RMetS
Royal Meteorological Society

Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System

Gregory C. Smith,[a]* François Roy,[a] Mateusz Reszka,[b] Dorina Surcel Colan,[b] Zhongjie He,[a] Daniel Deacu,[b] Jean-Marc Belanger,[a] Sergey Skachko,[c,d] Yimin Liu,[c] Frédéric Dupont,[b] Jean-François Lemieux,[a] Christiane Beaudoin,[a] Benoit Tranchant,[e] Marie Drévillon,[f] Gilles Garric,[f] Charles-Emmanuel Testut,[f] Jean-Michel Lellouche,[f] Pierre Pellerin,[a] Harold Ritchie,[g] Youyu Lu,[h] Fraser Davidson,[c] Mark Buehner,[i] Alain Caya[i] and Manon Lajoie[b]

[a]*Environment Canada, Recherche en Prévision Numérique Environnementale, Dorval, Québec, Canada*
[b]*Environment Canada, Meteorological Service of Canada, Dorval, Québec, Canada*
[c]*Fisheries and Oceans Canada, St. John's, Newfoundland, Canada*
[d]*Belgian Institute for Space Aeronomy, BIRA-IASB, Brussels, Belgium*
[e]*Collecte Localisation Spatiale, Toulouse, France*
[f]*Mercator-Océan, Toulouse, France*
[g]*Environment Canada, Recherche en Prévision Numérique Environnementale, Dartmouth, Nova Scotia, Canada*
[h]*Fisheries and Oceans Canada, Dartmouth, Nova Scotia, Canada*
[i]*Environment Canada, Data Assimilation and Satellite Meteorology Research, Dorval, Québec, Canada*

*Correspondence to: G. C. Smith, 2121 Transcanada Highway, Dorval, H9P 1J3, Canada. E-mail: gregory.smith@ec.gc.ca

Reproduced with the permission of the Minister of Environment Canada

Recent increases in marine traffic in the Arctic have amplified the demand for reliable ice and marine environmental predictions. This article presents the verification of ice forecast skill from a new system implemented recently at the Canadian Meteorological Centre called the Global Ice Ocean Prediction System (GIOPS). GIOPS provides daily global ice and ocean analyses and 10-day forecasts on a 1/4°-resolution grid. GIOPS includes a multivariate ocean data assimilation system that combines satellite observations of sea-level anomaly and sea-surface temperature (SST) together with *in situ* observations of temperature and salinity. Ice analyses are produced using a 3D-Var method that assimilates satellite observations from SSM/I and SSMIS together with manual analyses from the Canadian Ice Service. Analyses of total ice concentration are projected onto the thickness categories used in the ice model using spatially and temporally varying weighting functions derived from ice model tendencies. This method may reduce deleterious impacts on the ice thickness distribution when assimilating ice concentration, as it can directly modulate (and reverse) nonlinear processes such as ice deformation. An objective verification of sea ice forecasts is made using two methods: analysis-based error assessment focusing on the marginal ice zone, and a contingency table approach to evaluate ice extent as compared to an independent analysis. Together the methods demonstrate a consistent picture of skilful medium-range forecasts in both the Northern and Southern Hemispheres as compared to persistence. Using the contingency table approach, GIOPS forecasts show a significant open-water bias during spring and summer. However, this bias depends on the choice of threshold used. Ice forecast skill is found to be highly sensitive to the assimilation of SST near the ice edge. Improved observational coverage in these areas (including salinity) would be extremely valuable for further improvement in ice forecast skill.

Key Words: verification; Canadian; global; polar prediction; data assimilation; operational oceanography; ocean modelling; sea ice modelling

Received 28 February 2014; Revised 9 March 2015; Accepted 18 March 2015; Published online in Wiley Online Library 13 May 2015

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

As numerical weather prediction (NWP) systems become further refined, the interactions across the air–ice–ocean interface are becoming increasingly important (e.g. Smith *et al.*, 2013b). This is giving rise to the development of a new generation of fully integrated environmental prediction systems composed of atmosphere, ice, ocean and wave modelling and analysis systems. Such systems are in increasing demand as the utility of marine information products (e.g. for emergency response) becomes more widely recognized.

