exhibit a non-Gaussian behavior too as thickness of snow and sea ice layers (which have to be positive) and sea water temperature (which has to be above the freezing temperature). We limit ourselves to sea ice concentration as this is an observed variable and thus large corrections are expected for this parameter. Other variables with a non-Gaussian distribution are dealt with using an ad-hoc correction after the analysis step.

5.4. Analysis

The analysis scheme is based on the Kalman filter analysis, where the model forecast $\mathbf{x}^f$ (with error covariance $\mathbf{P}^f$) is updated by the observation $\mathbf{y}^o$ (with error covariance $\mathbf{R}$) resulting in the analysis state $\mathbf{x}^a$ and its error covariance ($\mathbf{P}^a$):

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{11}$$

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{12}$$

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f \tag{13}$$

where $\mathbf{H}$ is the observation operator extracting the observed part of the state vector. The mean state $\mathbf{x}^f$ and its covariance are computed from an ensemble of perturbed members $\mathbf{x}^{\mathbf{f}(k)}$ where $k = 1, \ldots, N$ (Evensen, 2007):

$$\mathbf{x}^f = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}^{f(k)} \tag{14}$$

$$\mathbf{P}^f = \frac{1}{N-1}\sum_{k=1}^{N}\left(\mathbf{x}^{f(k)} - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^{f(k)} - \overline{\mathbf{x}^f}\right)^T = \mathbf{S}^f\mathbf{S}^{fT} \tag{15}$$

where the columns of the matrix $\mathbf{S}^f$ are the difference between each member and the ensemble mean (multiplied by $\frac{1}{\sqrt{N-1}}$). The ensemble update used here is based on the Ensemble Transform Kalman Filter (Bishop et al., 2001). In order to avoid to form $\mathbf{P}^a$ explicitly, $\mathbf{P}^a$

is expressed also in terms of the square root matrix $\mathbf{S}^a$ ($\mathbf{P}^a = \mathbf{S}^a\mathbf{S}^{aT}$) which is possible when the following eigenvalue decomposition is made:

$$\left(\mathbf{H}\mathbf{S}^f\right)^T \mathbf{R}^{-1}\mathbf{H}\mathbf{S}^f = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{16}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and where $\mathbf{\Lambda}$ is diagonal. $\mathbf{U}$ and $\mathbf{\Lambda}$ are both of size $N \times N$.

Using this eigenvector decomposition and the Sherman-Morrison-Woodbury formula (Golub and Van Loan, 1996) one can compute the analysis $\mathbf{x}^a$ and the square root of the analysis error covariance $\mathbf{S}^a$ by:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{S}^f\mathbf{U}(\mathbf{I}+\mathbf{\Lambda})^{-1}\mathbf{U}^T(\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{17}$$

$$\mathbf{S}^a = \mathbf{S}^f\mathbf{U}(\mathbf{I}+\mathbf{\Lambda})^{-1/2}\mathbf{U}^T \tag{18}$$

Based on $\mathbf{x}^a$ and $\mathbf{S}^a$, an ensemble can be finally reconstructed:

$$\mathbf{x}^{a(k)} = \mathbf{x}^a + \sqrt{N-1}\,\mathbf{S}^{a(k)} \tag{19}$$

In order to filter spurious long-range correlations, a localization scheme in the observation space has been used with a length scale of 2000 km (approximately 20 grid points). This assimilation scheme is implemented in a tool called the Ocean Assimilation Kit and described in more detail in Vandenbulcke et al. (2006); Barth et al. (2008). The present setup is the first global implementation of the assimilation tool and it required some adaption in order to properly handle the periodic boundary conditions in the localization scheme.

5.5. Variant of assimilation experiments

Different assimilation experiments were conducted to assess the different choices that have been adopted during the implementation of the assimilation scheme. Table 1 shows the RMS error and skill-score of the model state forecast compared to the observations (not yet assimilated). The RMS values for SST and sea ice concentration are computed over the

entire globe. If the RMS values for sea ice concentration are to be compared with RMS values computed over the latitude range $[-90, -\phi]$ and $[\phi, 90]$, then the values reported here have to be multiplied by $1/(1 - \sin(\phi))$. The RMS values for the sea ice drift are computed only over the southern hemisphere over the grid cells where sea ice is present in the model and in the observations. These experiments were carried out for the year 1985 (the initial year of the study). The skill-score is defined as:

$$\text{skill score(experiment)} = 1 - \frac{\text{RMS}^2\text{(experiment)}}{\text{RMS}^2\text{(baseline)}} \tag{20}$$

Negative values of this skill score mean a deterioration of the results and positive values an improvement. The baseline experiment uses a state vector of 30 variables, Gaussian anamorphosis is applied to sea ice concentration, and the standard deviation error for the sea ice concentration is set to 0.1. In a first test, the Gaussian anamorphosis was disabled and the sea ice concentration was adjusted to the interval $[0, 1]$ after the analysis (row ExpFNA-0.1 in table 1). Compared to the baseline experiment, a small error increase in sea ice concentration was observed, while other variables are not affected. Given the strong non-Gaussian character of the sea ice concentration, one could have expected a larger impact of the anamorphosis transform. However, the Gaussian anamorphosis only transforms the marginal pdfs while the character of the multidimensional pdf of the state-vector might not be significantly changed by the transformation. In the following experiments, the Gaussian anamorphosis was kept.

By reducing the error standard deviation (ExpFA-0.07) of the sea ice concentration to 0.07, the sea ice concentration forecast was improved by 8 % (compared to the baseline experiment with an error standard deviation of 0.1). A larger positive impact on the model variables (and especially on sea ice concentration) was obtained by using a reduced state vector excluding snow thickness and temperature inside the ice/snow layer (ExpRA-0.05, ExpRA-0.07 and ExpRA-0.1). This indicates that the excluded variables are related in a non-linear way to the observations and that relationship cannot be represented by a covariance. However, for

this experiment the best results were obtained by using again 0.1 as error standard deviation of the sea ice concentration. Using a lower value did not result in an improvement contrary to the result with the full state vector.

In summary, the experiments lead to the configuration with a reduced state vector and Gaussian anamorphosis of sea ice concentration, where sea ice concentration was assimilated with an error standard deviation of 0.1.

6. Reanalysis

While the calibration of the assimilation setup was performed on a single year (2000), this section presents the model simulations with data assimilation from 1st January 1985 to 31st December 2006. The time period was determined to ensure the availability of all used data sets.

6.1. RMS with assimilated data sets

The comparison with the assimilated data set is instructive to get a first view of the behaviour of the assimilated variables. The RMS error for sea surface temperature and sea ice concentration are computed over the whole globe while the RMS error for sea ice drift is limited to the southern hemisphere. The free model is not influenced by the observation described in section 3. In particular, its wind forcing is the original NCEP wind forcing. As mentioned previously, the sea ice drift observations are used at two stages: for correcting the wind forcing and during the analysis.

Since the RMS errors showed a clear seasonal behaviour the 1606 assimilation cycles were aggregated on a monthly basis (Figure 11). In the free run, the sea surface temperature RMS error is on average 0.98 °C. This error is strongly reduced after the first assimilation cycle (not shown) and maintained at a relatively low level (about 0.5 °C) by the continuous assimilation of SST data. On average, the SST RMS error is highest during August and a secondary peak is observed in January. The seasonal behaviour of the SST RMS is significantly reduced by the assimilation. The behaviour of the sea ice concentration is similar, as a clear seasonal

cycle can be seen in the RMS error and the average RMS error is highest in September (the period with the maximum sea ice extent in the Southern Hemisphere). As expected from the previous result, the RMS error of the sea ice drift component (only in the southern hemisphere) shows also a strong seasonal cycle. While the model forecast RMS is still smaller than the RMS of the free run, the RMS error increases relatively fast after the analysis (not shown). This is attributed to the fact that the underlying time scales of the variability in sea ice drift are shorter than the 5-day assimilation cycle. The average of the RMS error over all assimilation cycles is given in table 2.

6.2. Validation with the World Ocean Database

All observations from the World Ocean Database from the period 1st January 1985 to 31st December 2006 have been collected. For model verification, the model results are usually interpolated on the location and date of the observations (*e.g* Alvera-Azcárate et al., 2007). As the vertical grid of the model is the same at every location, we decided to rather interpolate the observations vertically on the model levels. These vertically interpolated profiles are then compared to the model results interpolated horizontally.

The free-running model has the largest temperature error near the surface where the model has the most variability (Figure 12). As the model assimilates sea surface temperature, the largest impact of the assimilation is indeed at the surface where the RMS error and bias (which is partly included in the RMS error) are strongly reduced. The RMS error is improved by the assimilation over 200 m depth and the bias over 120 m. Below those depths there is a slight degradation of the temperature which is essentially a systematic error in form of a bias. One possible way forward for improvement of the assimilation scheme could be to include a temperature relaxation toward a climatology to control such error. As the ensemble is generated by perturbing the atmospheric fields, the resulting vertical correlation scale between the surface and the subsurface level is about 100 meters (as calculated by computing the standard deviation averaged over time and horizontal space of the analysis increment). As the error increase at depth is not introduced by the analysis step, it must be

introduced by the model reaction to an analyzed initial condition. In fact, it is well known that sequential analysis can produce shocks after restarting the model from an analysis (*e.g.* Malanotte-Rizzoli et al., 1989; Barth et al., 2007; Yan et al., 2014). Incremental update techniques are a promising approach to reduce such problems during the re-initialization of the model (Bloom et al., 1996; Yan et al., 2014).

The model does not assimilate salinity and therefore changes in salinity are only due to the multivariate covariance between the observed variables and salinity, and also due to the model adjustment after the analysis. The validation reveals that the assimilation reduces the salinity RMS error and bias everywhere with a diminishing impact at depth. Contrary to the temperature validation, no degradation at depth was observed.

6.3. Mean sea surface height

The mean model sea surface height was also compared with CNES-CLS09 MDT version 1.1 (Rio et al., 2011). The model sea surface height is related to the mean currents by the geostrophic relationship. The CNES-CLS09 MDT is essentially based on in situ dynamic heights, drifting buoy velocities and the geoid model computed from GRACE (Gravity Recovery and Climate Experiment) data. It is thus an independent data set. The objective of this comparison is to assess the impact of the assimilation on the mean sea surface height and the mean currents. As the focus of this study is the southern polar region, the comparison is limited to the area south of 40°S. A constant over this domain has been subtracted to remove any offset which is not dynamically significant. The RMS values represent thus centered RMS. The RMS error between the free running model and the MDT over this area is 0.218 m which is reduced to 0.165 m between analysis and MDT. The RMS of the 5-day forecast based on the analysis is essentially the same with 0.166 m (Figure 13). Overall the mean SSH gradient is more realistic in the analysis compared to the free model run leading to a more realistic representation of the Antarctic Circumpolar Current. The structure of the gradient is also more realistic in the model run with assimilation, especially in the Amundsen Sea and Ross sea sector.

Assuming a jet with a Gaussian velocity profile, one can determine the characteristics of the

polar front by fitting the error function on the mean sea surface height h (Gille, 1994):

$$h(y) = a + (b - a)\,\mathrm{erf}\left(\frac{y - p}{w}\right) \tag{21}$$

where y is the latitude, p is the position, w is the width, a and b are the hypothetical values of the sea surface height if y could tend to $-\infty$ and $+\infty$ (respectively). Tests were performed to include an additional term to separate the polar and the sub-polar front, however as SSH corresponds to mean over a long time period and given the coarse resolution of the model results, the sea surface height h did not contain sufficient details to distinguish these two fronts. The proposed fit corresponds thus to the overall change of SSH over the frontal system. This fit has been performed on the ORCA2 model grid (also for the CNES-CLS09 MDT) and repeated for all longitudes of the model grid. The model run without assimilation reproduces relatively well the position of the front (Figure 14). The RMS error of the position (averaged over all longitudes) is 1.70°. While the assimilation can locally degrade the position of the front, it reduces on average the RMS error to 1.61°. The overall structure of the width of the front agrees with the width determined from the CNES-CLS09 MDT. However, the width in the free model run is overestimated, indicating that the model is too smooth and the ACC (Antarctic Circumpolar Current) is too diffuse. While the width of the front in the analysis is still too large, the assimilation improves its representation and the RMS error is reduced from 5.96° to 3.27°.

