ERA report series

26 Sea Surface Temperature and Sea Ice Concentration for ERA5

Shoji Hirahara, Magdalena Alonso Balmaseda, Eric de Boisseson and Hans Hersbach

Series: ERA Report Series

A full list of ECMWF Publications can be found on our web site under:

http://old.ecmwf.int/publications/

Contact: library@ecmwf.int

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

Abstract

This report explores ways to describe Sea Surface Temperature (SST) and Sea Ice Concentration (SIC) suitable for next generation reanalyses at ECMWF. To make a long-term time series, we first evaluate two high-resolution SST/SIC products: HadISST2 and OSTIA. HadISST2 is considered for the historical period, while OSTIA would support the near-real-time reanalysis.

The two products reveal a broad agreement in the global mean SST, suggesting that the combined time series should be consistent over time. Differences are mainly found in small scale variabilities, where OSTIA better resolves the tropical instability waves and sub-mesoscale eddies in the mid-latitudes. The high-resolution features allow OSTIA for a better fit to scatterometer surface wind observations in the 4D-Var data assimilation (DA), improving forecast scores for up to 3 days.

We also discuss the existence of other differences between the two products. Particularly important are the discrepancies found in both SIC and related SST variabilities in the polar regions. By taking into account those uncertainties in SIC as well as SST, we show that ensemble DA produces reliable spread and improves the quality of subsequent 4D-Var DA that uses the background error.

Finally, we propose a plan for the use of SST and SIC in the next ECMWF reanalysis, ERA5.

1 Introduction

Sea surface temperature (SST) is an essential variable in climate studies. A number of products have been developed to facilitate those applications to provide a globally complete and temporally consistent description of the ocean state in history. This report is an attempt to create a continuous SST and sea ice concentration (SIC) record for reanalyses. A specific focus is put on ERA5, the next generation atmospheric reanalysis at ECMWF.

There have been similar efforts in recent years. Fiorino (2004) created a long-term SST dataset that covered the ERA-40 (Uppala et al., 2005) production period (1957-2002). SSTs were carefully evaluated and then combined together to make a single consecutive time series. This SST was used as the lower boundary condition in the ERA-40 atmospheric reanalysis. Its successor, ERA-Interim (Dee et al. 2011), also uses this dataset for the same period. Another example is the Atmospheric Model Intercomparison Project (AMIP) phase 2. Hurrell et al. (2008) constructed a long-term SST dataset by taking SST anomalies from several products and by overlaying them on top of a common high-resolution climatology. Sea Ice Concentration (SIC) also requires a careful and dedicated work to avoid any spurious climate jumps in history (Nomura, 1995; Fiorino, 2004).

These long-term datasets, however, revealed inconsistencies resulting from combining different products. For example, merging SST products with different horizontal resolution can imprint the inconsistency on atmospheric variables. Most noticeable are oceanic small-scale eddies and related surface wind anomalies. Numerical models have experienced such resolution changes with occasional SST upgrades (Chelton, 2005; Maloney and Chelton, 2006; Masunaga et al., 2015). Moreover, recent studies have shown that long-term trends in numerical models are very sensitive to subtle differences in the prescribed SSTs (Flannaghan et al., 2014; Fueglistaler et al., 2015).

Over the past centuries, SST observations are unevenly distributed over space and time. SST products have been developed and improved independently for dedicated purposes. Consistency in climate-oriented SST products tends to be achieved by selecting a subset of instruments with a long history. On the other hand, SST products for NWP try to collect as many observations as available within a given

cut-off time in order to provide the best estimate as possible. Reanalyses such as ERA5 usually stand in between the two demands, requiring a long-term consistent record as well as a high-quality estimate updated in near-real-time, although these are hardly met by a single dataset.

This report will evaluate the latest generation of high-resolution SSTs in use at ECMWF, HadISST version 2 (Kennedy et al., 2016) and OSTIA (Donlon et al., 2012). Our aim is to find their optimal usage in ERA5, considering its production period (1979 to present) and horizontal resolution (T639; $\simeq$ 32km). As their performance is well documented by the providers, we instead check the products through intercomparisons and through a targeted atmospheric 4D-Var data assimilation (Section 2). Even more efforts are needed for representing uncertainties in ensemble reanalysis. We thus develop an SST and SIC perturbation strategy that will provide consistent ensemble realisations over time (Section 3). Finally, we summarize the results and propose a strategy for the use of SST and SIC in ERA5 (Section 4).

2 Intercomparison

2.1 Sea Surface Temperature

2.1.1 Data

This section will briefly describe HadISST2 and OSTIA, together with three other SST products used for intercomparison. Key differences are summarised in Table 1. The HadISST version 2 is a long-term SST analysis developed by the UK Met Office Hadley Centre (Kennedy et al. 2016). There are two separate products for the same version number. The "pentad" product consists of spatially complete, 5-daily mean fields on a 0.25°x 0.25°latitude/longitude grid, while the "monthly" product uses a monthly data window with the same spatial resolution. Daily fields are created by temporal interpolation from neighbouring product dates. ERA5 will be the first application at ECMWF for the pentad version, while the monthly has been used in ERA-20C (Poli et al., 2013, 2015a) and ERA-20CM (Hersbach et al., 2015). HadISST2 assimilates in-situ observations as well as two infrared radiometers: the Along Track Scanning Radiometer (ATSR) and the Advanced Very High Resolution Radiometer (AVHRR). Uncertainty in the analysis is explicitly represented by a set of ten ensemble realisations. Each member reflects possible errors in observation sampling, bias adjustments and analysis parameters. The ten ensemble members (number 69, 137, 396, 115, 1346, 1194, 1059, 400, 1169 and 1466) are stored in MARS as numbers 0,1, $\cdots$ and 9 respectively. We select number = 0 as control member as it shows a medium trend in the

Table 1: List of sea surface temperature datasets used in this study

Name	Time Period	Resolution	
		Temporal	Horizontal
HadISST2†	1961-2010*	pentad	0.25°x 0.25°
	1899-2010*	monthly	
OSTIA	2007-	daily	0.05°x 0.05°
CCISST	Sep.1991-2010	daily	0.05°x 0.05°
COBE-SST2	1850-2014	monthly	1°x 1°
ERA-Interim	1979-	monthly-daily	1°x 1°-0.05°x 0.05°

† HadISST v2.0.0.0 and v2.1.0.0 are used in this study.

* V2.1.0.0 is not available for 2008-2010 at the time of writing.

global mean SST, even though all the realisations are equally possible by design. This dataset is under very active development. There are several sub-versions available within the version 2 family. Among them, this study will use v2.0.0.0 and v2.1.0.0 (hereafter v2.0 and v2.1). The main difference between the two is sea ice, which will be described in Section 2.2.1. We prefer to use the new version v2.1 for intercomparison. As this version ends in 2007, we use v2.0 instead in the 4D-Var data assimilation (DA) experiments conducted for 2009-2010 in Section 2.3.

OSTIA is a high resolution (0.05°x 0.05°) SST operationally produced at the UK Met Office (Donlon et al., 2012). The production started in 2007 and ECMWF has been using this dataset in operations since mid-2008. As an operational system, OSTIA has experienced occasional system upgrades, such as the ocean mask change that took place in November 2011 and the improved analysis error estimate available from February 2014 (Jonah Robert-Johns, personal communication). Various types of instruments are analysed, including in-situ, geostationary satellites, and microwave imagers like the Tropical Rainfall Measuring Mission Microwave Imager (TMI) and the Advanced Microwave Scanning Radiometer-EOS (AMSR-E) which should offer all-weather observing capability at ≈25km resolution (Wentz et al., 2000). In terms of the number of observation and hence the size of sampling errors, OSTIA is considered to be the most accurate analysis used in this study.

The European Space Agency - Climate Change Initiative (ESA CCI) SST v1.1 (CCISST; Merchant et al. 2014) is a successor of the OSTIA reanalysis (Roberts-Jones et al., 2012). It inherits the analysis system from OSTIA, but also adds various improvements to facilitate climate applications. CCISST uses only a few long-life high resolution infrared sensors to create a homogeneous satellite-based climatology that spans the period 1991-2010.

To evaluate the temporal consistency between HadISST2 and OSTIA, we also compare them with COBE-SST2 (Hirahara et al., 2013), where only in-situ observations are analysed. COBE-SST2 is a monthly 1°x 1°resolution dataset that is available for 1850 onwards. Like HadISST2, it provides a homogenised record over time to support climate studies.

In ERA-Interim (Dee et al., 2011), five SST and SIC products are combined together to make a single time series from 1979 to the present. These are HadISST (Rayner et al., 2003) for 1979-November 1981, NCEP 2D-Var for December 1981-June 2001, NOAA OIv2 (Reynolds et al., 2002) for July 2001-December 2001, NCEP RTG (Thiébaux et al., 2003) for January 2002-January 2009 and OSTIA for February 2009 onwards.

The SST/SIC products described above are interpolated onto a common 1°x1°grid for intercomparison (Section 2.1 and 2.2). Area averages are conserved in this process. In the DA experiments described in Section 2.3 and 3, we use a bilinear interpolation to preserve small-scale variability in the original products.

2.1.2 Results

The recent increase in computational power has allowed for much higher resolution reanalyses than previous generations. Accordingly, horizontal resolution now seems to have become an important aspect for SST products. Figure 1 compares the temporal evolution of SST from HadISST2 (v2.1) and OSTIA at 1.5°N. The plotted period shows a gradual transition from a weak El Niño to the following La Niña in 2006-2007 (see also Fig. 2).