While the potential benefit of atmosphere–ice coupling for NWP has been suggested by several studies (e.g. Pellerin *et al.*, 2004), it is currently not clear whether ice forecasting systems are sufficiently skilful to support their use within NWP systems. Moreover, the application of numerical ice forecasts within the ice services community has been limited, due in part to uncertainty regarding forecast skill and a lack of understanding of how to develop useful products.

While several systems are currently producing routine sea ice forecasts, relatively little has been published to date regarding their forecast skill. Perhaps the most significant study is that of Van Woert *et al.* (2004), who provide a detailed evaluation of the Polar Ice Prediction System (PIPS) using an analysis-based verification method that focuses on changes in the simulated and analysed sea ice concentration. Using this method, PIPS was shown to have only marginally significant skill, with correct forecasts only 25% of the time and with little or no skill in winter (November to January). However, it is not clear whether this forecast error is underestimated, as errors in the verifying analysis are not taken into account. Smith *et al.* (2013a) propose an alternative verification method using RADARSAT synthetic aperture radar image analyses. While this method provides a highly detailed evaluation of both ice concentration and thickness, it is limited in terms of both spatial and temporal coverage. Hernandez *et al.* (2009) provide a sample of sea ice forecast skill from the TOPAZ system (Towards an Operational Prediction system for the North Atlantic European coastal Zones; Bertino and Lisæter, 2008) as compared to persistence, in a region near Bering Strait. However, given the small size of the region it would be difficult to apply the results more generally. Similarly, Sakov *et al.* (2012) show ice concentration innovation statistics from the TOPAZ4 reanalysis over the Arctic and suggest that larger errors are present during the summer season, although little detail is provided. As such, the main source of information regarding ice forecast verification methods and skill estimates is currently found in various technical reports (Posey *et al.*, 2010; Melsom *et al.*, 2011).

Here, we provide an objective assessment of sea ice forecast skill, using two complementary verification methods, in a new system implemented recently at the Canadian Meteorological Centre (CMC) called the Global Ice Ocean Prediction System (GIOPS). The first method, based on that of Van Woert *et al.* (2004), is an analysis-based verification that employs a targeted spatial sub-sampling such that only areas of change in the analysed ice concentration are considered. The second method uses a contingency table approach verifying against an independent binary analysis of ice coverage (ice/water: Buehner *et al.*, 2013a).

Section 2 provides a description of GIOPS, including the ice–ocean model, the ocean and ice data assimilation methods and the method used to project changes in ice concentration onto the different ice thickness categories. Section 3 outlines the ice verification techniques used. The objective verification of GIOPS sea ice forecasts is presented in section 4. Conclusions and a discussion of sources of forecast errors and future directions are presented in section 5.

2. Description of GIOPS

The GIOPS system has been running in real time at CMC since 2011. The system details and verification scores presented here correspond to GIOPSv1.1.0 implemented on 25 June 2014.

2.1. Ice–ocean model

The numerical model used is the NEMO-CICE coupled ice–ocean model based on NEMO (Nucleus for European Modelling of the Ocean) version 3.1 (Madec *et al.*, 1998; Madec, 2008) and CICE (Community Ice CodE) version 4.0 (Hunke and Lipscomb, 2010). NEMO is a primitive equation z-level model making use of the hydrostatic and Boussinesq approximations. The model employs a linearized free surface (Roullet and Madec, 2000) with partial cell topography (Adcroft *et al.*, 1997). The version used here has a tri-polar 'ORCA' grid and 50 levels in the vertical, with vertical spacing increasing from 1 m at the surface to 500 m at the ocean bottom.