7. Identification of model errors

The aim of this section is the proposal of a technique for the identification of model errors during the assimilation cycle and its application on the proposed reanalysis for sea ice. Model errors can be traced by considering how the data assimilation system tends to pull the analysis away from the background towards the observations. Such approach was pioneered by Klinker and Sardeshmukh (1992) and further developed by Schubert and Chang (1996) and by Rodwell and Palmer (2007). The problem is approached here from a different perspective by the use of post-processing techniques and rigorous theoretical considerations. As argued

in Vannitsem and Nicolis (2008), forecasts at small lead times can be corrected using Model Output Statistics (MOS) techniques in case systematic model errors are present. Random initial-condition errors, on the other hand, cannot be corrected. Importantly, additional corrections can be obtained by consideration of an additional predictor (other than the one corresponding to the predictand) *in case this predictor is strongly correlated to the model error present*. The aim of this section is to diagnose the presence of model errors by seeking additional predictors that strongly correct the forecast. The identification of good predictors may then lead to an increased understanding of the source of error.

The data set considered consists of the ensemble-mean of the NEMO-LIM2 reanalysis. The observations against which we calibrate and compare the results are the aforementioned OSTIA data set.

7.1. Correction based on post-processing techniques

Three predictands or corrected forecasts are constructed: The *total Antarctic sea ice area*, the *Antarctic sea ice area as a function of longitude* and the *Antarctic sea ice area as a function of latitude*. Tests pointed out that the point-by-point or full-field sea ice concentrations could not be corrected with the post-processing methods. Full-field calibration methods based on EOF analysis might be an alternative that is worth investigating in the future (Di Giuseppe et al., 2013), but the present analysis is restricted to spatially integrated quantities.

The corrected forecast x^c is obtained by use of two predictors or model variables, x_1^f and x_2^f, based on the following regression relation:

$$x^c = \theta\left(\beta_0 + \beta_1 x_1^f + \beta_2 x_2^f\right). \tag{22}$$

Here the function θ ensures that the corrected sea ice area x^c is nonnegative:

$$\theta(z) = z \text{ when } z \geq 0, \text{ and, } \theta(z) = 0 \text{ when } z < 0. \tag{23}$$

The regression coefficients β_0, β_1 and β_2 are obtained by numerical minimization of the mean squared error, analogous to the technique of Linear Model Output Statistics (LMOS)

as discussed by Vannitsem and Nicolis (2008):

$$MSE = \left\langle (x_n^c - y_n^o)^2 \right\rangle_n , \tag{24}$$

where $\langle \cdot \rangle_n$ represents the average over the training data set and y^o denotes the observation. Three correction methods are used here:

- *Bias correction*: $\beta_1 = 1$ and $\beta_2 = 0$ while β_0 is an optimized parameter.
- *One-predictor correction*: $\beta_2 = 0$ while β_0 and β_1 are optimized parameters.
- *Two-predictor correction*: β_0, β_1 and β_2 are optimized parameters.

As the first predictor x_1^f the model variable corresponding to the predictand is taken. For the second predictand x_2^f, on the other hand, the following variables are considered, all taken from the 5-day forecast: ice thickness, ice y-velocity, ice x-velocity, sea surface height, barotropic stream function trends, sea surface height mean, sea surface salinity mean, sea surface temperature mean, sea surface x (meridional) and y (zonal) mean velocity, divergence and rotational components of horizontal velocity components, salinity, temperature, x-velocity, y-velocity and turbulent kinetic energy. As the model uses a leap-frog time stepping scheme, for some of the variable two consecutive time steps are available and have been used as predictors. Extra second predictors are constructed by full-field transformations of the aforementioned variables. More specifically, the totally advected, the longitudinally-advected and latitudinally-advected quantities are obtained by multiplying the variables with the total velocity or the longitudinal or latitudinal surface velocities, respectively. Also full-field multiplications are performed with the sea ice concentration SIC, with 1-SIC and with SIC(1-SIC) in order to obtain predictors that are only nonzero over sea ice, over open sea or near the sea ice edge, respectively. For non-surface variables we consider also the vertically-averaged (oceanic) quantities.

For correcting the predictand Antarctic sea ice area, all predictors are averages over the oceanic area south of 50°S. Similarly the predictors tested to correct the predictands that are a function of longitude or latitude are model variables averaged along the same latitudes

and longitudes, all south of 50°S. Note that for each of the three correction methods the regression coefficients are calculated separately. Also, longitude-by-longitude (latitude-by-latitude) analysis is performed for the predictands that are a function of longitude (latitude). Verification scores are obtained by correcting data subsets that are independent from the ones used to obtain the regression coefficients. More specifically, a cross validation is performed by which each single calendar year is corrected using the coefficients that were trained on the other 20 years.

7.2. Results of post-processing

The overall impact of all correction schemes on the Antarctic sea ice area is tabulated in Table 3 showing in the first column the RMSE values of the corrected forecasts, relative to the RMSE of the uncorrected 5-day forecast. Our post-processing technique was also applied on the analysis data (taking $x^f \rightarrow x^a$ in Eq. (22)) for which results are shown in the right-most column of Table 3.

A bias correction of the forecast amounts to a 5% RMSE reduction as compared to the uncorrected forecast while an additional variability correction (or one-predictor correction) yields 4% of extra reduction. By far the strongest correction (45% extra reduction) derives from the use of the best two-predictor correction. The situation is clearly different for post-processing applied to the analysis where all correction methods pull the analysis away from the observations (increased RMSE).

The main second predictor that reduces the RMSE associated with the 5-day *forecast* is model sea-surface temperature (SST). Almost all other (second) predictors that are not directly related to the model SST improve marginally or deteriorate upon the one-predictor forecast. Using the model SST (averaged south of 50°S) as a second predictor to correct the antarctic sea ice area, the RMSE is reduced with 50% as compared to the one-predictor forecast (see Table 3). Correcting the Antarctic sea ice area as a function of longitude using the model SST amounts to a reduction of 12%. This is a strong indication of a model error correlated with SST that considerably affects the forecast of the sea ice area. The global

view of the modeling impact on forecasts allows for emphasizing the dominant role played by model errors associated with sea surface temperature forecasts. Improvements will therefore be expected provided a better representation of sea surface temperature is achieved.

Fig. 15 shows the average Antarctic sea ice area for the different forecasts, the analysis and the observation as a function of the day of the year and the RMSE associated with these forecasts is given in Fig. 16. Clearly the RMSE of the two-predictor corrected forecast has the weakest seasonal cycle. Analogously, Fig. 17 depicts the longitudinal RMSE dependence of the sea ice area as function of longitude. The strongest two-predictor corrections are obtained in the Ross and Weddell seas and during Antarctic summer.

The best two-predictor correction scheme for the analysis leads to a larger value of the RMSE. As discussed in details in Vannitsem and Nicolis (2008), the absence of correction of the post-processing approach indicates that no model errors nor initial biases (related to the observations) are affecting the analyses, or in other words that the sole error present in the analysis is a random initial condition error and that the data assimilation scheme has made a proper use of the observations.

In turn the presence of a purely random initial condition error affecting the forecast step of the data assimilation scheme allows for concluding that the large biases of the five-day forecast are predominantly induced by a model error strongly correlated to the model SST. Even though so far no specific modeling scheme - such as horizontal turbulent transport - is pinpointed as the source of model error, we believe that progress can be made by considering other predictors more related to some specific parameterization schemes. This question is worth addressing in the future.

Once a second predictor providing substantial corrections is found, the variables or parameterization tendencies that strongly affect this predictor must be used to define more specific predictors for the post-processing scheme. Since, in our case, the model error is strongly correlated with SST, new predictors related to surface heat fluxes, ice melting or freezing, or the parameterization of eddy-induced mixing at sub-grid scales could be good candidates.

Once the observables responsible of the model error are isolated, the parameterization scheme should be reassessed, and sensitivity analyses based for instance on adjoint models could be performed.

Note that the post-processing approach as proposed here is not equivalent to finding variables that are highly correlated with the observations. In addition the use of different interpolation schemes could affect the amplitude of the absolute RMSE values, but the ratio between the best two-predictor scheme and the one-predictor scheme is not affected, suggesting the robustness of the conclusions.

8. Conclusions

This study shows that sea ice drift can be used to correct the wind field over the Southern Ocean as the model sea ice drift and 3-day mean surface wind field are strongly correlated. This relationship was used to adjust the wind field using pseudo-wind field observations based on sea ice drift data. As expected, the model using the adjusted wind field produces results closer to the sea ice drift data. But the adjusted wind field is also closer to the Cross-Calibrated Multi-Platform Ocean Surface Wind field (based on ERA-40 and observations). The impact on sea ice concentration was also assessed. However, only a small error reduction was found which suggests that only a small part of the model error in sea ice concentration is due to the wind fields. Comparison of the adjusted wind fields with direct wind observations would be useful to further assess the validity of the wind corrections.

Based on this adjusted wind, a reanalysis using the global NEMO model ORCA2 for the period 1985 to 2006 using 50 ensemble members has been presented. This model assimilates sea surface temperature, sea ice concentration and sea ice drift. The sea ice concentration assimilation used a Gaussian anamorphosis to transform this variable into a variable which follows more closely a Gaussian distribution. This transformation resulted in an improvement of the sea ice forecast. Despite this improvement being relatively small, the cost in term of CPU time is vanishingly small compared to the ensemble forecast and the analysis.

Finally, the reanalysis was compared to the World Ocean Database which is an independent data set. The assimilation was able to reduce the overall RMS error and bias of the

model compared to *in situ* temperature and salinity profiles. As the focus of the reanalysis is the Southern Ocean, the impact of the assimilation on the ACC (Antarctic Circumpolar Current) was also assessed by comparing the mean sea surface height of the model to the mean dynamic topography derived from various observations. The assimilation improved in general the mean surface height of the model in the Southern Ocean. In particular, the overall position and strength of the ACC was closer to observations after the assimilation.

Data assimilation is not directly suited to correct or diagnose consistent model errors since these are usually considered as random uncorrelated processes (except when the model bias is related to errors in the model parameters which can be estimated using variational assimilation or using a Kalman filter with an augmented state vector). The post-processing technique known as model output statistics attempts to relate a series of past forecast variables with the corresponding observations and is commonly used in numerical weather predictions for improving forecasts by reducing the impact of model errors. This technique can also be used to identify the presence of model errors by means of the analysis of the forecast improvements obtained using multiple predictors (Vannitsem and Nicolis, 2008). Indeed the potential forecast improvement based on a predictor reflects the presence of model errors (systematic or not) strongly correlated with this specific predictor. In the present investigation, the cross-validated RMS error of the 5-day forecast for the total Antarctic sea ice area could be halved using the SST forecast (averaged south of 50°S) as predictor. This indicates that SST is an important predictor strongly affected by the modelling error. This finding constitutes a first step to the identification of the underlying modeling scheme at the origin of the model error affecting the forecast. The post-processing technique was also applied on the analysis but was unable to reduce the RMS error, indicating that there is no obvious systematic error affecting the sea ice analysis.