In OSTIA, the tropical instability waves (TIWs) are clearly visible in the eastern tropical Pacific (Fig.1). The wave period is apparently shorter than one month since two or more wave peaks pass through 120°W

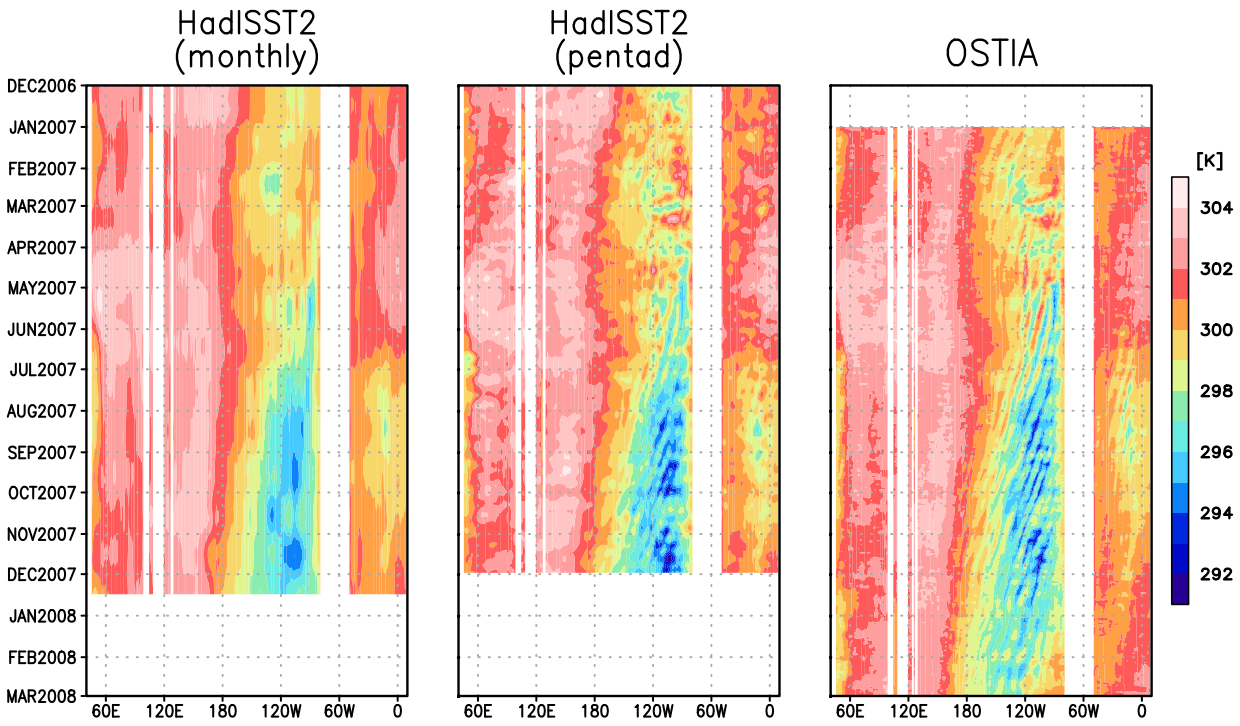

Figure 1: Longitude-time section of sea surface temperature (K) at 1.5°N for HadISST2 (v2.1) monthly (left), pentad (centre) and OSTIA (right). Data are not available for the masked periods.

within a month (e.g., October 2007). Recent observational studies denote that TIWs consist of two major physical modes that propagate at different latitudes. At the 1.5°N in the plot, it appears that 17-day Yanai waves are more likely captured rather than 33-day unstable Rossby waves (Lyman et al., 2007).

HadISST2 has a fine enough resolution (0.25°x 0.25°) to represent the 17-day mode, which has a typical wave length of about ≃1400km (Lyman et al., 2007). However, the monthly product fails to reproduce its westward propagation, simply because the data window is too long to resolve such fast moving waves. This implies that the grid resolution does not exactly match what is actually resolved (Reynolds and Chelton, 2010). In contrast, the pentad has a resolution comparable to OSTIA and reproduces the TIWs.

A number of studies have pointed out the importance of SST variability in numerical modeling. In the context of the IFS, de Boisséson et al. (2012) assessed impacts of SST on the Madden-Julian oscillation forecasts. Temporal resolution turned out to be a vital factor for its representation in the model. Poli et al. (2015b) compared ERA-20C humidity with satellite observations, suggesting that a higher temporal resolution SST would be required to improve the lack of intra-monthly atmospheric variability. For these reasons, the use of the pentad version instead of the monthly will bring a clear benefit to high-resolution reanalyses. However, also note that before the first AVHRR sensor becomes available in September 1981, even the pentad HadISST2 cannot resolve these tropical signals as it does in recent years. Atmospheric reanalysis users may need to be aware of such resolution changes due to the continuously changing observing network in history.

Interannual variability is another concern in combining multiple SST products. Figure 2 shows the monthly mean NINO.3 (120°- 90°W, 5°S-5°N) SSTs over the El Niño/La Niña period of 2006-2007 (same as Fig. 1). Overall, OSTIA agrees very well with HadISST2. Aside from the lower values in the beginning of 2007, OSTIA falls within the range of the ensemble spread in HadISST2. Furthermore, the

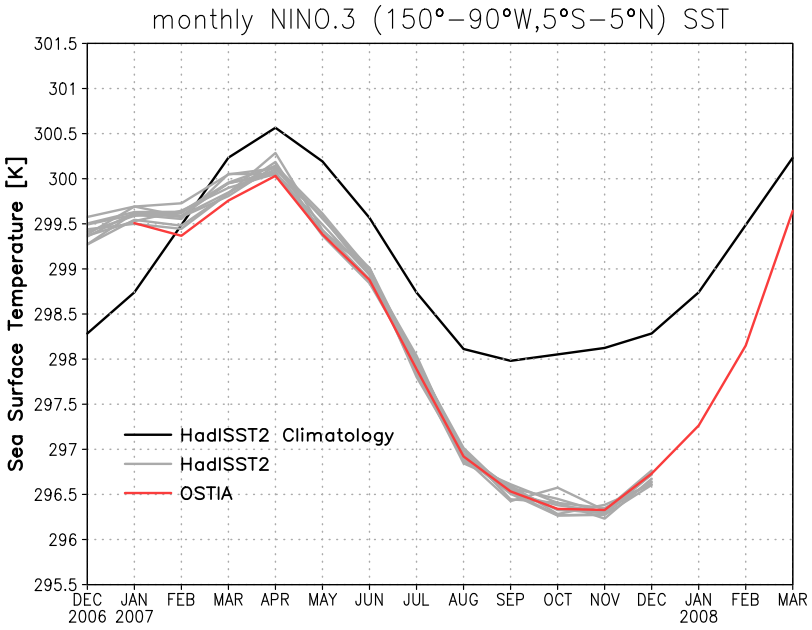

Figure 2: Monthly mean NINO.3 (150°- 90°W, 5°S-5°N) SST [K] for HadISST2 monthly (grey) and OSTIA (red). Also plotted is climatology (black) computed from HadISST2 ensemble means in 1979-2007. Data are not available for the masked months.

SST differences between the products are far smaller than the monthly mean anomalies themselves. El Niño and Southern Oscillation (ENSO) is the principal component of interannual variability in the ocean sub-surface. Its consistent representation is equally important for atmospheric modeling. In spite of the small scale differences revealed in Fig. 1, we expect that the two products would give a similar response in the large scale variability in the atmosphere.

Figure 3 compares annual mean SSTs averaged over the low-to-mid latitudes (60°S - 60°N). In higher latitudes, SIC is commonly used as a proxy SST observation to help fill up the data-sparse region. Therefore, the usage, quality and long-term trend of SIC can affect the observation-based SST record. This is separately evaluated in the next section.

HadISST2 provides a direct measurement of uncertainty with its ensemble spread (Fig. 3). It becomes smaller from 1979 onwards, suggesting its improving accuracy over time. The spread among SST products (about 0.05K), however, is larger than the HadISST2 ensemble spread. It is also important to note that the inter-product spread does not seem to decrease over time.

During the overlapping period in 2007, HadISST2 and OSTIA are consistent with each other, being colder than the other two (COBE-SST2 and CCISST). HadISST2 is also colder than the other two data sets during most part of the record. Given that HadISST2 uncertainty is estimated within a given set of observation and an analysis scheme in the product, the systematic differences to the others should be explained by other factors that are not taken into account. The exact reason is unclear, though the negative deviation of HadISST2 could be explained by the respective definitions of ”sea surface temperature”. CCISST defines SST as the 20cm depth water temperature at 9 and 21 hours local time. At that time of the day, sea water is half warmed up and gives a roughly equal temperature to the daily mean. COBE-SST2

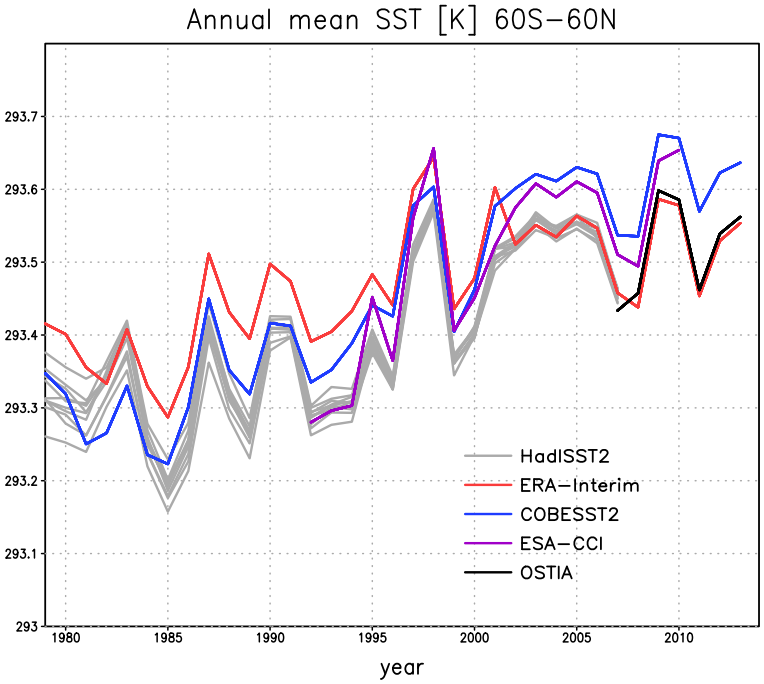

Figure 3: Annual mean 60°S-60°N average SSTs [K] from HadISST2 (grey; 10 members), ERA-Interim (red), COBE-SST2 (blue), CCISST (purple) and OSTIA (black).

uses a mixture of daytime and night-time observations (in-situ only), which are meant to provide a "bulk" SST. In contrast, HadISST2 does not include daytime observations for infrared measurements (AVHRR and ATSR), which generally outnumbers in-situ observations. This avoidance of daytime observation leads to colder SSTs by construction. On the other hand, OSTIA adopts another established way of defining SST, which is "foundation depth" temperature. The depth is defined as the uppermost layer that is free from the diurnal cycle. This roughly corresponds to 4-10m depth below the sea surface and the temperature at that level is roughly equivalent to night-time minimum temperatures at the layers above (Donlon et al., 2007). The difference between HadISST2 and OSTIA is much smaller than the inter-product spread. This could be a result of the above-mentioned physical consistency in the way of defining SST.

It is also important to evaluate the long-term trend of this semi-global SST index. Visual inspection of Figure 3 indicates that from 2000s onwards, the combined time series of HadISST2 and OSTIA exhibits similar long-term trend as in COBE-SST2 and CCISST. During the 90s, CCISST exhibits stronger warming than any of the other products. And before the 90s the warming trend is larger in COBE-SST2 than in HadISST2. Among the different products, the ERA-Interim is the one showing the weakest warming trend, although the behaviour before 2000 is consistent with HadISST2. The weak trend of ERA-Interim can be attributed to the switch from NOAA OIv2 to NCEP RTG SST as the lower boundary condition in 2001-2002. This resulted in a sudden SST cool down due to the discontinuity between the two products as shown in Simmons et al. (2014).

All the three intercomparisons showed a modest consistency between HadISST2 and OSTIA. By combining these products along the time, it seems to be possible to construct a single consistent record from 1979 onwards.