The model configuration has an eddy-permitting global 1/4° resolution (referred to as ORCA025). The ORCA025 configuration has been developed through the DRAKKAR Consortium (Barnier *et al.*, 2007) and uses model parameter settings as defined in Barnier *et al.* (2006) and Penduff *et al.* (2010). An energy–enstrophy conserving momentum advection scheme (Barnier *et al.*, 2006) is used, with Laplacian diffusion. A biharmonic operator is used for horizontal mixing of momentum. Vertical eddy diffusivity and viscosity coefficients are computed from a level 1.5 turbulent closure scheme using a prognostic equation for the turbulent kinetic energy (Blanke and Delecluse, 1993). More details can be found in Barnier *et al.* (2006) and Penduff *et al.* (2007, 2010).

CICE is a dynamic–thermodynamic sea ice model. It is a continuum-based model, that is, it does not track individual ice floes but rather calculates the evolution of a thickness distribution of the ice pack within grid cells. The thickness distribution evolves because of both thermodynamic processes (vertical growth/melt and lateral melt) and dynamic processes (advection and redistribution). Here, we use ten thickness categories, with boundaries between categories at 10, 15, 30, 50, 70, 120, 200, 400 and 600 cm. The additional categories used here (the default value in CICE is five categories) allow a more detailed representation of both very thin ice (less than 1 m) and thicker multi-year ice. The dynamic component calculates the velocity field by solving explicitly the two-dimensional (2-D) sea ice momentum equation using the Elastic Viscous Plastic (EVP) approach (Hunke, 2001). The common elliptical yield curve of Hibler (1979) is used and the ice strength is calculated according to Rothrock (1975) for a multi-thickness category model (see also Lipscomb *et al.* (2007)). A detailed description of the equations describing the dynamics and the thermodynamics of CICE can be found in Hunke (2001), Lipscomb *et al.* (2007) and Hunke and Lipscomb (2010).

The thermodynamic component calculates growth/melt of snow and ice, and the temperature profile in the vertical, by solving a heat diffusion equation. In GIOPS, the default number of layers (four ice layers and one for the snow) is used. The upper boundary condition is the net heat flux exchanged with the atmosphere, whose components are the latent, sensible, incoming and outgoing long-wave and short-wave radiation fluxes. We use the default CCSM3 (Community Climate System Model version 3) scheme to calculate the albedo and the attenuation of the absorbed short-wave radiation. In vertical salt exchange the ice is assumed to have a salinity of 4 psu (practical salinity units). However, we neglect the effect of brine pockets on heat storage to prevent instabilities in regions of low salinity. The lower boundary condition is the heat flux from the ocean to the ice. Based on the temperature profile and the boundary conditions, growth (at the base) or melt (at the base and/or at the top) are calculated. Lateral melting depends on the average diameter of the ice floes (Steele, 1992), a parameter specified here to be 300 m. New ice that forms over open water has a specified thickness of 5 cm. Sensitivity tests (not shown) suggest that this

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

value has little impact on ice volume but can affect short-range forecast skill as it changes the concentration along the marginal zone.

Atmospheric forcing for NEMO-CICE is taken from the lowest dynamic model level of CMC forecasts from the Global Deterministic Prediction System (GDPS: Charron *et al.*, 2012), which is currently at a resolution of 25 km. Fluxes are calculated using the Coordinated Ocean Reference Experiments (CORE) bulk formula (Large and Yeager, 2004) adapted for application at the height of the bottom GDPS model level (roughly 40 m). For consistency in the calculation of momentum, heat and moisture fluxes, the surface roughness of sea ice is set to the same value used in the GDPS (1.6×10^{-4} m). Use of the default value from CICE (5.0×10^{-4} m) was found to lead to an excessive transfer of momentum from the atmosphere to the ice, resulting in exaggerated ice drift speeds and an accumulation of thick ice in the Beaufort Gyre (not shown). As is commonly done in NWP, the GDPS persists the initial sea ice concentration and SST fields. This is expected to result in some inconsistency in fluxes applied to GIOPS.