9. Acknowledgments

This work was funded by the project PREDANTAR (SD/CA/04A) from the federal Belgian Science policy and the Sangoma FP7-SPACE-2011 project (grant 283580). François

Massonnet is a F.R.S. - FNRS Post-Doctoral Fellow and Alexander Barth a F.R.S. - FNRS Research Associate. NCEP/NCAR Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/. The UK Met Office, EUMETSAT OSI-SAF and MyOcean are acknowledged for providing the OSTIA/OSI-SAF sea surface temperature and sea ice concentration. We also thank the National Snow and Ice Data Center for providing the ice drift data, the World Ocean Data Base for the in situ temperature and salinity profiles, CNES and CLS for the mean dynamic topography. This is a MARE publication.

D. G. Vaughan, J. C. Comiso, I. Allison, J. Carrasco, R. Kwok, P. Mote, T. Murray, F. Paul, J. Ren, E. Rignot, O. Solomina, K. Steffen, T. Zhang, Observations:Cryosphere, in: T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

I. Eisenman, W. N. Meier, J. R. Norris, A spurious jump in the satellite record: has Antarctic sea ice expansion been overestimated?, The Cryosphere 8 (4) (2014) 1289–1296, doi:10.5194/tc-8-1289-2014.

J. Turner, J. Overland, Contrasting climate change in the two polar regions, Polar Research 28 (2) (2009) 146–164, ISSN 1751-8369, doi:10.1111/j.1751-8369.2009.00128.x.

S. Solomon, Stratospheric ozone depletion: A review of concepts and history, Rev. Geophys. 37 (3) (1999) 275–316, doi:10.1029/1999RG900008.

C. M. Bitz, L. M. Polvani, Antarctic climate response to stratospheric ozone depletion in a fine resolution ocean climate model, Geophys. Res. Lett. 39 (20), doi:10.1029/2012GL053393.

780 K. L. Smith, L. M. Polvani, D. R. Marsh, Mitigation of 21st century Antarctic sea ice loss
781 by stratospheric ozone recovery, Geophys. Res. Lett. 39 (20), doi:10.1029/2012GL053325.

782 M. Sigmond, J. C. Fyfe, The Antarctic Sea Ice Response to the Ozone Hole in Climate
783 Models, Journal of Climate 27 (3) (2013) 1336–1342, doi:10.1175/JCLI-D-13-00590.1.

784 J. Zhang, Increasing Antarctic Sea Ice under Warming Atmospheric and Oceanic Conditions,
785 Journal of Climate 20 (2007) 2515–2529, doi:10.1175/JCLI4136.1.

786 S. E. Stammerjohn, D. G. Martinson, R. C. Smith, X. Yuan, D. Rind, Trends in Antarctic
787 annual sea ice retreat and advance and their relation to El Niño Southern Oscillation and
788 Southern Annular Mode variability, J. Geophys. Res. 113 (C3), doi:10.1029/2007JC004269.

789 H. Goosse, W. Lefebvre, A. de Montety, E. Crespin, A. Orsi, Consistent past half-century
790 trends in the atmosphere, the sea ice and the ocean at high southern latitudes, Climate
791 Dynamics 33 (7) (2009) 999–1016–1016, doi:10.1007/s00382-008-0500-9.

792 C. H. Kirkman, C. M. Bitz, The Effect of the Sea Ice Freshwater Flux on Southern Ocean
793 Temperatures in CCSM3: Deep-Ocean Warming and Delayed Surface Warming, Journal
794 of Climate 24 (9) (2011) 2224–2237, doi:10.1175/2010JCLI3625.1.

795 L. Landrum, M. M. Holland, D. P. Schneider, E. Hunke, Antarctic Sea Ice Climatology,
796 Variability, and Late Twentieth-Century Change in CCSM4, Journal of Climate 25 (14)
797 (2012) 4817–4838, doi:10.1175/JCLI-D-11-00289.1.

798 P. R. Holland, R. Kwok, Wind-driven trends in Antarctic sea-ice drift, Nature Geoscience 5
799 (2012) 872–875, doi:10.1038/ngeo1627.

800 R. Bintanja, G. J. van Oldenborgh, S. S. Drijfhout, B. Wouters, C. A. Katsman, Important
801 role for ocean warming and increased ice-shelf melt in Antarctic sea-ice expansion, Nature
802 Geoscience 6 (5) (2013) 376–379, doi:10.1038/ngeo1767.

803 H. Goosse, V. Zunz, Decadal trends in the Antarctic sea ice extent ultimately controlled by
804 ice–ocean feedback, The Cryosphere 8 (2) (2014) 453–470, doi:10.5194/tc-8-453-2014.

C. de Lavergne, J. B. Palter, E. D. Galbraith, R. Bernardello, I. Marinov, Cessation of deep convection in the open Southern Ocean under anthropogenic climate change, Nature Clim. Change 4 (4) (2014) 278–282, doi:10.1038/nclimate2132.

I. Mahlstein, P. R. Gent, S. Solomon, Historical Antarctic mean sea ice area, sea ice trends, and winds in CMIP5 simulations, Journal of Geophysical Research: Atmospheres 118 (2013) 1–6, ISSN 2169-8996, doi:10.1002/jgrd.50443.

V. Zunz, H. Goosse, F. Massonnet, How does internal variability influence the ability of CMIP5 models to reproduce the recent trend in Southern Ocean sea ice extent?, The Cryosphere 7 (2) (2013) 451–468, doi:10.5194/tc-7-451-2013.

L. M. Polvani, K. L. Smith, Can natural variability explain observed Antarctic sea ice trends? New modeling evidence from CMIP5, Geophys. Res. Lett. 40 (12) (2013) 3195–3199, ISSN 1944-8007, doi:10.1002/grl.50578.

N. C. Swart, J. C. Fyfe, The influence of recent Antarctic ice sheet retreat on simulated sea ice area trends, Geophysical Research Letters 40 (16) (2013) 4328–4332, doi:10.1002/grl.5082.

C. L. Parkinson, D. J. Cavalieri, Antarctic sea ice variability and trends, 1979–2010, The Cryosphere 6 (4) (2012) 871–880, ISSN 1994-0424, doi:10.5194/tc-6-871-2012.

K. E. Taylor, R. J. Stouffer, G. A. Meehl, An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society 93 (4) (2011) 485–498, doi:10.1175/BAMS-D-11-00094.1.

J. Turner, T. J. Bracegirdle, T. Phillips, G. J. Marshall, J. S. Hosking, An Initial Assessment of Antarctic Sea Ice Extent in the CMIP5 Models, Journal of Climate 26 (5) (2013) 1473–1484, doi:10.1175/JCLI-D-12-00068.1.

T. Fichefet, M. A. Morales Maqueda, Modelling the influence of snow accumulation and snow-ice formation on the seasonal cycle of the Antarctic sea-ice cover, Climate Dynamics 15 (1999) 251–268, doi:10.1007/s003820050280.

T. Fichefet, H. Goosse, M. A. Morales Maqueda, A hindcast simulation of Arctic and Antarctic sea ice variability, 1955–2001, Polar Research 22 (2003) 91–98, doi:10.1111/j.1751-8369.2003.tb00100.x.

P. R. Holland, N. Bruneau, C. Enright, M. Losch, N. T. Kurtz, R. Kwok, Modeled Trends in Antarctic Sea Ice Thickness, J. Climate 27 (10) (2014) 3784–3801, ISSN 0894-8755, doi:10.1175/JCLI-D-13-00301.1.

D. H. Bromwich, R. L. Fogt, K. I. Hodges, J. E. Walsh, A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions, Journal of Geophysical Research 112 (2007) D10111, doi:10.1029/2006JD007859.

K. Hines, D. H. Bromwich, G. J. Marshall, Artificial Surface Pressure Trends in the NCEP/NCAR Reanalysis over the Southern Ocean and Antarctica, Journal of Climate 13 (2000) 3940–3952, doi:10.1175/1520-0442(2000)013⟨3940:ASPTIT⟩2.0.CO;2.

M. Vancoppenolle, R. Timmermann, S. F. Ackley, T. Fichefet, H. Goosse, P. Heil, K. C. Leonard, J. Lieser, M. Nicolaus, T. Papakyriakou, J.-L. Tison, Assessment of radiation forcing data sets for large-scale sea ice models in the Southern Ocean, Deep Sea Research Part II: Topical Studies in Oceanography 58 (2011) 1237–1249, ISSN 0967-0645, doi:10.1016/j.dsr2.2010.10.039.

R. Timmerman, A. P. Worby, H. Goosse, T. Fichefet, Utilizing the ASPeCt sea ice thickness data set to evaluate a global coupled sea ice–ocean model, Journal of Geophysical Research 109 (2004) C07017, doi:10.1029/2003JC002242.

A. Stössel, Z. Zhang, T. Vihma, The effect of alternative real-time wind forcing on Southern Ocean sea ice simulations, Journal of Geophysical Research 116 (2011) C11021, doi:10.1029/2011JC007328.

D. Stammer, C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C. N. Hill, J. Marshall, Global ocean circulation during 1992–1997, estimated from ocean

observations and a general circulation model, J.-Geophys.-Res. 107 (C9) (2002) 3118, ISSN 2156-2202, doi:10.1029/2001JC000888.

A. Stössel, Employing Satellite-Derived Sea Ice Concentration to Constrain Upper-Ocean Temperature in a Global Ocean GCM, J. Climate 21 (17) (2008) 4498–4513, ISSN 0894-8755, doi:10.1175/2008JCLI2256.1.

F. Massonnet, P. Mathiot, T. Fichefet, H. Goosse, C. K. Beatty, M. Vancoppenolle, T. Lavergne, A model reconstruction of the Antarctic sea ice thickness and volume changes over 1980-2008 using data assimilation, Ocean Modelling 64 (2013) 67–75, ISSN 1463-5003, doi:10.1016/j.ocemod.2013.01.003.

N. Ferry, L. Parent, G. Garric, C. Bricaud, C. E. Testut, O. Le Galloudec, J.-M. Lellouche, M. Drévillon, E. Greiner, B. Barnier, J. M. Molines, N. Jourdain, S. Guinehut, C. Cabanes, L. Zawadski, GLORYS2V1 Global ocean reanalysis of the altimetric era (1993-2009) at meso scale, Tech. Rep., Mercator Ocean, 2012.

M. A. Balmaseda, A. Vidard, D. L. T. Anderson, The ECMWF Ocean Analysis System: ORA-S3, Mon. Wea. Rev. 136 (8) (2008) 3018–3034, ISSN 0027-0644, doi:10.1175/2008MWR2433.1.

J. A. Carton, B. S. Giese, A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA), Mon. Wea. Rev. 136 (8) (2008) 2999–3017, ISSN 0027-0644, doi:10.1175/2007MWR1978.1.

T. Janjić, J. Schröter, R. Savcenko, W. Bosch, A. Albertella, R. Rummel, O. Klatt, Impact of combining GRACE and GOCE gravity data on ocean circulation estimates, Ocean Science 8 (1) (2012) 65–79, doi:10.5194/os-8-65-2012.

P. Sakov, F. Counillon, L. Bertino, K. A. Lisæter, P. R. Oke, A. Korablev, TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Science 8 (4) (2012) 633–656, doi:10.5194/os-8-633-2012.

880 P. Mathiot, C. König Beatty, T. Fichefet, H. Goosse, F. Massonnet, M. Vancoppenolle, Better
881 constraints on the sea-ice state using global sea-ice data assimilation, Geosci. Model Dev.
882 5 (6) (2012) 1501–1515, ISSN 1991-9603, doi:10.5194/gmd-5-1501-2012.

883 L. Bertino, G. Evensen, H. Wackernagel, Sequential Data Assimilation Techniques
884 in Oceanography, International Statistical Review 71 (2003) 223–241, doi:10.1111/j.
885 1751-5823.2003.tb00194.x.

886 E. Simon, L. Bertino, Application of the Gaussian anamorphosis to assimilation in a 3-D cou-
887 pled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment,
888 Ocean Science 5 (4) (2009) 495–510, doi:10.5194/os-5-495-2009.