2.2 Sea Ice Concentration

2.2.1 Data

The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) provides high-quality SIC products that have been developed under the Ocean and Sea Ice Satellite Application Facilities (OSI-SAF). OSTIA uses its operational version (OSI-401-a; hereafter OSI-SAF), and CCISST and HadISST2 use the reprocessed version (OSI-409-a; Eastwood et al. 2014; hereafter, OSI-SAFr). See Table 2 for a brief summary.

For this study, we obtained the original OSI-SAFr (1978-2014) from the provider. It comes as a daily level-3 product on a 10km polar stereographic grid. We interpolate this onto a spatio-temporally complete, daily 0.25°x 0.25°latitude/longitude grid as used in HadISST2. Uncertainty in the product is estimated by the provider and this is used in our quality control procedure to reject poor quality dates. If the hemispheric mean analysis uncertainty exceeds 15%, such dates are not used in the following process. When the sea ice field is spatially incomplete, these dates are also rejected. Missing dates are then temporally interpolated linearly from the closest neighbouring dates. If no data is available within $\pm$30 days, SIC is relaxed toward the climatology in 1993-2002. Additionally, a land-spillover correction is applied following Markus and Cavalieri (2009). This mostly removes spurious coastal sea ice, which is typically observed in the Barents Sea and along the Russian coast in boreal summer. Only coastal grid points are subject to this quality control. This pre-processed SIC dataset will be evaluated in the next section.

For HadISST2, we use two sub-versions: v2.0 and v2.1. The main difference is in the SIC component, where Rayner et al. (2003) and Titchner and Rayner (2014) (hereafter RA03 and TR14) are used in v2.0 and v2.1, respectively. RA03 is based on a microwave-derived sea ice from the Goddard Space Flight Center (GSFC) and the National Centers for Environmental Prediction (NCEP). These two products are intercalibrated and a melt-pond correction is applied before combined. TR14 is also a special effort for creating a homogeneous, historical SIC record that extends back to the pre-satellite period. To achieve this goal, OSI-SAFr is bias-adjusted toward ice charts produced by the US National Sea Ice Center (NIC). Summertime melt pond bias is essentially removed from this dataset since it refers to such a visual chart analysis. The spatio-temporal resolution is monthly 1°x 1°in both products. To be used with HadISST2, the SIC fields are interpolated into the same grid as SST, i.e., daily 0.25°x 0.25°.

Table 2: List of sea ice concentration datasets.

Name	Resolution		related SST dataset
	Temporal	Horizontal	
OSI-SAF (OSI-401-a)	daily	≈10km	OSTIA
OSI-SAF reprocessing (OSI-409-a)	daily	≈10km	CCISST, HadISST v2.1.0.0*
Rayner et al. (2003);RA03	monthly	1°x 1°	HadISST v2.0.0.0
Titchner and Rayner (2014);TR14	monthly	1°x 1°	HadISST v2.1.0.0

* Bias adustements applied.

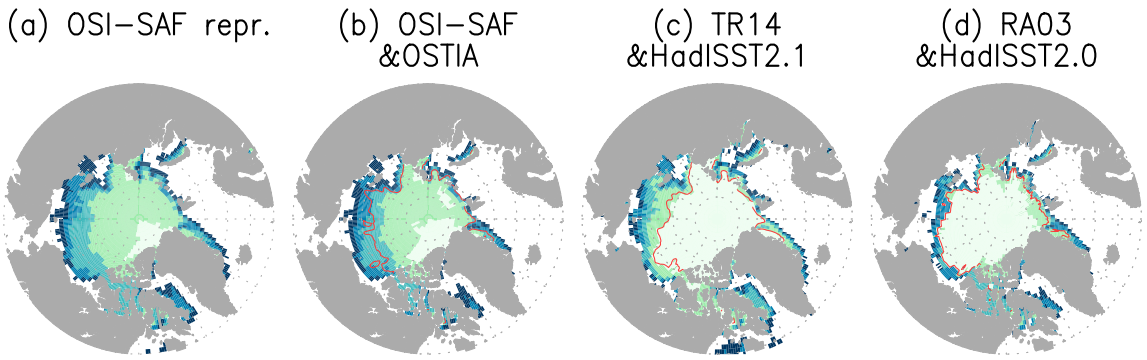

Figure 4: Sea ice concentration (%; shade) in July 2007 in (a) OSI-SAF reprocessing, (b) OSI-SAF, (c) TR14 and (d) RA03. The red contour indicates the SST of −0.5 °C in (b) OSTIA, (c) HadISST v2.1 and (d) HadISST v2.0. See Table 2 for details.

2.2.2 *Results*

First, we compare the spatial distribution of monthly mean SICs from OSI-SAFr, OSI-SAF, TR14 and RA03 for July 2007 (Fig. 4). The two OSI-SAF products (Fig.4a and b) are almost identical, but the ice edge is slightly expanded towards the shore in OSI-SAFr (e.g., in the Baffin Bay and to the east of Greenland). As noted, TR14 is derived from OSI-SAFr, where SICs are empirically adjusted to match the ice charts by NIC. The adjustment is done with a function of SIC itself in a multiplicative manner. If compared with the original OSI-SAFr, TR14 mostly preserves the ice edge locations, while bearing a steeper meridional gradient towards the pole. This results in an overall increase in sea ice area. The high SIC in TR14 partially arises from the fact that ice charts are generally less affected by summertime melt-pond bias (negative SIC bias) than passive microwave retrievals (Titchner and Rayner, 2014). In this respect, TR14 is considered to be more accurate than OSI-SAFr during summer. Much sharper gradients are found in RA03, which also tries to correct such melt-pond biases in a different method. In this particular month, we find considerably large differences between those products, locally reaching ± 40%.

Uncertainties in SIC analysis have been and will be an active research area (Andersen et al., 2007; Inoue et al., 2008; Ivanova et al., 2014; Nomura, 1995). The example above clearly suggests the existence of large structural uncertainties in the products we use, which presumably result from different SIC retrievals, the use of hand-drawn ice charts and bias adjustment processes.

Also plotted on Figure 4 is the −0.5°C SST isotherm. SIC is negatively correlated with SST, and such statistics is employed in polar grids to derive SST from SIC where direct SST observation is limited (e.g. Rayner et al. 2003). In HadISST2, the contour closely follows 90% SIC, suggesting the SST in the region is strongly constrained by sea ice. However, OSTIA allows relatively free SST variation around the same SIC (Roberts-Jones et al., 2012), indicating that the assumed SIC-SST relationship and the observation usage at ice-covered regions are substantially different.

Figure 5 shows long-term trends of Arctic sea ice extent (SIE) from the four sea ice products. The Arctic SIE takes its annual maximum (minimum) in March (September) in climatology. September has a relatively stronger downward trend. In the plot the datasets show a good agreement in interannual fluctuations but with constant differences over time. In the low SIE (SIC≥10%), a large spread is found in September especially before 2005. This is consistent with the fact that September is more prone to melt-pond biases than March.

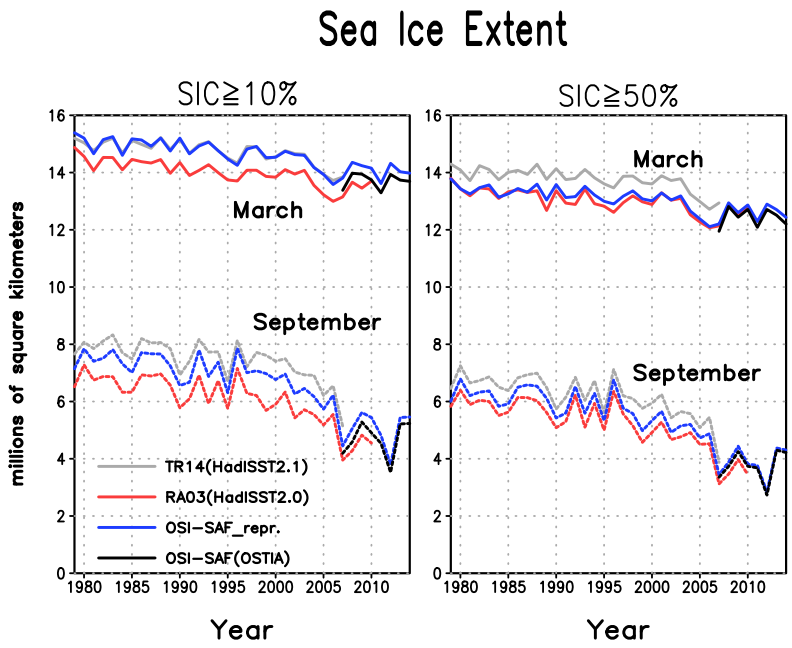

Figure 5: Arctic sea ice extent ($\times 10^6 km^2$) in March and in September from TR14 (grey), RA03 (red), OSI-SAFr (blue) and OSI-SAF (black). The extent is defined as the oceanic area with 10% (50%) or more SIC in the left (right) panel.

Discrepancies are less evident in a higher SIE (SIC$\geq$50%) but still exist among the datasets. The positive deviations of TR14 from OSI-SAFr in March are rather noticeable. As melt ponds are unlikely to form up in this month (e.g., Rösel et al. 2012), the difference implies that melt ponds are not the only source of uncertainty that we need to consider.

We can estimate possible atmospheric response to different sea ice prescriptions. Two IFS 4D-Var experiments (not shown) are run with HadISST v2.0 and v2.1, showing $\simeq -1$K response in two-meter temperature to appear locally for $+5$% SIC difference. This result emphasises the importance of using a historically consistent SIC product in reanalyses. Due to the relatively good agreement between the OSI-SAF products and the technical difficulties to obtain real-time updates of TR14, we propose the original OSI-SAFr to be used for ERA5, followed by OSI-SAF. As the names suggest, these products are largely consistent in terms of the analysis system and spatiotemporal resolution. If used together with HadISST2.1, spatial co-variability with SST is preserved at the first order, as TR14 originates from OSI-SAFr. In Section 3, we revisit structural uncertainties in SIC products to explore a way to consider them in EDA.