2.2. Ocean assimilation system

The ocean analysis system used in GIOPS is the Système d'Assimilation Mercator version 2 (SAM2). SAM2 is a data assimilation tool designed for the regional and global oceans, which has been developed for different NEMO configurations by Mercator-Océan (Tranchant *et al.*, 2008; Cummings *et al.*, 2009; Lellouche *et al.*, 2013). The analysis method is based on a Reduced Order Kalman Filter using a SEEK (Singular Evolutive Extended Kalman) formulation (Pham *et al.*, 1998) derived from the Kalman filter. Background error covariances are modelled by an ensemble of multivariate three-dimensional anomalies derived from a multi-year hindcast simulation (Lellouche *et al.*, 2013). SAM2 assimilates observations of sea-level anomaly (SLA), sea-surface temperature (SST), as well as *in situ* temperature and salinity profiles. Satellite observations of SLA are obtained from the Archiving, Validation and Interpretation of Satellite Oceanographic data (AVISO) Ssalto/Duacs near-real-time product and currently include Jason2, Cryosat2 and Saral/Altika data. SLA data are assimilated using the mean dynamic topography of Rio and Hernandez (2004). *In situ* observations are taken from a variety of sources including the Argo network of autonomous profiling floats (Gould, 2005), moorings, ships of opportunity, marine mammals and research cruises. SST is assimilated using the CMC operational SST analysis (Brasnett, 2008; Martin *et al.*, 2012) for consistency with NWP systems at CMC.

A detailed description of the SAM2 system design and performance is provided in Lellouche *et al.* (2013). The version used here is similar to that referred to as IRG_V0 (or PSY3V2R1) in Lellouche *et al.* (2013) with several important differences. First, we use a more recent version of NEMO (v3.1; i.e. similar to that used in IRG_V1V2) with an adapted version of the CORE bulk formula (mentioned above). Second, in place of the single ice category model used by Mercator-Océan, we use the multi-category sea ice model CICE. CICE is used within GIOPS in order to maintain consistency with other CMC systems, such as the Regional Ice Prediction System (Lemieux *et al.*, 2014). Most importantly, GIOPS ice fields are constrained by the assimilation of sea ice concentration observations (described in the following section), whereas those in IRG_V0 are not.

The assimilation of SST is also significantly modified from that in IRG_V0. Here, the operational CMC SST analysis is interpolated onto the ORCA025 grid and assimilated directly with a constant error of 0.3 °C. The use of a smaller SST observation error than IRG_V0 is applied as it corresponds more closely to the estimated error from the CMC SST analysis (Brasnett, 2008) and also to provide a more tightly constrained SST to reduce initialization shock when using GIOPS analyses in coupled medium-range forecasts with the GDPS (Smith *et al.*, 2013b). In contrast to IRG_V0, no damping of increments or SST error amplification is used near the ice edge. Rather, the CMC SST analysis is set to the freezing point in locations where the ice analysis has a concentration greater than 20%. Using this method the ice analysis is used as a proxy for freezing-point temperature under ice and thus can impact the full model state through the multivariate covariances. In particular, this is expected to provide a better representation of mixed-layer properties in the marginal ice zone.

Finally, the system implementation is somewhat different than that used by Mercator-Océan. As is done for IRG_V0, a delayed-mode analysis is produced every Tuesday valid 6 days prior (i.e. the previous Wednesday), using a 7-day assimilation window. This allows the assimilation of a greater number of *in situ* observations and satellite altimetry data. This is followed by a real-time analysis valid every Wednesday at 0000 UTC also using a 7-day assimilation window and initialized from the delayed-mode cycle. In addition to the IRG_V0 weekly cycles, GIOPS employs a 1-day real-time analysis cycle for other days of the week, which assimilate only SST observations. These analyses are then blended with ice analyses and used to produce daily 10-day ice–ocean forecasts in real time. The combination of weekly 7-day and daily 1-day analysis cycles allows a compromise between high-quality delayed analyses with more observations and real-time availability, in a manner which maintains tightly constrained surface fields of temperature and sea ice concentration.