889 D. Béal, P. Brasseur, J.-M. Brankart, Y. Ourmières, J. Verron, Characterization of mixing
890 errors in a coupled physical biogeochemical model of the North Atlantic: implications
891 for nonlinear estimation using Gaussian anamorphosis, Ocean Science 6 (2010) 247–262,
892 doi:10.5194/os-6-247-2010.

893 K. A. Lisæter, J. Rosanova, G. Evensen, Assimilation of ice concentration in a coupled
894 ice-ocean model, using the Ensemble Kalman filter, Ocean Dynamics 53 (2003) 368–388,
895 doi:10.1007/s10236-003-0049-4.

896 F. Massonnet, H. Goosse, T. Fichefet, F. Counillon, Calibration of sea ice dynamic parame-
897 ters in an ocean-sea ice model using an ensemble Kalman filter, J. Geophys. Res. Oceans
898 119 (7) (2014) 4168–4184, ISSN 2169-9291, doi:10.1002/2013JC009705.

899 A. Barth, A. Alvera-Azcárate, J.-M. Beckers, J. Staneva, E. V. Stanev, J. Schulz-Stellenfleth,
900 Correcting surface winds by assimilating High-Frequency Radar surface currents in the
901 German Bight, Ocean Dynamics 61 (5) (2011) 599–610, doi:10.1007/s10236-010-0369-0,
902 URL http://hdl.handle.net/2268/83330.

903 J. Marmain, A. Molcard, P. Forget, A. Barth, Y. Ourmières, Assimilation of HF radar surface
904 currents to optimize forcing in the northwestern Mediterranean Sea, Nonlinear Processes
905 Geophysics 21 (2014) 659–675, doi:10.5194/npg-21-659-2014.

H. Ngodock, M. Carrier, A 4DVAR System for the Navy Coastal Ocean Model. Part I: System Description and Assimilation of Synthetic Observations in Monterey Bay, Mon. Wea. Rev. 142 (6) (2014) 2085–2107, ISSN 0027-0644, doi:10.1175/MWR-D-13-00221.1.

N. Kimura, Sea ice motion in response to surface wind and ocean current in the Southern Ocean, Journal of the Meteorological Society of Japan 82 (2004) 1223–1231, doi:10.2151/jmsj.2004.1223.

P. Malanotte-Rizzoli, R. E. Young, D. B. Haidvogel, Initialization and data assimilation experiments with a primitive equation model, Dynamics of Atmospheres and Oceans 13 (1989) 349–378, doi:10.1016/0377-0265(89)90046-8.

A. Barth, J.-M. Beckers, A. Alvera-Azcárate, R. H. Weisberg, Filtering inertia-gravity waves from the initial conditions of the linear shallow water equations, Ocean Modelling 19 (2007) 204–218, doi:10.1016/j.ocemod.2007.06.007, URL http://hdl.handle.net/2268/4266.

Y. Yan, A. Barth, J. Beckers, Comparison of different assimilation schemes in a sequential Kalman filter assimilation system, Ocean Modelling 73 (0) (2014) 123–137, ISSN 1463-5003, doi:10.1016/j.ocemod.2013.11.002, URL http://hdl.handle.net/2268/163077.

G. Madec, NEMO ocean engine, no. 27 in Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, 2008.

T. Fichefet, M. A. Morales Maqueda, Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, Journal of Geophysical Research 102 (1997) 12609–12646, doi:10.1029/97JC00480.

R. Timmermann, H. Goosse, G. Madec, T. Fichefet, C. Ethe, V. Dulière, On the representation of high latitude processes in the ORCA-LIM global coupled sea ice-ocean model, Ocean Modelling 8 (1-2) (2005) 175–201, ISSN 1463-5003, doi:10.1016/j.ocemod.2003.12.009.

S. Bouillon, M. A. Morales Maqueda, V. Legat, T. Fichefet, An elastic-viscous-plastic sea ice model formulated on Arakawa B and C grids, Ocean Modelling 27 (2009) 174–184, doi:10.1016/j.ocemod.2009.01.004.

P. Mathiot, H. Goosse, T. Fichefet, B. Barnier, H. Gallée, Modelling the seasonal variability of the Antarctic Slope Current, Ocean Science 7 (4) (2011) 455–470, doi:10.5194/os-7-455-2011.

E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, D. Joseph, The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society 77 (1996) 437–471, doi:10.1175/1520-0477(1996)077⟨0437:TNYRP⟩2.0.CO;2.

H. Goosse, T. Fichefet, Importance of ice-ocean interactions for the global ocean circulation: A model study, Journal of Geophysical Research 104 (C10) (1999) 23337–23355, ISSN 2156-2202, doi:10.1029/1999JC900215.

J. Roberts-Jones, E. K. Fiedler, M. J. Martin, Daily, Global, High-Resolution SST and Sea Ice Reanalysis for 1985-2007 Using the OSTIA System, J. Climate 25 (2012) 6215–6232, doi:10.1175/JCLI-D-11-00648.1.

C. Fowler, Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors, Tech. Rep., NASA DAAC at the National Snow and Ice Data Center, Boulder, Colorado USA, 2003.

P. K. Kundu, J. S. Allen, Some three-dimensional characteristics of low-frequency current flucutuations near the Oregon coast, Journal of Physical Oceanography 6 (1976) 181–199, doi:10.1175/1520-0485(1976)006⟨0181:STDCOL⟩2.0.CO;2.

A. Alvera-Azcárate, A. Barth, D. Sirjacobs, J.-M. Beckers, Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF, Ocean Science 5 (2009) 475–485, doi:10.5194/os-5-475-2009, URL http://www.ocean-sci.net/5/475/2009/.

A. Barth, A. Alvera-Azcárate, R. H. Weisberg, Assimilation of high-frequency radar currents in a nested model of the West Florida Shelf, Journal of Geophysical Research 113 (2008) C08033, doi:10.1029/2007JC004585, URL http://hdl.handle.net/2268/26171.

R. Atlas, R. N. Hoffman, J. Ardizzone, S. M. Leidner, J. C. Jusem, D. K. Smith, D. Gombos, A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications, Bulletin of the American Meteorological Society 92 (2011) 157–174, doi:10.1175/2010BAMS2946.1.

A. Barth, J.-M. Beckers, C. Troupin, A. Alvera-Azcárate, L. Vandenbulcke, divand-1.0: n-dimensional variational data analysis for ocean observations, Geoscientific Model Development 7 (1) (2014) 225–241, doi:10.5194/gmd-7-225-2014, URL http://www.geosci-model-dev.net/7/225/2014/.

P. Brasseur, J.-M. Beckers, J.-M. Brankart, R. Schoenauen, Seasonal Temperature and Salinity Fields in the Mediterranean Sea: Climatological Analyses of an Historical Data Set., Deep–Sea Research 43 (2) (1996) 159–192, doi:10.1016/0967-0637(96)00012-X.

P. Mathiot, B. Barnier, H. Gallée, J. M. Molines, J. L. Sommer, M. Juza, T. Penduff, Introducing katabatic winds in global ERA40 fields to simulate their impacts on the Southern Ocean and sea-ice, Ocean Modelling 35 (3) (2010) 146 – 160, ISSN 1463-5003, doi:10.1016/j.ocemod.2010.07.001.

R. Massom, P. Harris, K. Michael, M. Potter, The distribution and formative processes of latent-heat polynyas in East Antarctica, Annals of Glaciology 27 (1998) 420–426.

M. A. Morales Maqueda, A. J. Willmott, N. R. T. Biggs, Polynya Dynamics: a Review of Observations and Modeling, Reviews of Geophysics 42 (1) (2004) RG1004, ISSN 1944-9208, doi:10.1029/2002RG000116.

H. Sumata, T. Lavergne, F. Girard-Ardhuin, N. Kimura, M. A. Tschudi, F. Kauker, M. Karcher, R. Gerdes, An intercomparison of Arctic ice drift products to deduce un-

981 certainty estimates, Journal of Geophysical Research: Oceans 119 (8) (2014) 4887–4921,
982 ISSN 2169-9291, doi:10.1002/2013JC009724.

983 C. H. Bishop, B. Etherton, S. J. Majumdar, Adaptive Sampling with the Ensemble Trans-
984 form Kalman Filter. Part I: Theoretical Aspects, Monthly Weather Review 129 (2001)
985 420–436, doi:10.1175/1520-0493(2001)129⟨0420:ASWTET⟩2.0.CO;2.

986 L. Vandenbulcke, M. Rixen, J.-M. Beckers, A. Alvera-Azcárate, A. Barth, An analysis of the
987 error space of a high-resolution implementation of the GHER hydrodynamic model in the
988 Mediterranean Sea, Ocean Modelling 24 (1-2) (2008) 46–64, doi:10.1016/j.ocemod.2008.
989 05.007, URL http://hdl.handle.net/2268/4263.

990 F. Lenartz, C. Raick, K. Soetaert, M. Gregoire, Application of an Ensemble Kalman filter
991 to a 1-D coupled hydrodynamicecosystem model of the Ligurian Sea, Journal of Marine
992 Systems 68 (2007) 327–348, doi:10.1016/j.jmarsys.2006.12.001.

993 Y. Chen, C. Snyder, Assimilating Vortex Position with an Ensemble Kalman Filter, Monthly
994 Weather Review 135 (2007) 1828–1845, doi:10.1175/MWR3351.1.

995 G. Evensen, Data assimilation: the Ensemble Kalman Filter, Springer, 279pp, 2007.

996 G. H. Golub, C. F. Van Loan, Matrix Computations, Johns Hopkins University Press, Bal-
997 timore, 3rd ed. edn., 1996.

998 L. Vandenbulcke, A. Barth, M. Rixen, A. Alvera-Azcárate, Z. B. Bouallegue, J.-M. Beck-
999 ers, Study of the combined effects of data assimilation and grid nesting in ocean
1000 models. Application to the Gulf of Lions., Ocean Science 2 (2) (2006) 213–222, doi:
1001 10.5194/os-2-213-2006.

1002 A. Alvera-Azcárate, A. Barth, Z. B. Bouallègue, M. Rixen, J.-M. Beckers, Forecast Verifi-
1003 cation of a 3D model of the Ligurian Sea. The use of Discrete Wavelet Transforms in the
1004 skill assessment of spatial forecasts, Journal of Marine Systems 65 (1–4) (2007) 460–483,
1005 doi:10.1016/j.jmarsys.2005.09.015, URL http://hdl.handle.net/2268/4264.

S. C. Bloom, L. L. Takacs, A. M. D. Silva, D. Ledvina, Data assimilation using incremental analysis updates, Monthly Weather Review 124 (1996) 1256–1271, doi:10.1175/1520-0493(1996)124⟨1256:DAUIAU⟩2.0.CO;2.

M. H. Rio, S. Guinehut, G. Larnicol, New CNES-CLS09 global mean dynamic topography computed from the combination of GRACE data, altimetry, and in situ measurements, Journal of Geophysical Research: Oceans 116 (C7) (2011) C07018, ISSN 2156-2202, doi:10.1029/2010JC006505.

S. T. Gille, Mean sea surface height of the Antarctic Circumpolar Current from Geosat data: Method and application, Journal of Geophysical Research: Oceans 99 (C9) (1994) 18255–18273, ISSN 2156-2202, doi:10.1029/94JC01172.

E. Klinker, P. Sardeshmukh, The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements, Journal of the atmospheric sciences 49 (7) (1992) 608–627, doi:10.1175/1520-0469(1992)049⟨0608:TDOMDI⟩2.0.CO;2.

S. Schubert, Y. Chang, An objective method for inferring sources of model error, Monthly Weather Review 124 (2) (1996) 325–340, doi:10.1175/1520-0493(1996)124⟨0325:AOMFIS⟩2.0.CO;2.