2.3 Deterministic 4D-Var experiment

2.3.1 Experiment setup

SST products are evaluated through atmospheric Data Assimilation (DA) experiments. The IFS Cycle 40r3 is run at T511L91 ($\approx$40km, 91 vertical levels), a deterministic 4D-Var configuration with three incremental trajectory updates (T95-T159-T511). The same set of atmospheric observations is used as in operations, but two different SSTs, HadISST v2.0.0.0 (EXPHADISST2) and OSTIA (EXPOSTIA), are given as the lower boundary condition. The DA cycles start from the identical initial condition and

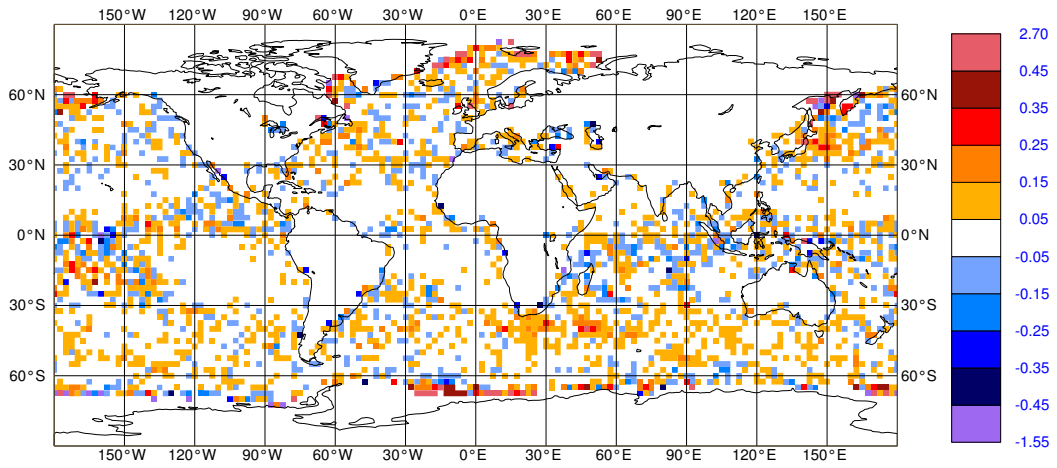

Figure 6: Standard deviation of first guess departures of 10-m zonal wind with respect to the Advanced Scatterometer (ASCAT), shown as a difference between EXPHADISST2 and EXPOSTIA [m/s]. The difference is positive (negative) when EXPOSTIA better (worse) fits to ASCAT. The statistics period is 22 January - 10 March in 2010.

after seven days of spin-up, the results are compared for August 2009 and February 2010.

2.3.2 Results

Atmospheric DA is useful in assessing the quality of SST products. SST is allowed to force the atmosphere, while DA keeps the rest of the atmosphere close to the observed state. In our experiment setting, the difference between EXPHADISST2 and EXPOSTIA should highlight such impacts from given SSTs. Figure 6 shows the difference (EXPHADISST2-EXPOSTIA) of the standard deviation of first guess departures of 10-m zonal wind with respect to the Advanced Scatterometer (ASCAT). A smaller deviation means that the forecast agrees more with the independent satellite observation in the next analysis cycle. This index can be considered as an indirect measure of the quality of underlying SST.

Differences tend to appear in cloudy areas such as the Inter-Tropical Convergence Zone (ITCZ). In contrast, subtropical oceans show virtually no difference. As both products use in-situ and infrared measuarements in the same way, this difference is probably a result of the additional use of microwave sensors in OSTIA, which are able to observe sea surface condition through clouds. In the South Pacific Convergence Zone for example, OSTIA manages to show a better fit to ASCAT while the benefit is little in subtropical cloud-free regions. One must be careful for the positive impacts around the ice edge in OSTIA because this is due to different prescriptions of sea ice. The statistics in those regions relies much on small samples as scatterometers do not observe ice-covered areas.

Overall, the Figure 6 is covered by positive values, which corresponds to smaller forecast departures in EXPOSTIA than in EXPHADISST2. Positive impacts from the use of OSTIA are dominant in the southern hemisphere (SH) mid-latitudes (30°-60°S). Clear improvements are also found along subtropical oceanic fronts, such as the Kuroshio Extension and the Agulhas Current. These regions are characterised by strong oceanic eddy activity and atmospheric westerly jets aloft.

It is possible that the results can change depending on models and 4D-Var configurations. To further

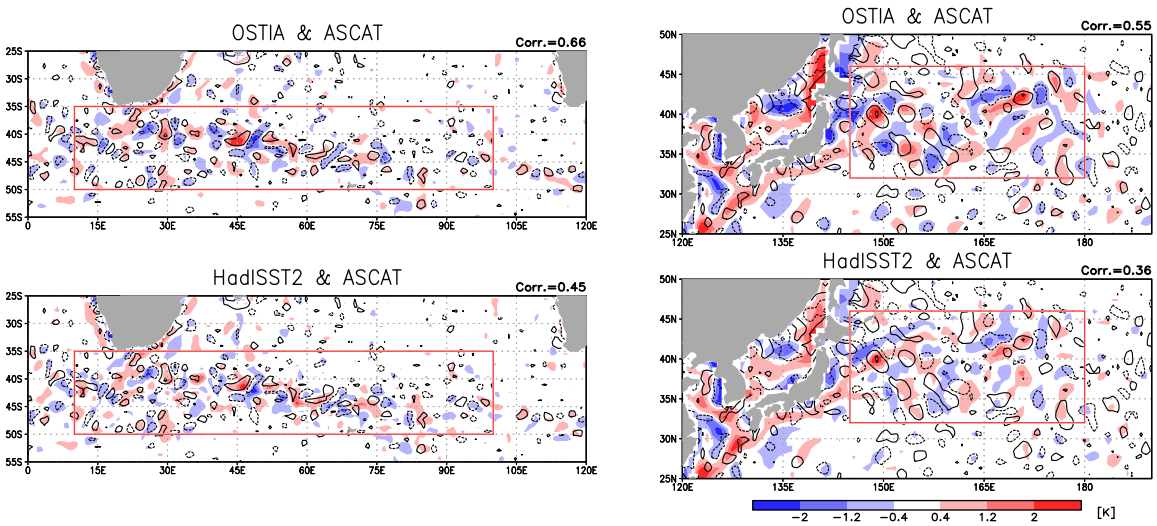

Figure 7: Spatially high-pass-filtered ASCAT wind speed (contour) and SST (shade) from OSTIA (top) and HadISST2 (bottom). The dotted(solid) line indicates negative(positive) wind anomalies at −1.25, −0.75 and −0.25 (0.25, 0.75 and 1.25) [m/s]. Shown on the top right of each panel is the anomaly correlation coefficient between wind speed and SST anomalies in the boxed region.

confirm the result, we also compare the SSTs directly with ASCAT without using the model (Fig. 7). Our focus is on the two mid-latitude ocean fronts mentioned above. In order to extract high-resolution features, we apply a LOESS high-pass filter with a half-power radius of 30°in longitude by 10°in latitude (Schlax and Chelton, 1992; Chelton et al., 2004). The software is obtained from Netlib (`http://www.netlib.org/`) and is slightly adapted to allow the rectangular filtering domain used in this study. ASCAT level-2 winds are first binned at a T255 reduced gaussian grid at each hour and only high-frequency components are extracted and averaged over time. The same procedure is applied to the gridded SST products.

In the Agulhas current region, both SSTs successfully represent sub-mesoscale eddies represented as a train of positive and negative anomalies (Fig.7). The surface wind also exibits anomalies that geographically collocate with SST anomalies. OSTIA, however, has a slight more positive correlation with ASCAT than HadISST2 over the eddy-active area. A positive correlation appears also along the Kuroshio Extension along 35°N. Warm ocean destabilises the atmospheric boundary layer, enhancing the vertical mixing and pulling strong upper wind down to the surface. On the other hand, cold surface stabilises the boundary layer and decouples the surface from the upper layers. This results in a strong positive correlation between wind and SST (Chelton and Xie 2010 and references therein). As the 4D-Var DA experiments have a fine enough resolution (≃40km) to respond to those oceanic eddies, it is concluded that the improved first guess fits in OSTIA are brought through atmospheric dynamical response to the high-resolution features better represented in the product.

Figure 8 shows anomaly correlation of analysed and forecasted temperature in the upper and the lower troposphere. The difference is defined as EXPOSTIA−EXPHADISST2 so that a positive (negative) value corresponds to a better(worse) score for EXPOSTIA. Note that the difference in correlation is normalised. At short lead times up to three days, scores are improved in most of the regions, where a particularly clear signal is found in the Southern Hemisphere (20°S-90°S) lower troposphere. This result is consistent with the better surface wind representation in the initial field (Fig. 6). However, the scores appear to be worse in the later forecast days, particularly in the upper troposphere in tropics (Day 6 and later). It turned out that the HadISST2 control member used for this experiment is slightly warmer than

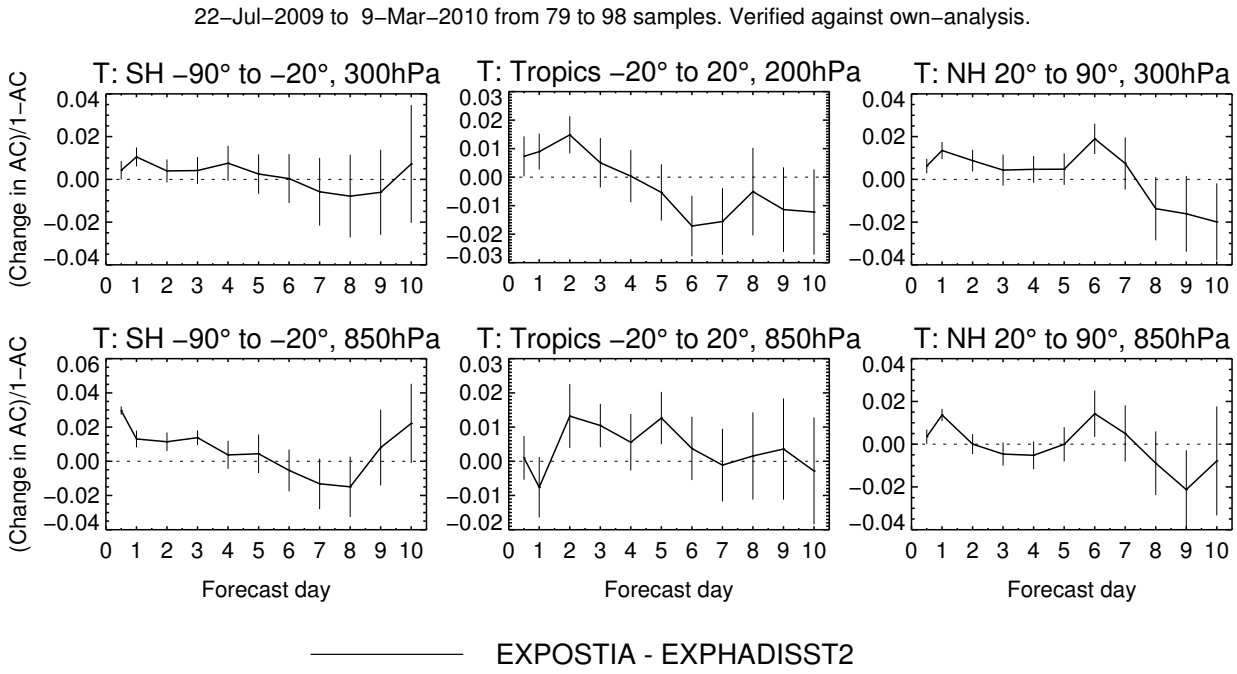

Figure 8: Anomaly correlation between analysed and forecasted temperature for the Southern Hemisphere (90°S-20°S; left), tropics (20°S-20°N; centre) and for the Northern Hemisphere (20°N-90°N; right). The forecast scores are represented as normalised difference of EXPOSTIA with respect to EXPHADISST2 (see legend in y-axis). Top and bottom panels show the upper (300hPa for the NH and the SH and 200hPa for tropics) and the lower (850hPa) troposphere respectively. The score is the sum of summer (22 July - 10 September 2009) and winter (22 January - 10 March 2010) experiments.

its ensemble average during the summer and winter experiment periods on average (not shown). As this model cycle has a cold bias in the upper troposphere, warm SST is considered to cancel the bias through the moist adiabatic adjustment, warming up the upper troposphere and resulting in better scores at those time steps. In fact, another experiment confirms that the scores can be simply improved by artificially raised OSTIA SST in the tropics by the same amount (not shown). Also note that HadISST2 is a probabilistic product by design. Given that the warm control member is just one of the possible realisations, we must be careful not to over-interpret these deterministic experiments.