2.3. Sea ice concentration assimilation system

The sea ice assimilation system used by GIOPS is based on the 3D-Var method developed for regional applications at CMC (Buehner *et al.*, 2013a, 2015) implemented on a global domain at ~10 km resolution (Buehner *et al.*, 2013b). The ice analysis system uses a 6 h window centred at 0000, 0600, 1200 and 1800 UTC. Due to practical considerations, the system currently does not use the ice model forecast as background, but rather the ice analysis produced 6 h earlier is used instead.

The 3D-Var analysis system assimilates both passive microwave satellite observations (SSM/I and SSMIS) and Canadian Ice Service (CIS) manual analyses. CIS manual analyses are produced daily using a combination of remotely sensed data, ship and aircraft reports and RADARSAT image analyses. The retrieved ice concentration from passive microwave sensors is calculated from observed brightness temperatures using the National Aeronautics and Space Administration (NASA) Team 2 (NT2) sea ice algorithm (Markus and Cavalieri, 2000). To account for the difference between the larger footprint of the passive microwave channels used in the NT2 algorithm and the analysis grid, a footprint operator is applied in the observation operator (Buehner *et al.*, 2013a).

An important aspect of the ice analysis system is the quality control performed on the observations. In particular, passive microwave observations are rejected wherever the surface air temperature is above freezing, as well as where the SST > 4 °C. Buehner *et al.* (2013b) show that as compared to manual analyses from the National Ice Center (NIC), standard deviations of differences in ice concentration are roughly 20% lower in 3D-Var analyses than in the previous operational ice analyses produced at the CMC. An improved fit with NIC analyses is a strong indication that the 3D-Var analyses provide a more robust estimate of ice concentration that the previous operational ice analysis produced at CMC. However, as the manual NIC analyses also contain errors, differences from NIC analyses can not necessarily be considered as errors.

2.4. Ice–ocean blending algorithm

Prior to model initialization (e.g. either as part of an assimilation cycle or for a forecast), the 3D-Var total ice concentration analysis is blended with fields from the ice–ocean model. As the ice model

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

uses ten ice thickness categories, a special treatment is required in order to determine how the increment in total ice concentration is projected onto the various ice thickness categories. A description of two options and a new method developed for GIOPS is described below.

Here we seek to update a set of model partial sea ice concentrations, C_{fi} $(i = 1, N)$, at a particular grid point (where N is the number of thickness categories) using a total concentration analysis increment, dA_a, where $dA_a = A_a - A_f$, A_a is the total analysis concentration from the 3D-Var ice analysis and A_f is the total forecast concentration, such that:

$$A_f = \sum_{i=1}^{N} C_{fi}. \tag{1}$$

To do so, we require a set of partial concentration analysis increments, dC_{ai}. We thus need to specify a set of weights, w_i, such that $dC_{ai} = dA_a * w_i$, whereby the sum of all weights equals 1.

Without any *a priori* knowledge concerning the source of the model error (e.g. formation, melt, advection), we could simply rescale the existing ice thickness distribution (referred to hereafter as the RED method) as:

$$w_i = \frac{C_{fi}}{A_f}. \tag{2}$$

The RED method will correctly address errors associated with the advection of a homogenous ice cover or with the amplitude of new ice formation in a grid cell. However, with this method additional ice formation, melt and deformation will be represented incorrectly, as these processes act nonlinearly on the different ice categories. In particular, systematic (repeated) corrections could lead to an accumulation of errors in higher-thickness categories.

For example, let us consider the case of a model bias resulting in excessive melt in a two ice category model. We assume that ice from the thinnest category melts completely and that only ice in the thicker category remains. If we apply the RED method to correct the total ice concentration, the partial concentration of the thicker category would be increased but no thin ice would be added. As a result, if this correction were applied repeatedly at each assimilation cycle a further increase of the partial concentration of thick ice would be applied resulting in an unphysical increase in ice volume.