M. Rodwell, T. Palmer, Using numerical weather prediction to assess climate models, Quarterly Journal of the Royal Meteorological Society 133 (622) (2007) 129–146, doi:10.1002/qj.23.

S. Vannitsem, C. Nicolis, Dynamical properties of model output statistics forecasts, Monthly Weather Review 136 (2) (2008) 405–419, doi:10.1175/2007MWR2104.1.

F. Di Giuseppe, F. Molteni, A. M. Tompkins, A rainfall calibration methodology for impacts modelling based on spatial mapping, Quarterly Journal of the Royal Meteorological Society 139 (674) (2013) 1389–1401, doi:10.1002/qj.2019.

Table 1: Calibration of the assimilation configuration. A dash means no significant change.

RMS	state vec.	anam.	$\sqrt{\mathbf{R}_{ice}}$	SST	ice conc.	ui_ice	vi_ice
baseline experiment	full	yes	0.1	0.632	0.087	0.080	0.070
skill score (%)				SST	ice conc.	ui_ice	vi_ice
ExpFNA-0.1	full	no	0.1	-	-2.766	-	-
ExpFA-0.07	full	yes	0.07	-2.556	8.094	-	2.254
ExpRA-0.05	reduced	yes	0.05	-7.477	12.492	4.000	3.069
ExpRA-0.07	reduced	yes	0.07	-2.794	10.046	2.390	2.474
ExpRA-0.1	reduced	yes	0.1	-	13.454	-	2.279

Table 2: Total RMS error relative to assimilated data

	Free	Forecast	Analysis
SST [°C]	0.98	0.59	0.47
ice conc.	0.22	0.085	0.033
u_{ice} [m/s]	0.088	0.069	0.041
v_{ice} [m/s]	0.074	0.060	0.039

Table 3: Root Mean Squared Error (RMSE) of the Antarctic sea ice area of different correction methods applied on the 5-day forecasts (second column) and analysis (third column), all divided by the RMSE of the uncorrected 5-day forecast for Antarctic sea ice area.

Antarctic sea ice area	Forecast RMSE (%)	Analysis RMSE (%)
Uncorrected	100%	15.7%
Bias-corrected	95%	17.2%
One-predictor corrected	91%	18.2%
Two-predictor corrected (min. RMSE)	46%	16.6%

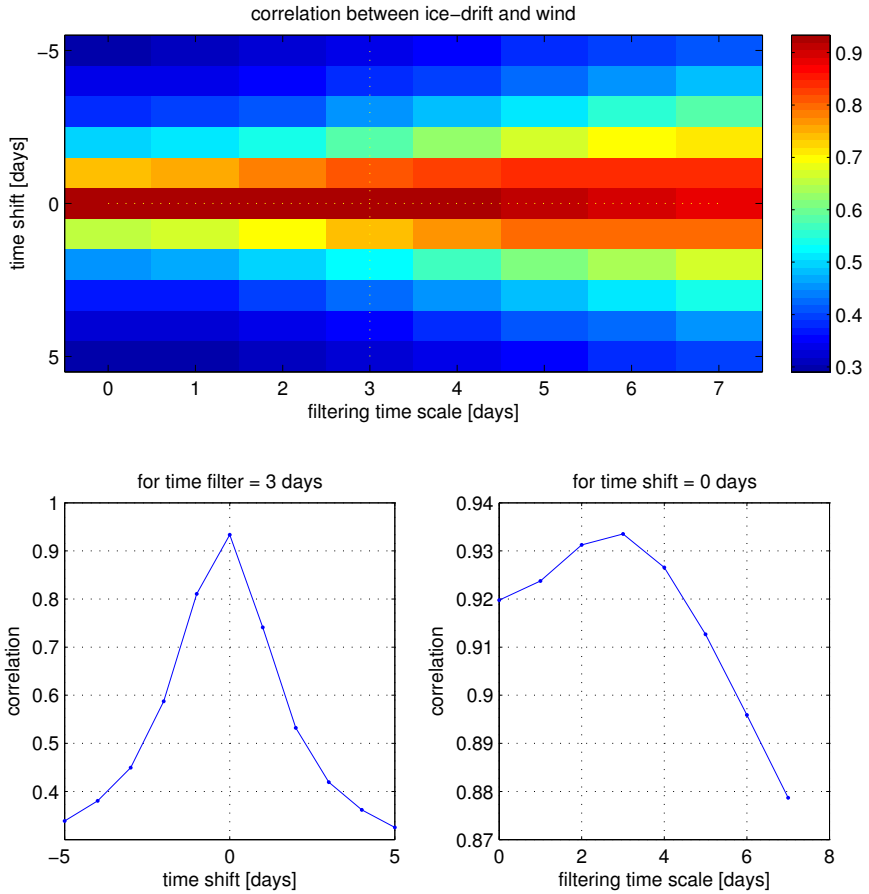

Figure 1: Magnitude of correlation coefficient (for the year 2000) as function of time shift and filtering time scale (panel a). Panel b show the magnitude of correlation coefficient as a function of the time shift for a filtering time scale sets to 3 days (vertical dotted line in panel a) and panel c show represents the magnitude of correlation coefficient as a function of filtering time scale for a the time shift set to 0 days (horizontal dotted line in panel a).

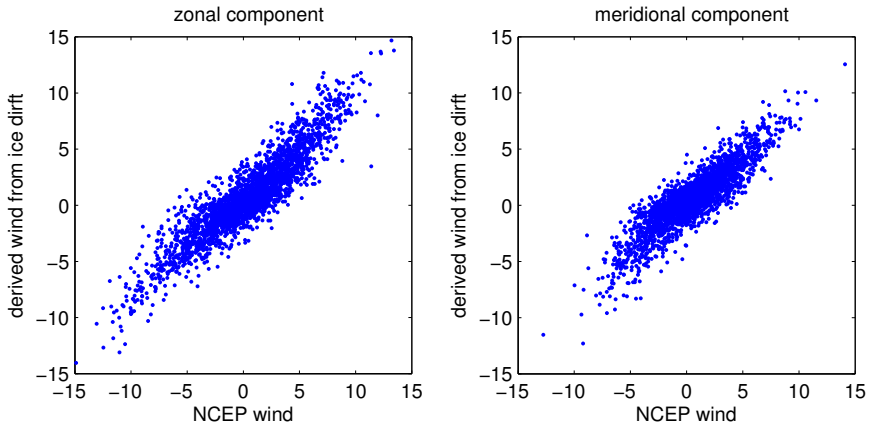

Figure 2: Scatter plot of NCEP wind versus wind estimated from ice drift (both in m/s) using the complex regression coefficient (with a filter of 3 days and without timeshift). The dots corresponds to the data from the year 2000 and to model grid points covered by ice in the Southern Hemisphere.

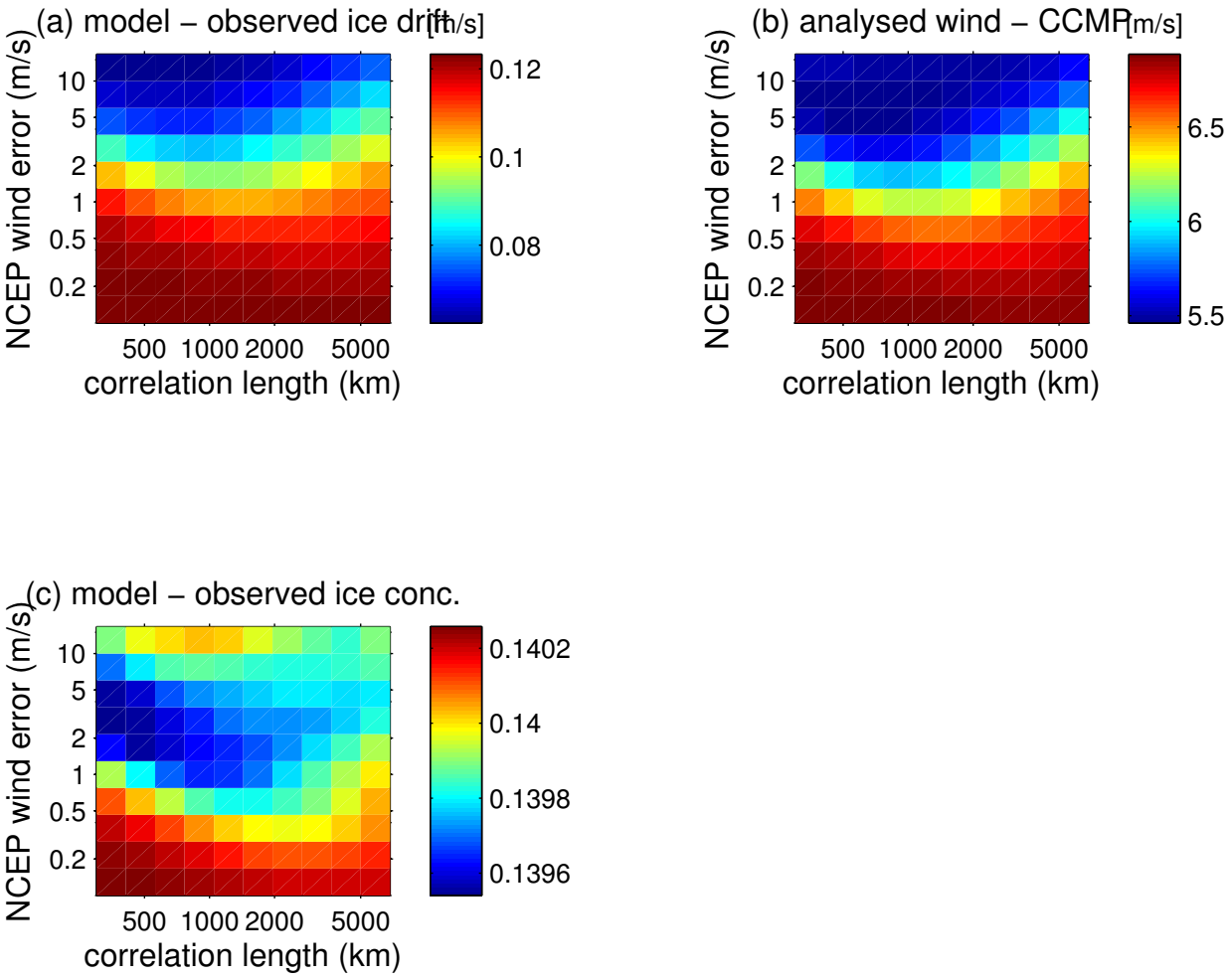

Figure 3: RMS difference of the model and observed sea ice drift (panel a), the analyzed winds and CCMP winds (panel b) and model and observed sea ice concentration (panel c) for different values of the correlation length and the NCEP wind error.