It seems that spatial/temporal resolution is the main source of differences between HadISST2 and OSTIA in the weather time scale. The benefit of the high-resolution SST is estimated to be 2-3 percent and it lasts for up to three days. Although not as sharply as OSTIA, HadISST2 is able to represent high resolution features such as SST/wind relationships. The key advantage of HadISST2 is its homogenised time series, which extends back to the past. It also allows for a real-time extension with the OSTIA in a consistent manner. In addition, the ensemble representation of uncertainty in HadISST2 is a unique feature that is not available in OSTIA. This should in principle benefit EDA, as discussed in the next section.

3 Representation of uncertainty

3.1 Method

The previous section revealed various differences between SST and SIC products that we will use. Some uncertainties can appear randomly and some systematically. As these will eventually propagate into the atmosphere through DA cycles, it is important to characterise and quantify them in the EDA. The aim of this section is to design a perturbation method that represents the perceived uncertainty in SST and SIC products and to quantify its impact on analyses and forecasts.

The use of SST perturbations was first introduced to IFS at Cycle 36r2 in the context of seasonal forecasting (Vialard et al., 2005). The same methodology was later extended to the EDA. This scheme randomly picks out a set of pre-calculated error patterns and then combines these to create a perturbed field of the day. By repeating the procedure, the scheme is able to produce an arbitrary number of ensemble realisations. Although the method is generic enough, the error patterns used in the scheme are now considered to be obsolete and probably over-represent uncertainties for recent SST products. Given the relatively small size of structural uncertainties seen in the previous section as well as the ensemble information in HadISST2, we decided to revise the perturbation scheme as described below.

We denote $\varepsilon(d)$ as a single realisation of the estimated error valid for a date d. This is, in general, a vector that represents the uncertainty of all the different variables in consideration. In our case ε is a two-dimensional error comprising SST and SIC. Let us assume now that we have N independent sources of uncertainty, and therefore the total error $\varepsilon(d)$ can be expressed in an additive form as:

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

where $\varepsilon^{i}(d)$ is an error realisation taken randomly from the i-th category of uncertainty at a date d. We will discuss the different uncertainty categories below, but before that it is pertinent to describe the temporal aspects of the perturbations.

3.2 Temporal treatment: interpolation and correlation

Error patterns are sampled at various time intervals (monthly or pentad for example) depending on the temporal resolution of original products. On the other hand, an application may require a different temporal discretization. In our particular case, we would like to generate daily perturbations from 5-daily and monthly products (see next section). Here, we perform a temporal interpolation as follows.

We randomly draw a set of errors from the two adjacent dates, d_- and d_+, and estimate a weighted interpolation to the date d as:

$$\varepsilon^i(d) = \frac{1-w}{\sqrt{w^2+(1-w)^2}}\varepsilon^i(d_-) + \frac{w}{\sqrt{w^2+(1-w)^2}}\varepsilon^i(d_+). \quad (2)$$

In case of monthly products, d_- and d_+ represent indices for the previous and the next calendar month. The values are assumed to represent the 15th of each month. For pentad mean products, these are defined as the third day in five. We use a temporal weight, $w = (d-d_-)/(d_+-d_-)$, as typically used in the bilinear interpolation, although this is inflated by $1/\sqrt{w^2+(1-w^2)}$. The inflation factor is introduced to homogenise the error spread over time. Random errors tend to cancel out when combined. If the two errors are statistically independent and have the same size of uncertainty, σ, such a bilinear interpolation results in an artificially reduced spread $\sqrt{w^2+(1-w^2)}\sigma$ in those interpolated dates.

This procedure also allows preserving the temporal correlation of the original sampling in our daily perturbation record.

3.3 Sources of uncertainty

The advantage of sampling realisations from existing datasets is that it provides a natural way of specifying the spatial and temporal correlation of the perturbations. We will consider two types of uncertainty: analysis uncertainty/error (AE) and structural uncertainty/error (SE). AE refers to the estimation of error within a single analysis system. SE are not sampled by AE, but still exists and results from structural differences between analysis systems. To represent correlation between SST and SIC, we also introduce the concept of multivariate uncertainty, which is the uncertainty in one variable induced by its correlation with the uncertainty in other variables.

The error realisations, ε, are empirically derived from existing datasets as described below. These pre-computed error patterns are randomly picked out and combined with Eq. (1) and (2). Table 3 describes a specific configuration designed for HadISST2 and OSTIA, the SSTs used in ERA5.

Table 3: Definition of perturbation. For HadISST2, the ensemble mean and individual members are represented with brackets (<>) and subscript($_i$), respectively. The fourth column denotes the temporal sampling interval of each error field. Climatological differences are subtracted beforehand. The statistics period is 1997-2006.

Variable	Perturbation	Definition	Sampling interval
SST	Analaysis error (AE)	HadISST2.1$_i$ - <HadISST2.1>	pentad
	Structural error (SE)	CCISST - <HadISST2.1>	monthly
	Ice Analysis error (IE)	<HadISST2.0>-<HadISST2.1>	monthly
SIC	Ice Analysis error (IE)	<HadISIC2.0>-<HadISIC2.1>	monthly

For ERA5, AE is sampled using the ten different realisations of HadISST2 ensemble available at every pentad. As we are interested in applying this perturbation to OSTIA, we take errors from a recent 10-year period (1997-2006). A set of ten perturbations is created per each pentad during this period, calculated as the difference between each ensemble member and the ensemble mean. Now, we group all the realisations into calendar pentads (73 pentad bins in a year). With the ten years and the ten ensemble members available in the sampling period, we can populate each of the pentad bins with 100 realisations to choose from. Moreover, the sign ($+$ or $-$) of the field is also perturbed. This doubles the number of samples and makes the probability distribution centred around zero. For any given date d, we identify the pentad bin (d_-, d_+) that we should choose samples from. Although this method does not produce flow dependent errors, it allows climatological seasonal cycle of errors. Fig. 9b shows the standard deviations of the SST AE generated with this method for the pentad corresponding to the 1st-5th of March. Large spread tends to appear in the Eastern Tropical Pacific and in the mid-latitude oceanic frontal zones. The magnitude and location of this pattern are consistent with the estimated error given by the OSTIA as shown in Fig. 9a. (OSTIA provides its own analysis uncertainty as the standard deviation). Rescaling of the AE spread to match the actual OSTIA error might be beneficial, but this is not tried to avoid extreme amplification. This is to avoid extreme amplification of erros that can happen when our AE has very little or no spread. For example, OSTIA often estimates large uncertainties within the ice-packed regions where HadISST2 usually gives far less spread. In such cases, rescaling may end up in an unnaturally large perturbation by dividing with a small number. Nevertheless, we expect that the AE is a good approximation of uncertainties in OSTIA. The previous section observed a gradual decrease of HadISST2 analysis uncertainty over time (Fig. 3). In the years when we create the error patterns (1997-2006), the spatial pattern and the size of ensemble spread show a close resemblance to the one estimated in OSTIA (Fig. 9a and b).

For the structural error (SE), we use differences between CCISST and ensemble means of HadISST2. We use monthly means to reduce the influence of random uncertainties, which is already taken into account in the AE. CCISST is used as an alternative to OSTIA because it adopts a similar analysis system (see Section 2.1.1) and provides a long overlapping period with HadISST2 to take statistics for. The ensemble average of HadISST2 is performed before taking a difference. This is meant to further reduce random uncertainties and help extract systematic components. All errors are defined as anomalies with respect to their own climatology so that the mean differences among the products (Fig. 3) are removed beforehand.

For SIC, we use the differences between RA03 and TR14. RA03 is used in HadISSTv2.0 and TR14 is used in HaISSTv2.1, respectively. In the products, SST is negatively correlated with SIC. To consider the co-variability of SST and SIC, we also create SST patterns from the two HadISST2s (v2.0 and v2.1) for the same year of sampling and perturb the pair of variables altogether. Figure 9c shows the SST spread resulting from the additive contribution from all sources of uncertainty. In the tropics to mid-latitudes, the SE increases the spread over broad areas. Although slightly smaller in amplitude than AE, it tends to represent large-scale errors probably because of the monthly mean sampling. In the polar regions, SSTs are further perturbed with the SIC uncertainty.

Figure 10 shows an example of SST and SIC perturbation generated from the three types of uncertainty. The SIC anomalies appear around the ice edge and bring negatively correlated SST anomalies as expected. The spread of the global mean SST is about 0.02K in this example, though it locally exceeds 1K around eddy-active regions (Fig. 9c).

Changes from Vialard et al. (2005) include (1) consistent error patterns with ERA5 deterministic SST and SIC, (2) homogenised error spread over time with Eq. (2), and (3) the introduction of SIC uncertainty. Next section will investigate possible impacts on EDA.

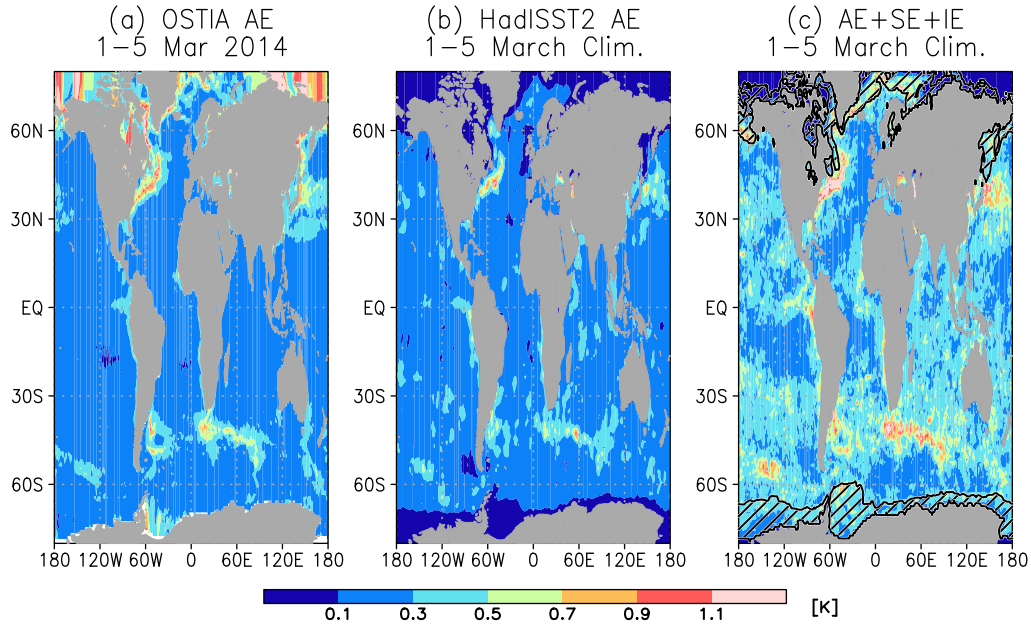

Figure 9: (a) Estimated analysis uncertainty (provided as the standard deviation) in OSTIA for 1-5 March 2014, (b) ten member ensemble spread of SST perturbation based on climatological analysis error computed from HadISST2 (AE; see Table 3), and (c) ten member SST and SIC spread based on the AE and structural errors (SE and IE in the Talble 3). The shades indicate SST uncertainty and the hatch indicates the SIC uncertainty is equal or larger than 1%.