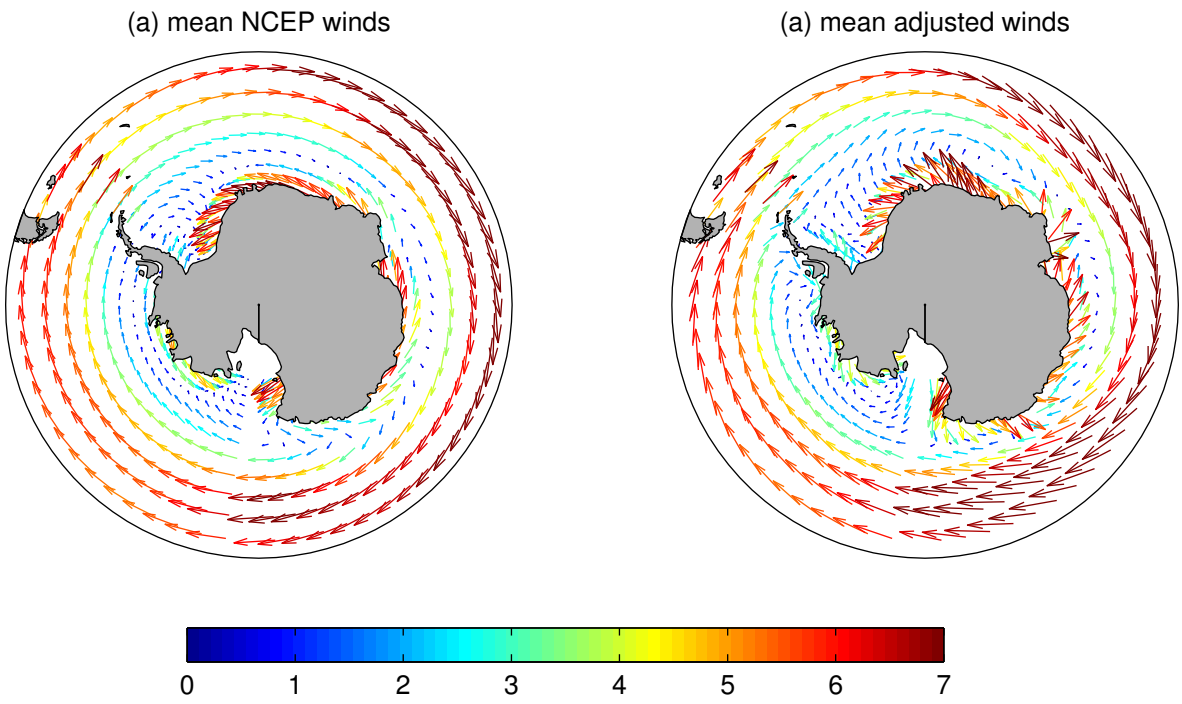

Figure 4: The mean NCEP wind vector (panel a) and the mean adjusted wind (panel b) averaged over the period from 1985 to 2006. The color represents the norm of the wind vector in m/s.

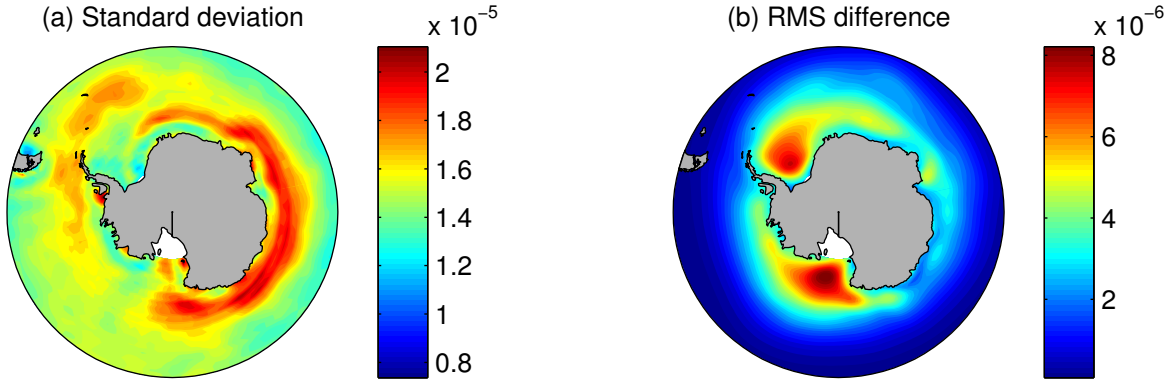

Figure 5: Standard deviation of NCEP wind curl (panel a) and RMS of the difference between the original NCEP wind curl and the adjusted wind curl (panel b). The units are s^{-1}.

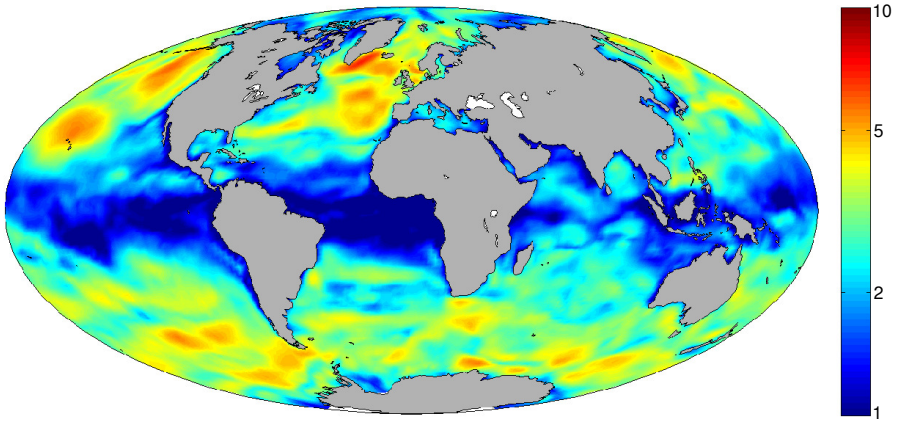

Figure 6: Wind speed ensemble standard deviation in m/s (2007-02-21).

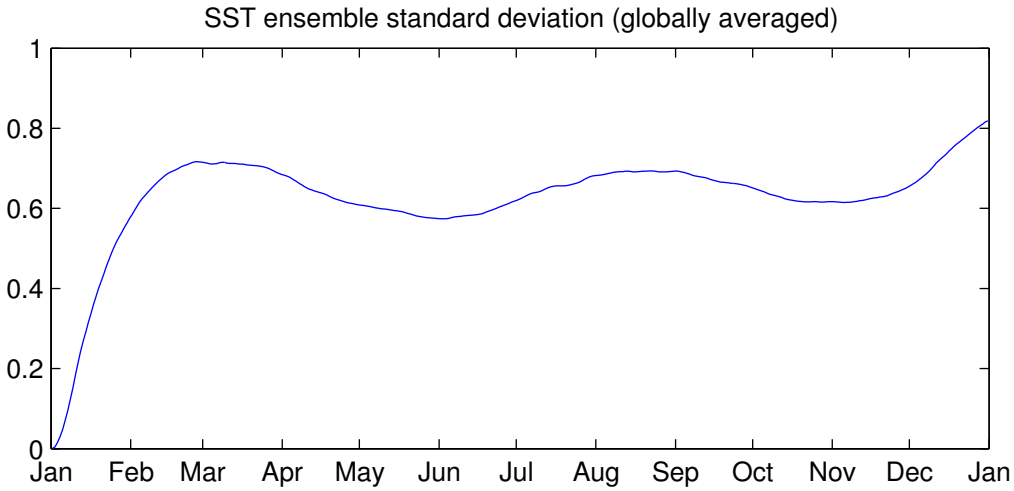

Figure 7: Globally-averaged SST ensemble standard deviation (°C). The spread is first computed for every model grid point and then averaged spatially.

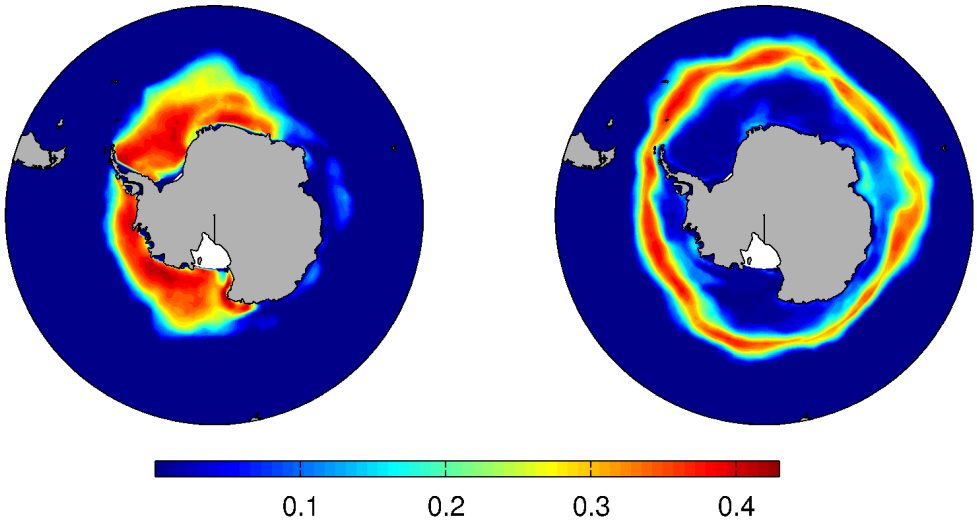

Figure 8: Sea ice concentration standard deviation at the minimum sea ice extent (2007-02-21, left panel) and during the maximum sea ice extent (2007-09-06, right panel)

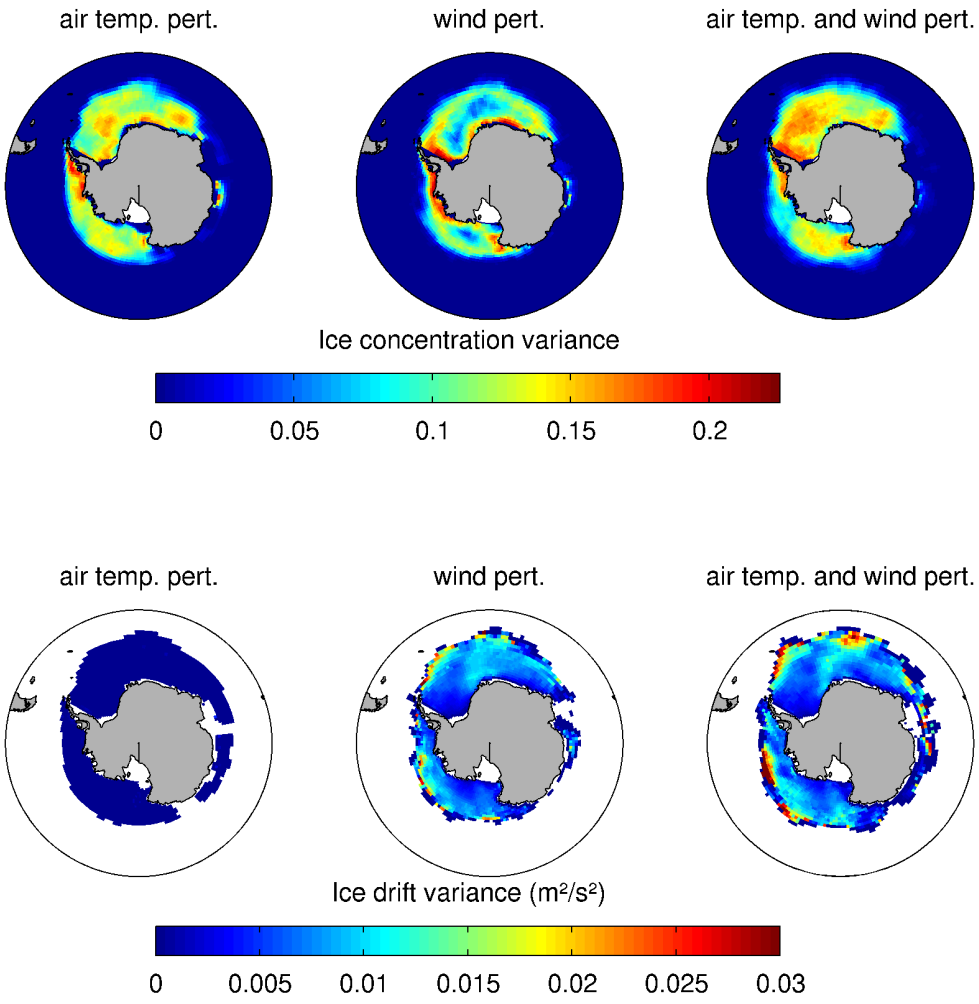

Figure 9: Ensemble variances of sea ice concentration (upper row) and sea ice drift (lower row) based on only air temperature perturbations, wind field perturbations or both. The ensemble variance corresponds to the starting day of the assimilation experiment (1985-01-01).

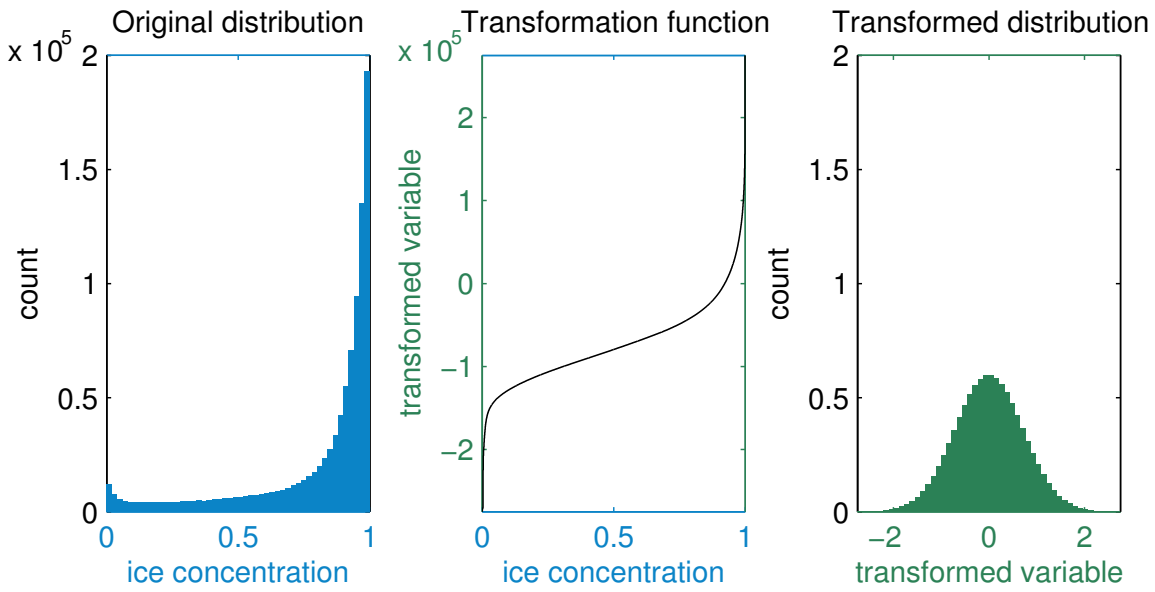

Figure 10: Illustration of the Gaussian anamorphosis: The left panel shows the original histogram of sea ice concentration, the middle panel represents the transformation function and the right panel is the resulting histogram.

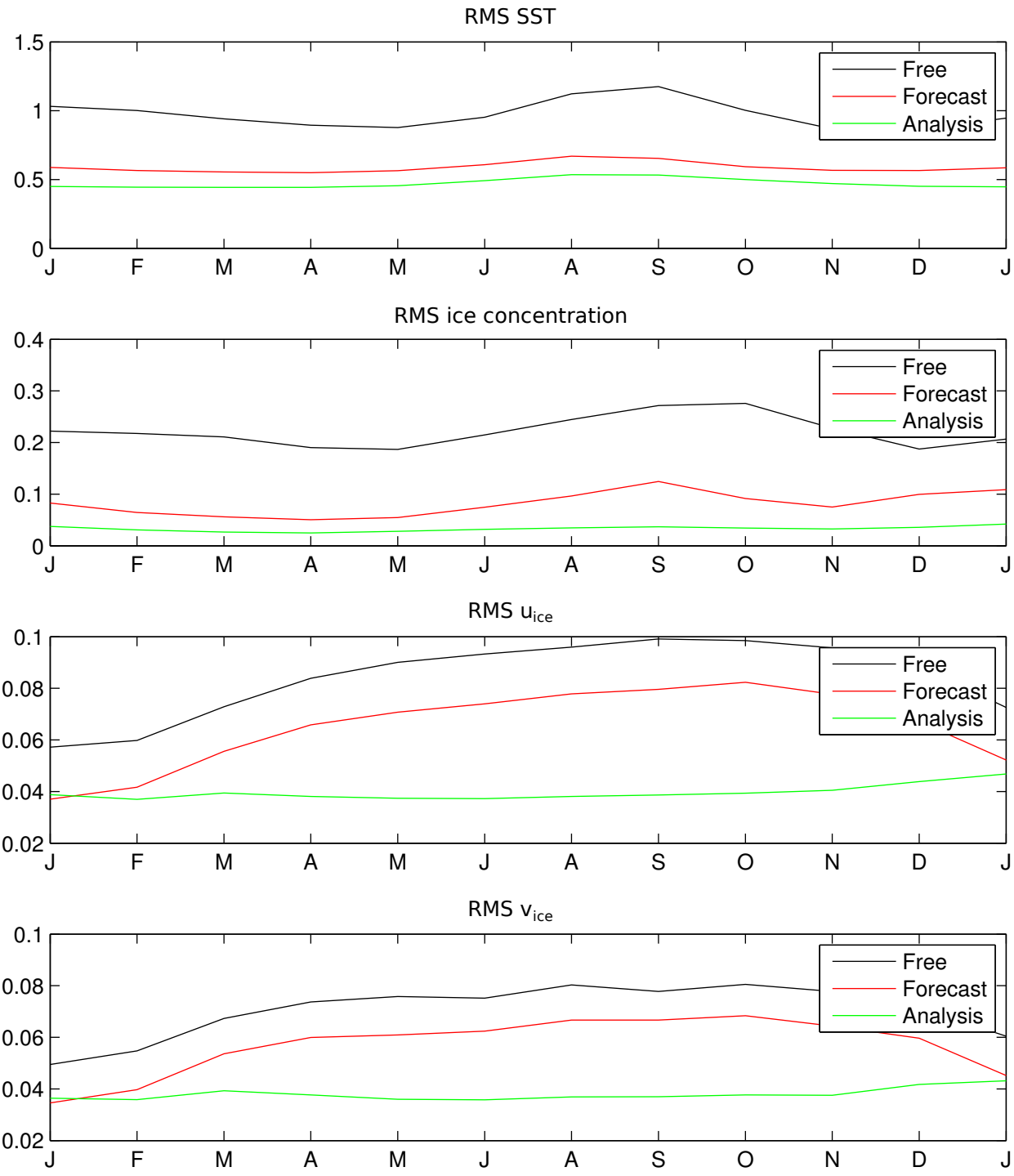

Figure 11: RMS error of the free running model and the model with assimilation (forecast and analysis) compared to the assimilated data for every month (x-axis) and averaged over all years.

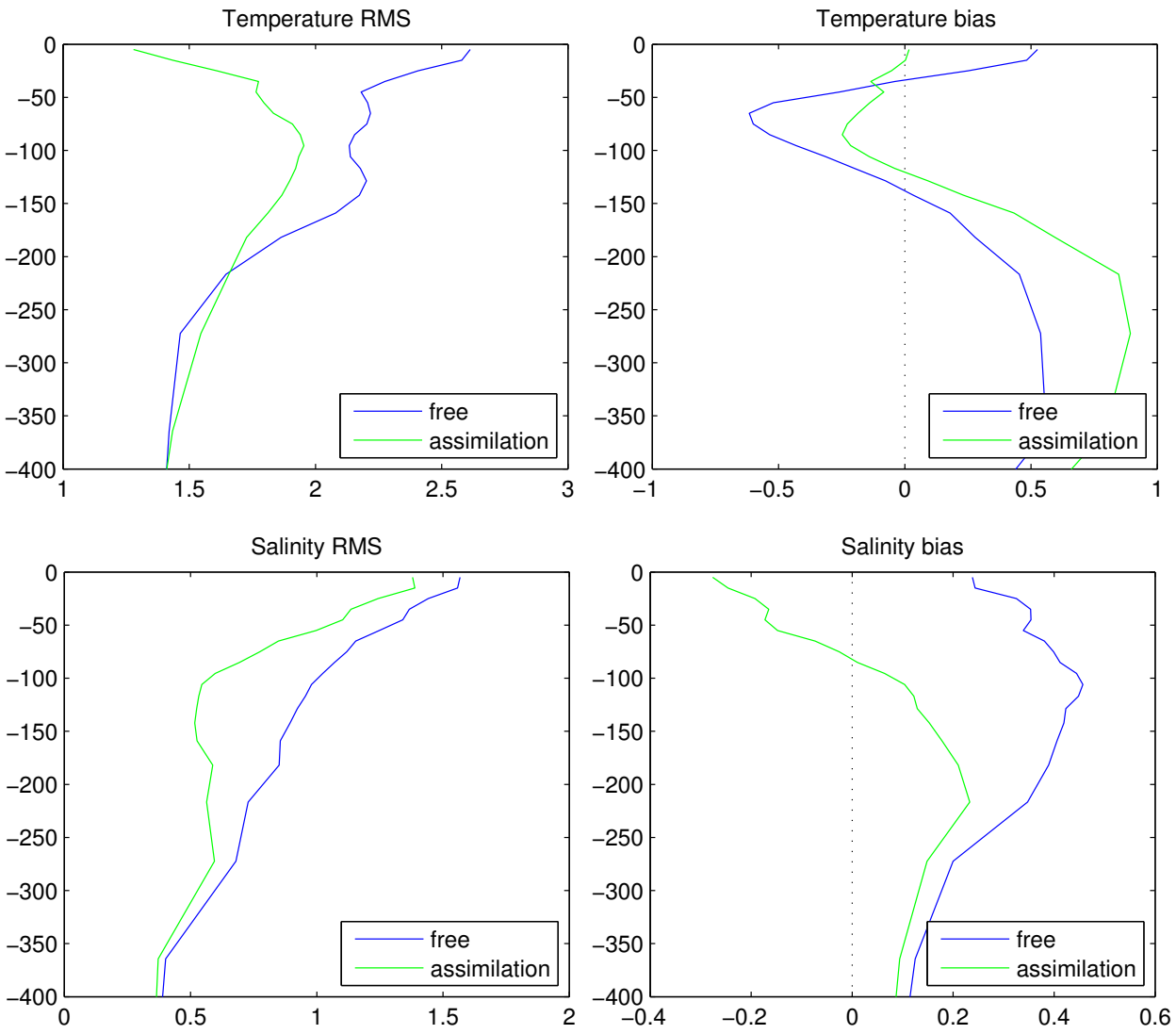

Figure 12: Validation with World Ocean Database using all observation from 1985-2006. The x-axis is temperature (upper row) or salinity (lower row) and the y-axis is depth.