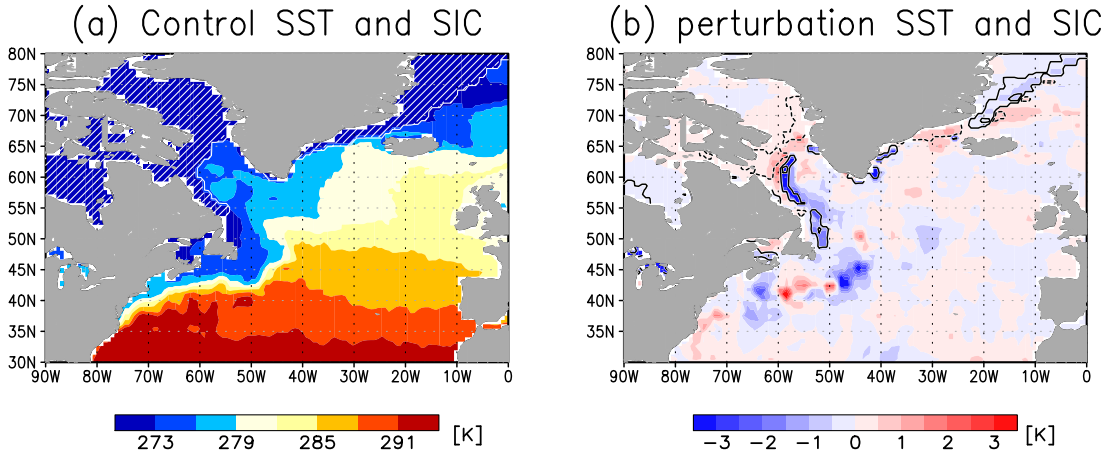

Figure 10: An example of SST and SIC perturbation. (a) SST (shaded) and ice-covered area (SIC≥1%; hatched). (b) SST (shaded) and SIC (contour) perturbation. The contours show SICs at -15, -5, 5 and 15% and solid/dotted lines indicate positive/negative.

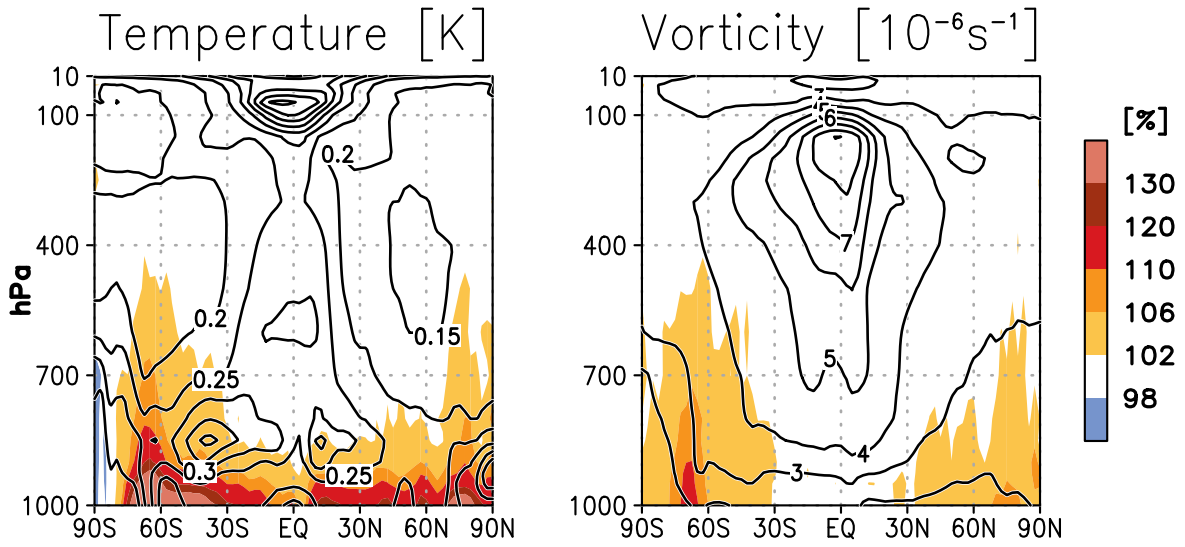

Figure 11: Temperature (left) and vorticity (right) ensemble analysis spread in EXPV2PERT (contour) and its ratio to EXPZEROPERT (shade). The spread is zonally and temporally averaged over 20 February - 10 May in 2010.

3.4 Experiment setup

Ensemble Data Assimilation (EDA) experiments are run with Cycle 41r1 at T255L137 resolution (≈80km, 137 vertical levels) with two trajectory updates (T159-T255). The ten member EDAs provide background error information to be fed into high-resolution deterministic 4DVar DAs (T511L137). Two different SST/SIC perturbation strategies are tested and compared. Ensemble members are perturbed either with the IFS's existing method (EXPV1PERT; Vialard et al. 2005) or with the new method (EXPV2PERT) developed in the previous section. As a reference, we also show EXPZEROPERT, where no SST/SIC perturbation is added to any of the members. The other aspects, such as model physics and atmospheric observations, are perturbed as in operations. Due to the limited computational resource, Jb wavelets are not updated.

The control SST and SIC are given by OSTIA and OSI-SAF. These are not perturbed either in the EDA nor in the subsequent deterministic 4D-Var DAs. With this experiment design, the performance of 4D-Var DAs can only change through error variance updates from the EDAs, which in turn are due to different SST/SIC perturbation methods.

3.5 Impacts

In applying the new SST and SIC perturbation scheme, it is interesting to know how it affects the atmospheric variables. Figure 11 shows the zonal mean spread of temperature and vorticity in EXPV2PERT. The EDA experiments perturb observations as well as model physics. Additional spread from the lower boundary condition can be assessed by comparing the results with EXPZEROPERT, where SST and SIC are not perturbed.

We find clear impacts in the lower tropospheric temperature. The EDA spread increases by 30% or more near the surface, although impacts are relatively small above 700hPa. This suggests that the uncertainty created by the surface boundary condition does not propagate into the upper troposphere, either because of physical reasons or the presence of other observations that strongly constrains the DA solution.

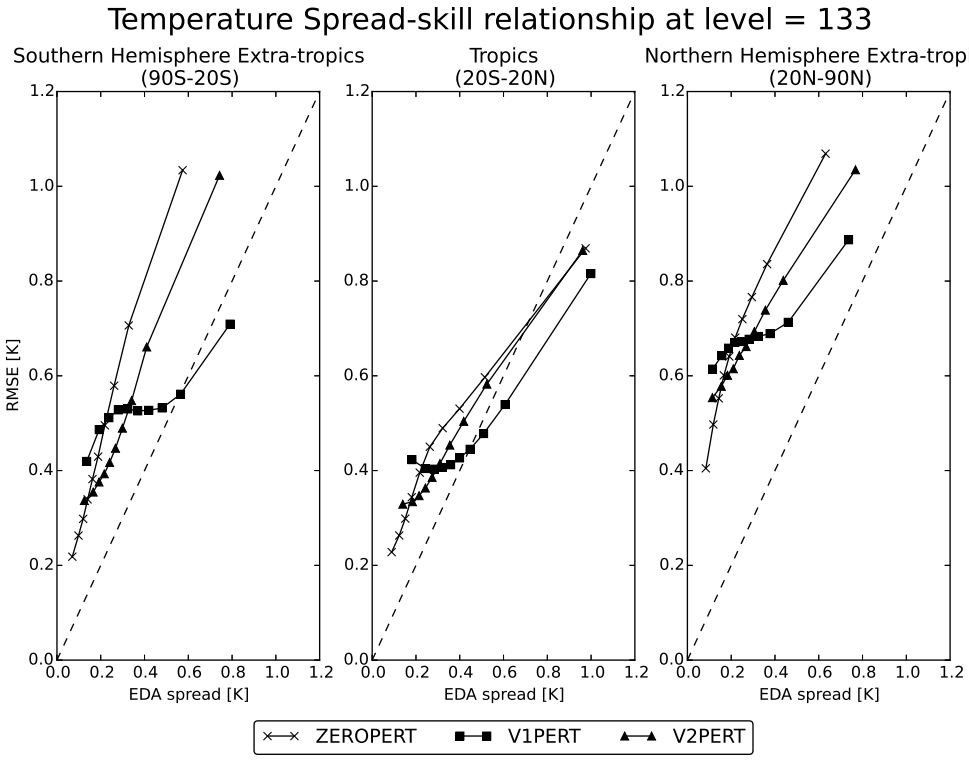

Figure 12: Spread-skill relationship of three-hour temeperature forecasts at the model level 133 (≈1000hPa) in EXPZEROPERT (cross), EXPV1PERT (square) and EXPV2PERT (circle). The abscissa indicates the ensemble forecast spread and the ordinate is the Root Mean Square Error (RMSE) of the ensemble mean forecast with respect to a high resolution 4D-Var analysis. Samples are collected geographically for the Southern Hemisphere(left), Tropics(centre) and the Northern Hemisphere(right) following Bonavita et al. (2012) and then averaged from the 20th of February 2010 to the 10th of May 2010.

The mid-to-high latitudes exhibit stronger response than the tropics. Also, the Southern Hemisphere shows relatively large response because the SST uncertainties are larger and more ocean grid points are perturbed than the Northern Hemisphere (Fig. 9c). Latitudinally confined peaks are found around 80°-70°S and 70°- 80°N. Both of them collocate with the sea ice edge in each of the hemispheres, suggesting a close link with the SIC perturbation. Additional diagnostics confirmed that these peaks mostly disappear when SIC is not perturbed (not shown). Another feature in the ice edge latitudes is the strong vorticity response. Providing that our SIC perturbation changes the surface condition, this can be considered as a dynamical response of the atmosphere since vorticity is generated when uniform wind blows over the surface with roughness/stability gradient.

The increased spread works to lower the weight of ice-affected first guess fields in DA, putting more weights on atmospheric observations when available. Note, however, that these results depend on the configuration of DA. If a longer data window is used, this would allow the atmosphere to carry the uncertainty further upward. A sparse atmospheric observation network would not be able to constrain the upper levels. This is typically the case for century-long reanalyses like ERA-20C. In this respect, impacts of SST/SIC perturbation may require a separate assessment for each DA system.

Figure 12 shows the spread-skill relationship for the surface temperature. The 3-hour forecast spread is plotted against its ensemble mean forecast error, which is verified against an independent high-resolution analysis (Bonavita et al. 2012; here T511L91).

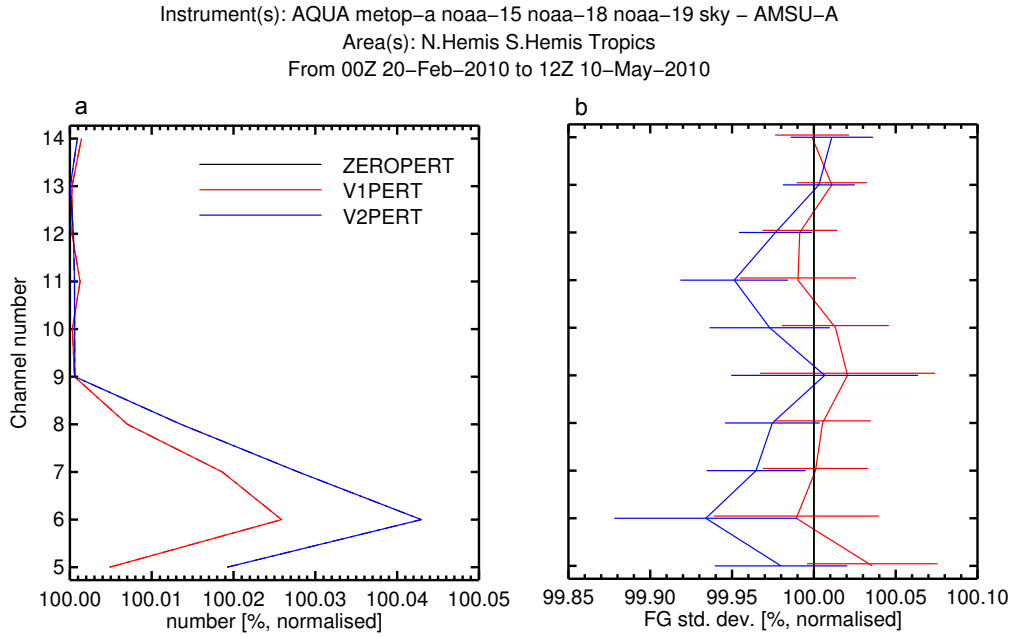

Figure 13: (a) Number of AMSU-A observation used in EXPV2PERT (blue) and EXPV1PERT (red), expressed as a ratio to EXPZEROPERT [%]. (b) same as (a) but for standard deviation of first guess departure. The statistics are based on 00UTC and 12UTC from 20th February 2010 to 10th May 2010.

Ensemble spread should serve as a measure of forecast uncertainty. Ideally, it needs to have a positive correlation with forecast skill so that they lie along the diagonal line in the scatter plot. In EXPZEROPERT, EDA forecast spread (abscissa) is generally smaller than the estimated error by 0.3 - 0.4K in the NH and the SH on average. EXPV2PERT improves this under-spread bias and help create a more correlated spread-skill relationship. Conversely, the EXPV1PERT does not distribute along the diagonal axis, indicating that it struggles to provide meaningful spread in forecasts. We recall here that EXPV1PERT are derived from differences between NCEP 2DVar and NOAA OIv2. These are irrelevant with OSTIA, the control SST used in this study. From this comparison, consistency in SST and the additional consideration for SIC uncertainty are considered important in providing reliable uncertainty.

The EDA spread is then used to update the background error in deterministic DAs (Bonavita et al., 2012). Figure 13 shows the observation feedback from the Advanced Microwave Sounding Unit-A (AMSU-A), which measures global temperature profiles. Compared with EXPZEROPERT, both of the perturbed EDAs accept more observations in tropospheric channels (channel 5-7; from the surface up to $\simeq$ 300hPa), where EXPV2PERT shows a slight more increase than EXPV1PERT. We also find significantly better first guess fits in most of the channels in EXPV2PERT. Improvements are evident in the lower troposphere as expected. The exception is the middle stratosphere (channel 11; $\simeq$30-20hPa). The reason for this is not clearly understood at the time of writing, though the impacts on stratospheric channels are also found in a different model cycle with a different DA configuration. This is interesting because it may be implying the existence of a physical mechanism, such as gravity waves, that brings this fast and remote impact.

Forecasts can be improved with the updated initial conditions. Figure 14 shows zonal mean, Root Mean Square Error (RMSE) of temperature forecasts. Differences are relatively small and confined to lower tropospheric levels at day 1. Later, the near-surface signal expands and spreads towards upper levels with time. At day 5, EXPV2PERT reduces RMSE in most of the troposphere, indicating an overall improvement over EXPZEROPERT. Comparably, EXPV2PERT shows more positive impacts than EXPV1PERT

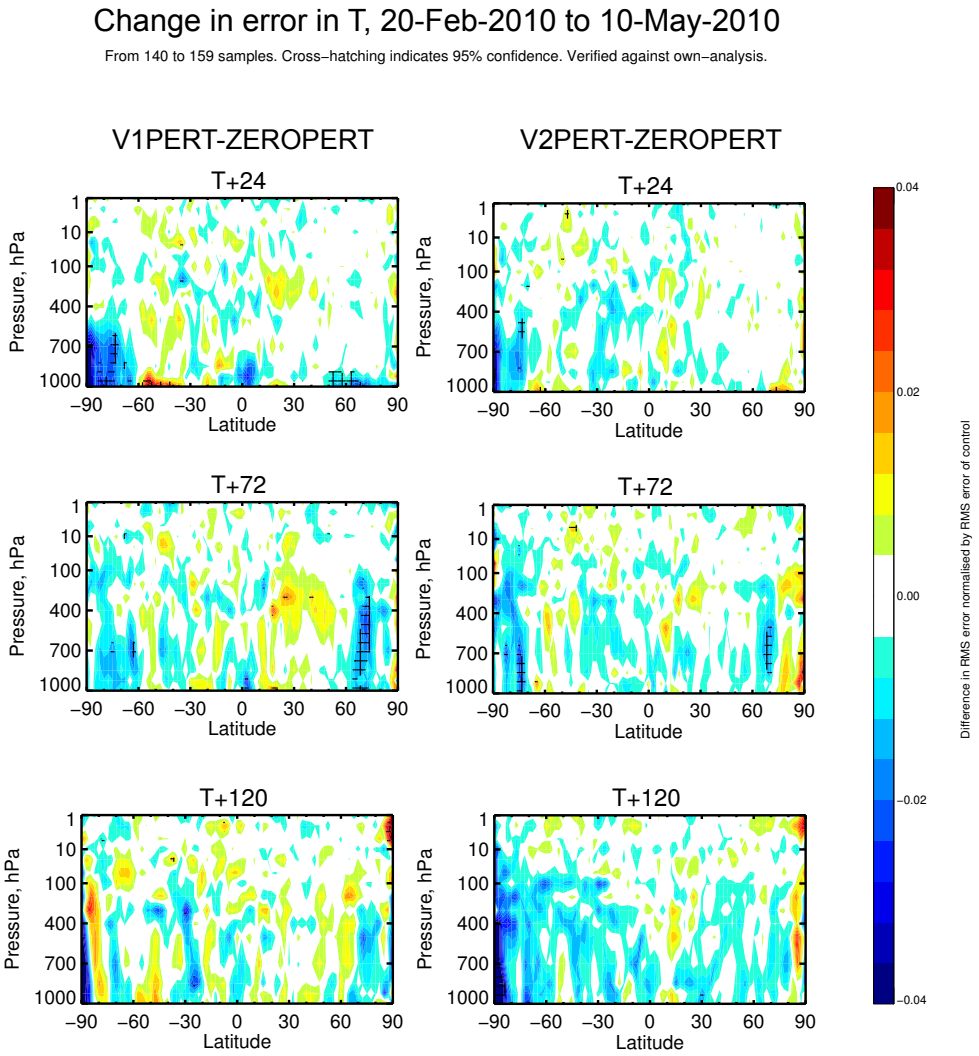

Figure 14: Zonal mean Root Mean Square Error (RMSE) of temperature forecasts for EXPV1PERT (left) and EXPV2PERT (right) at day 1 (top), day 3 (middle) and day 5 (bottom). The values are represented as a normalised difference to EXPZEROPERT, and are hatched if they are statistically significant with 95% confidence level. The forecasts are verified against own analysis.

at later time steps. Regionally averaged scores (not shown) are significantly improved in the lower troposphere for the first 3 days, and in the upper troposphere at day 4-5. The lagged improvement, from the low to the upper troposphere, could be explained by the time scale of synoptic scale disturbances. The SST/SIC perturbations can only change the background error used in the DA and affect the quality of analysis in the lower troposphere (Fig. 11). A low-level disturbance, if there is, takes a few days to grow and to carry the information towards the upper troposphere.

4 Summary and Proposal

We assessed consistency between HadISST2 and OSTIA in order to create a temporally consistent record of SST/SIC, which is updated in near-real-time. Intercomparisons showed that OSTIA has a similar climatology to HadISST2 in the global mean but better resolves small-scale variabilities. A slight cold bias was found in the first half of 2007, though the difference is negligible if compared with the analysis-to-analysis spread among the SST products used in this study. HadISST2 explicitly represents uncertainties in the analysis as ensemble realisations. The ensemble spread suggests that the uncertainty is reduced over time, though stable differences to the other products remain unexplained.

4D-Var DA experiments were conducted to show that, compared with HadISST2, OSTIA allows better use of surface wind observations such as ASCAT and this results in higher forecast scores in the troposphere. The improvements are mostly found around mid-latitude oceanic fronts. Recent studies (Chelton, 2005; Maloney and Chelton, 2006; Masunaga et al., 2015) have shown that the atmospheric boundary layer has a sensitivity to underlying SST anomalies associated with oceanic eddies. This study confirmed those findings in the context of atmospheric 4D-Var DA.

HadISST2 aims to create a climate record. Yet, it is not continued in real time. It has shown a comparable performance with OSTIA for a given DA configuration (T511L91). Considering also the good agreements in large scale variabilities, it seems acceptable to combine these products to make a single time series for ERA5.

Sea ice concentration also exhibits large discrepancies between products. HadISST2.1 comes with TR14 SIC dataset, which is historically consistent and is free from summertime melt-pond bias. However, this product is not available in near-real-time. We recommend that ERA5 should use the most consistent sea ice for the past period. The reprocessed OSI-SAF sea ice can be used together with OSTIA sea ice, while considering maximum possible uncertainties in them in the EDA.