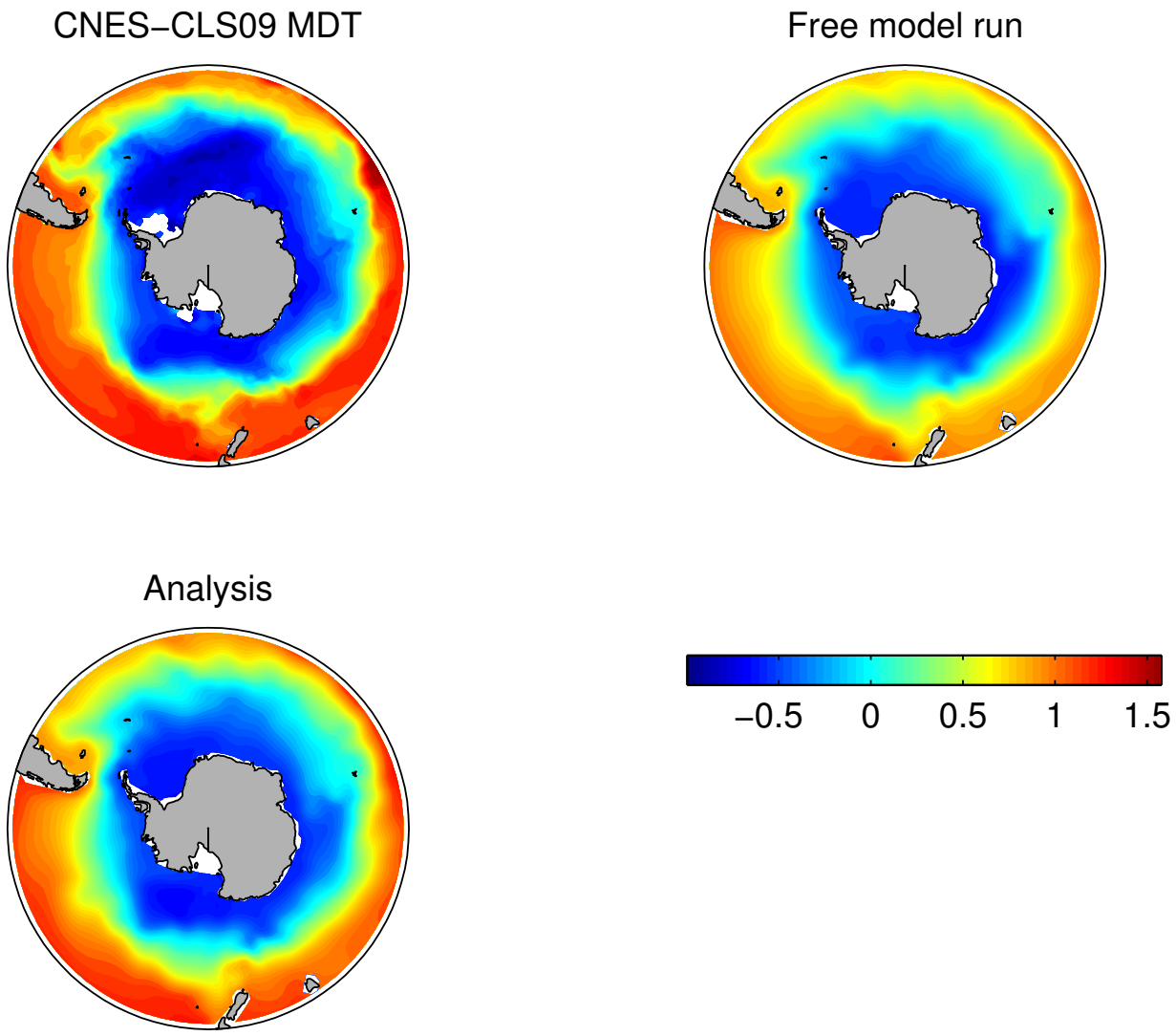

Figure 13: Comparison of mean SSH from observations and from the model (without and with assimilation). The spatial average of the shown domain was removed.

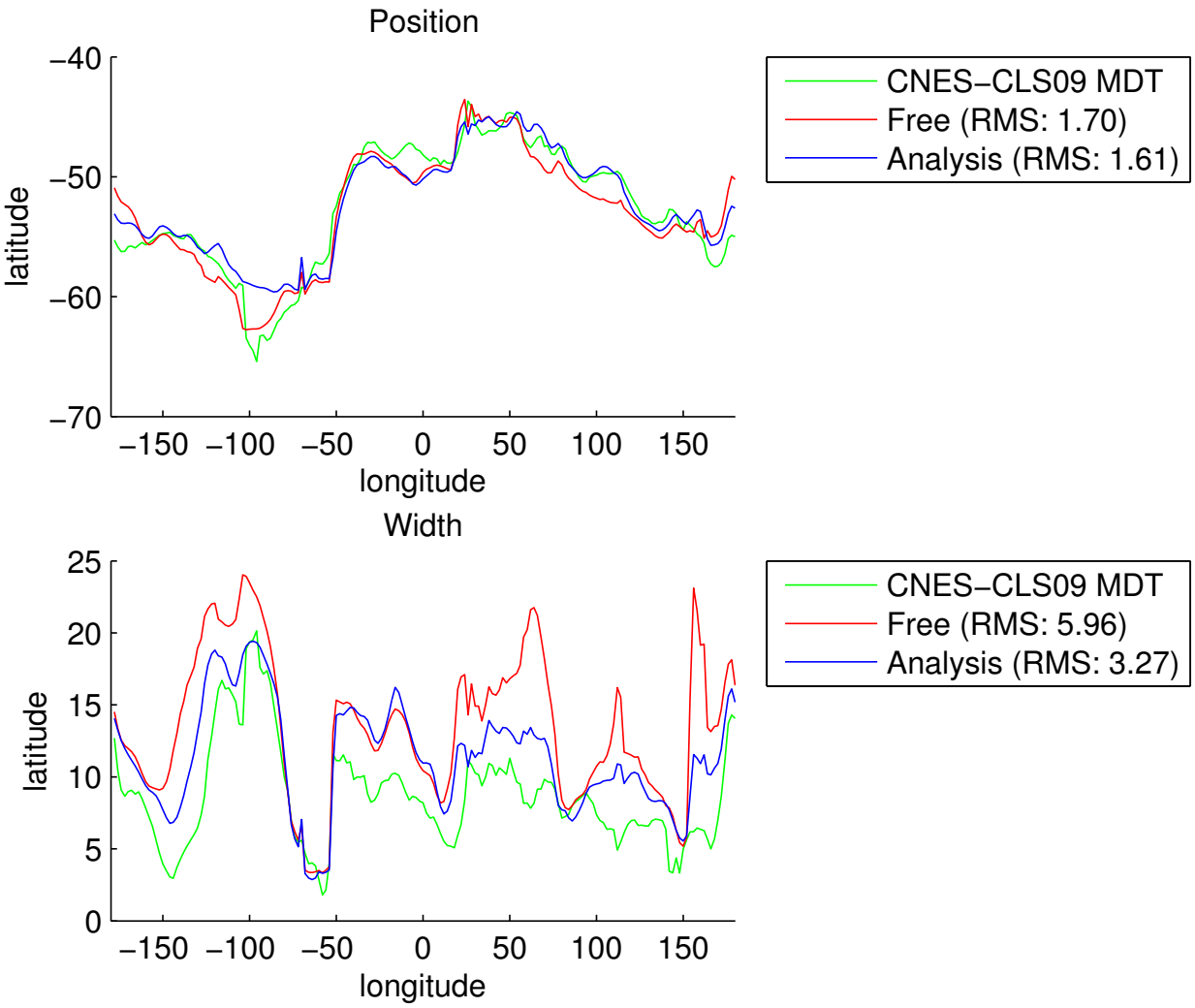

Figure 14: Position and width of the mean SSH front

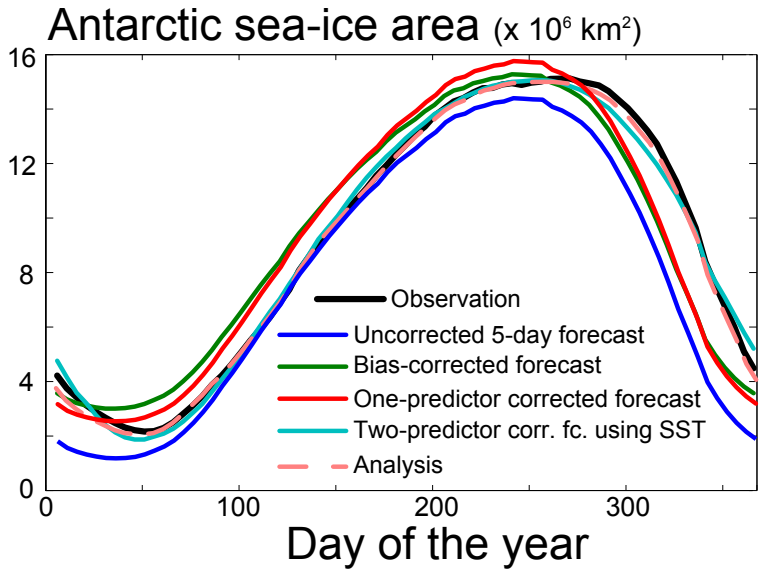

Figure 15: Antarctic sea ice area as a function of the day of the year averaged over the period 1985-2007. Shown are the observation, the reanalysis, the uncorrected 5-day forecast and different forecasts corrected with post-processing techniques.

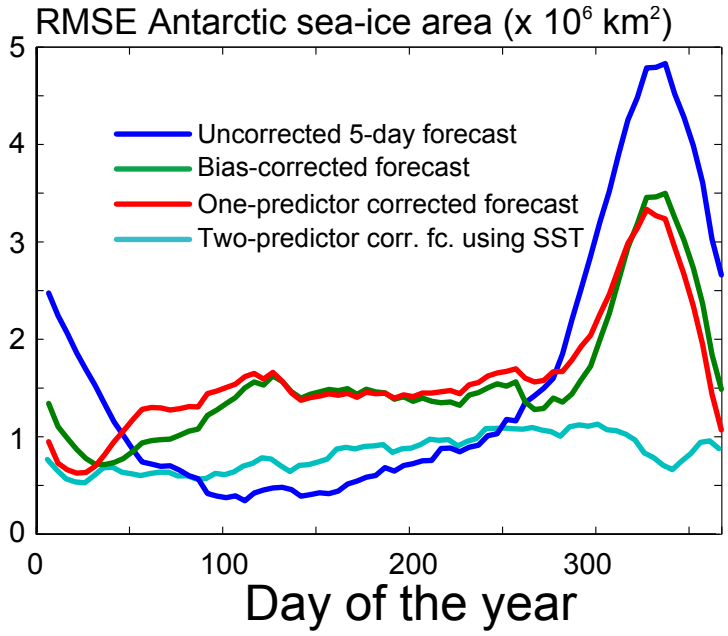

Figure 16: Root Mean Squared Error (RMSE) of the Antarctic sea ice area a function of the day of the year, averaged over the period 1985-2007, for the uncorrected 5-day forecast and different forecasts corrected with post-processing techniques.

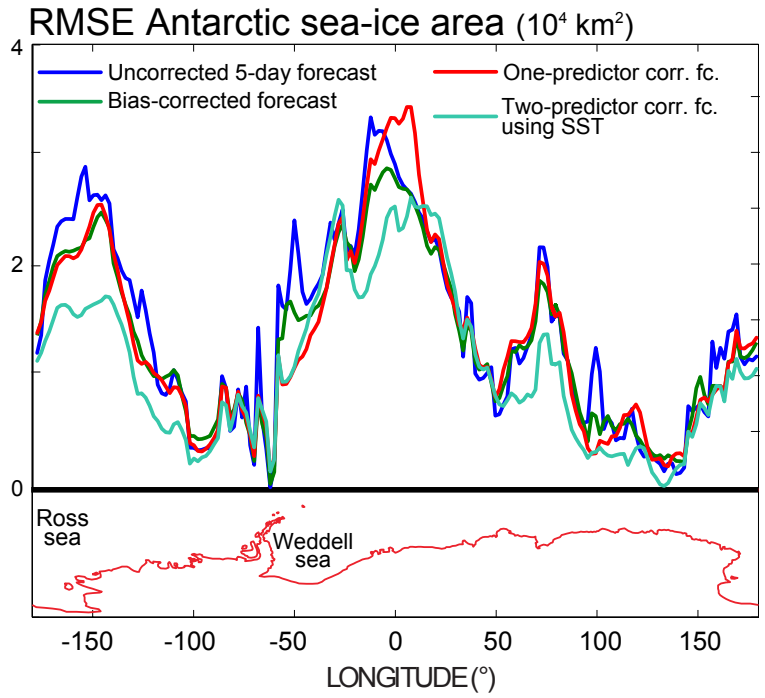

Figure 17: Root Mean Squared Error (RMSE) of the Antarctic sea ice area as a function of longitude for the uncorrected 5-day forecast and different forecasts corrected with post-processing techniques. The longitudinal spacing is 2° and the sea ice area at a certain longitude is the total sea ice area in a range of 2° east from that longitude (all south of 50°S).