The uncertainties found in this study motivated us to develop a new SST and SIC perturbation scheme. The main purpose is to provide consistent and relevant uncertainty statistics with the control SSTs/SICs. This is achieved in an empirical manner, using existing products to derive realistic error patterns to be used in ensemble applications. The error patterns do not strictly represent the flow dependency but take into account the seasonality of the uncertainty.

The new perturbation method was then evaluated with EDA experiments. We showed that with the new method the EDA produces more reliable spread and improves the quality of subsequent 4D-Var DA that uses the background error.

Table 4 proposes a plan for the use of SST and SIC in ERA5. For SST, HadISST2-pentad is used as the control SST until August 2007, and OSTIA follows thereafter. It is not yet determined which HadISST2 subversion to be used. The change of the datasets can take place at any point within the overlapping period from 2007 to 2010. The switching date, 1st of September 2007, is determined with two considerations: migrating to the higher resolution at the earliest possible date while avoiding the

cold period found at the beginning of OSTIA production. For SIC, two OSI-SAF products are used. The reprocessed version is used with HadISST2 and then switched to the operational one. In the ensemble ERA5 production, the control member uses the same SST and SIC as the deterministic system. For the other members, HadISST2 ensemble is further perturbed to account for structural uncertainties in SST and SIC. OSTIA is first perturbed with a proxy analysis error sampled from HadISST2 and then perturbed likewise. In this way SST and SIC are perturbed with a consistent set of uncertainties throughout the ERA5 period.

The plan above focuses on consistency in SST and SIC. To maintain this strength, the operational SST and SIC, OSTIA, requires a continuous monitoring. The resulting atmospheric reanalysis needs to be treated as an ensemble product rather than a deterministic when used. A similar effort should be followed in the context of coupled data assimilation. SST and sea ice lie on the boundary between the ocean and the atmosphere. It is expected that in the future a dynamically coupled reanalysis will improve flow dependent aspects of uncertainty in the oceanic and the atmospheric boundary layer. Future works also include a two-way approach - using a coupled reanalysis to help improve SST and SIC products and their uncertainty estimates through atmospheric and ocean observations.

Table 4: SST/SIC datasets to be used in ERA5. See Table 3 for reference.

Time period	SST		SIC	
	control	ensemble	control	ensemble
Jan.1979 - Aug.2007	HadISST2	HadISST2 + SE + IE	OSI-SAF reprocessing	IE
Sep.2007 -	OSTIA	AE + SE + IE	OSI-SAF	IE

Acknowledgements

The authors would like to thank John Kennedy, Nick Rayner and Holly Titchner (The UK Met Office Haley Centre) for giving us generous access to the latest HadISST2 versions. We appreciate comments from Chris J. Merchant (The University of Reading) on CCISST and its uncertainty estimation method. Dick Dee encouraged us to start this work. David Tan (St. Augustine's church in Hammersmith), Paul Poli (Météo-France), Massimo Bonavita and Elias Holm helped us set up and understand the data assimilation experiments in this study. Alan Geer developed and maintains a DA verification software "iver", which we fully exploited in this study. SH was seconded from the Japan Meteorological Agency from April 2014 to March 2016. He would like to express his sincere gratitude to all the staff at ECMWF.

References

Andersen, S., R. Tonboe, L. Kaleschke, G. Heygster, and L. T. Pedersen, 2007: Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration arctic sea ice. *Journal of Geophysical Research: Oceans (1978–2012)*, **112 (C8)**.

Bonavita, M., L. Isaksen, and E. Hólm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **138 (667)**, 1540–1559.

Chelton, D. B., 2005: The impact of SST specification on ECMWF surface wind stress fields in the eastern tropical Pacific. *Journal of Climate*, **18 (4)**, 530–550.

Chelton, D. B., M. G. Schlax, M. H. Freilich, and R. F. Milliff, 2004: Satellite measurements reveal persistent small-scale features in ocean winds. *Science*, **303 (5660)**, 978–983.

Chelton, D. B., and S.-P. Xie, 2010: Coupled ocean-atmosphere interaction at oceanic mesoscales. *Oceanography*, **23 (4)**, 52–69, doi:10.5670/oceanog.2010.05.

de Boisséson, E., M. Balmaseda, F. Vitart, and K. Mogensen, 2012: Impact of the sea surface temperature forcing on hindcasts of Madden-Julian Oscillation events using the ECMWF model. *Ocean Science*, **8 (6)**, 1071–1084.

Dee, D., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597.

Donlon, C., and Coauthors, 2007: The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bulletin of the American Meteorological Society*, **88 (8)**, 1197–1213.

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

Eastwood, S., T. Lavergne, and R. Tonboe, 2014: *Algorithm Theoretical Basis Document for the OSI SAF Global Reprocessed Sea Ice Concentration Product OSI-409*. The EUMETSAT Network of Satellite Appl. Fac.

Fiorino, M., 2004: A multi-decadal daily sea surface temperature and sea ice concentration data set for the ERA-40 reanalysis. *ERA-40 Project Report Series*, **12**, 16pp.

Flannaghan, T., S. Fueglistaler, I. Held, S. Po-Chedley, B. Wyman, and M. Zhao, 2014: Tropical temperature trends in atmospheric general circulation model simulations and the impact of uncertainties in observed SSTs. *Journal of Geophysical Research: Atmospheres*, **119 (23)**, 13–327.

Fueglistaler, S., C. Radley, and I. Held, 2015: The distribution of precipitation and the spread in tropical upper tropospheric temperature trends in CMIP5/AMIP simulations. *Geophysical Research Letters*, **42 (14)**, 6000–6007.

Hersbach, H., C. Peubey, A. Simmons, P. Berrisford, P. Poli, and D. Dee, 2015: ERA-20CM: A twentieth-century atmospheric model ensemble. *Quarterly Journal of the Royal Meteorological Society*, **141 (691)**, 2350–2375.

Hirahara, S., M. Ishii, and Y. Fukuda, 2013: Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate*, **27 (1)**, 57–75, doi:10.1175/JCLI-D-12-00837.1.

Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *Journal of Climate*, **21 (19)**, 5145–5153.

Inoue, J., J. A. Curry, and J. A. Maslanik, 2008: Application of aerosondes to melt-pond observations over Arctic sea ice. *Journal of Atmospheric and Oceanic technology*, **25 (2)**, 327–334.

Ivanova, N., O. M. Johannessen, L. T. Pedersen, and R. T. Tonboe, 2014: Retrieval of Arctic sea ice parameters by satellite passive microwave sensors: A comparison of eleven sea ice concentration algorithms. *Geoscience and Remote Sensing, IEEE Transactions on*, **52 (11)**, 7233–7246.

Kennedy, J., N. Reyner, S. C. Millington, and M. Saunby, 2016: The Met Office Hadley Centre sea ice and sea-surface temperature data set, version 2, part 2: seasurface temperature analysis. *in prep.*, URL http://www.metoffice.gov.uk/hadobs/hadisst2/.

Lyman, J. M., G. C. Johnson, and W. S. Kessler, 2007: Distinct 17-and 33-day tropical instability waves in subsurface observations. *Journal of physical oceanography*, **37 (4)**, 855–872.

Maloney, E. D., and D. B. Chelton, 2006: An assessment of the sea surface temperature influence on surface wind stress in numerical weather prediction and climate models. *Journal of Climate*, **19 (12)**, 2743–2762.

Markus, T., and D. J. Cavalieri, 2009: The AMSR-E NT2 sea ice concentration algorithm: Its basis and implementation. *Journal of The Remote Sensing Society of Japan*, **29 (1)**, 216–225.

Masunaga, R., H. Nakamura, T. Miyasaka, K. Nishii, and Y. Tanimoto, 2015: Separation of climatological imprints of the Kuroshio Extension and Oyashio fronts on the wintertime atmospheric boundary layer: Their sensitivity to SST resolution prescribed for atmospheric reanalysis. *Journal of Climate*, **28 (5)**, 1764–1787.

Merchant, C. J., and Coauthors, 2014: Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geosci. Data J.*, **1 (2)**, 179–191, doi:10.1002/gdj3.20.

Nomura, A., 1995: Global sea-ice concentration data set for use with the ECMWF re-analysis system. *ECMWF Re-Analysis Project (ERA) Technical Report*, **76**, 25pp.

Poli, P., H. Hersbach, P. Berrisford, D. Dee, A. Simmons, and P. Laloyaux, 2015a: ERA-20C deterministic. *ERA Report Series*, **20**, 48pp.

Poli, P., C. Peubey, K. Fennig, M. Schroeder, R. Roebelling, and A. Geer, 2015b: Pre-assimilation feedback on a fundamental climate data record of brightness temperatures from special sensor microwave imagers: A step towards mips4obs? *ERA Report Series*, **19**, 50pp.

Poli, P., and Coauthors, 2013: The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C). *ERA Report Series*, **14**, 59pp.

Rayner, N., D. E. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **108 (D14)**.

Reynolds, R. W., and D. B. Chelton, 2010: Comparisons of daily sea surface temperature analyses for 2007-08. *Journal of Climate*, **23 (13)**, 3545–3562.

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

Roberts-Jones, J., E. K. Fiedler, and M. J. Martin, 2012: Daily, global, high-resolution SST and sea ice reanalysis for 19852007 using the OSTIA system. *J. Climate*, **25 (18)**, 6215–6232, doi:10.1175/JCLI-D-11-00648.1.

Rösel, A., L. Kaleschke, and G. Birnbaum, 2012: Melt ponds on arctic sea ice determined from MODIS satellite data using an artificial neural network. *The cryosphere*, **6 (2)**, 431–446.

Schlax, M. G., and D. B. Chelton, 1992: Frequency domain diagnostics for linear smoothers. *Journal of the American Statistical Association*, **87 (420)**, 1070–1081.

Simmons, A., P. Poli, D. Dee, P. Berrisford, H. Hersbach, S. Kobayashi, and C. Peubey, 2014: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Quarterly Journal of the Royal Meteorological Society*, **140 (679)**, 329–353.

Thiébaux, J., E. Rogers, W. Wang, and B. Katz, 2003: A new high-resolution blended real-time global sea surface temperature analysis. *Bulletin of the American Meteorological Society*, **84 (5)**, 645–656.

Titchner, H. A., and N. A. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119 (6)**, 2013JD020 316, doi:10.1002/2013JD020316.

Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131 (612)**, 2961–3012.

Vialard, J., F. Vitart, M. A. Balmaseda, T. N. Stockdale, and D. L. T. Anderson, 2005: An ensemble generation method for seasonal forecasting with an oceanatmosphere coupled model. *Monthly Weather Review*, **133**, 441–453.

Wentz, F. J., C. Gentemann, D. Smith, and D. Chelton, 2000: Satellite measurements of sea surface temperature through clouds. *Science*, **288 (5467)**, 847–850.