Ideally, one would like to adjust the model thicknesses based on knowledge of which physical processes have led to the error in total concentration. Diagnosing this directly from the model is not straightforward as many processes may be active simultaneously. However, if we assume that the errors developed entirely over the forecast period, then we can use the differences between the partial concentrations from the forecast (C_{fi}) and those from the previous analysis (C_{0i}) to construct our weighting function, namely:

$$w_i = \frac{C_{fi} - C_{0i}}{A_f - A_0}, \tag{3}$$

where A_0 is the total ice concentration in the analysis used to initialize the forecast.

In practice, model errors will include spatial errors (e.g. a shift in the location of a formation event), the misrepresentation of physical processes (e.g. ridging) or missing processes altogether (e.g. a lack of mixing due to overly warm SSTs). As such, approximating these errors using only the change in partial concentrations is a fairly crude approach. Nonetheless, it does provide a means to enhance or reverse nonlinear processes captured by the model.

The assumption that model changes are representative of errors is only valid for sufficiently large changes over the model forecast period (an overly small denominator in Eq. (2) could lead to noisy and unphysical weights). To account for this limitation, we only apply this method when model tendencies over the forecast period are sufficiently large. In particular, we limit its use based on the ratio of the magnitude of model changes over the forecast period over the analysis increment, namely:

$$R = \left| \frac{A_f - A_0}{dA_a} \right|. \tag{4}$$

For large values of R, the evolution of model partial concentrations can be considered representative of random model error, while for small values of R the errors are likely to be due to other factors (e.g. missing model physics, errors in model forcing) and an alternate method must be applied (such as RED, Eq. (1)). Here we use a threshold value of $R = 0.5$, below which the RED method is used. This value ensures that model changes are at least half the size of differences from the analysis. We hereafter refer to this method as the Rescaled Forecast Tendencies (RFT) method. Sensitivity tests showed that this threshold allowed the widespread application of the RFT method but prevented errors associated with the amplification of very small model tendencies.

It is also important to note a few additional details regarding the ice blending algorithm. First, the blending is only applied when both the model field and the ice analysis have a concentration of less than 90%. This allows the model to maintain its representation of leads in the pack ice, which are not well observed. Additionally, no blending is applied when the difference between the model field and the ice analysis is less than 1%. Moreover, to remove small concentrations of ice that are residuals of the 3D-Var analysis procedure, ice blending is not applied in grid cells where both model fields and the ice analysis have less than 10% ice cover. When ice is present in the analysis but not in the model, the ice thickness is derived from the previous ice analysis (i.e. we assume erroneous melt). If there is no ice in the previous analysis, then it is assumed this is newly formed ice with a nominal thickness of 20 cm. This value is higher than the value used for newly formed ice in CICE (5 cm) to account for continued growth as the ice may have formed at any time during the analysis cycle (i.e. up to 7 days prior).

To evaluate the impact of this thickness modification method, we examine the evolution of ice area and volume over a 1-year period. The ocean model is initialized on 8 December 2010 from a PSY3V2R1 analysis produced by Mercator, while the ice fields are initialized from a 9-year spin-up (2002–2010) forced by the Canadian GDPS Reforecasts (CGRFs: Smith *et al.*, 2014). Experiments are then performed over the year 2011 with the full ocean and sea ice concentration analysis systems. Three experiments are performed: a free run with no assimilation (FREE) and two assimilative runs using the RED and RFT methods. The aim here is not to provide a detailed evaluation of the methods but to provide an indication of the impact that the choice of method can have on the total ice volume.

Results are shown for the Northern and Southern Hemispheres in Figure 1. The impact of assimilation provides an overall increase in the ice area of several million square kilometres over both hemispheres with respect to the FREE experiment. This is accompanied by a commensurate increase in ice volume. For the Arctic, all three experiments show roughly the same ice area in autumn, while an increase in ice volume for RFT and RED with respect to the FREE experiment can be seen. This suggests that the modification to the ice thickness distribution in the assimilation procedure may be introducing errors. When the RFT method is used in place of RED, a significant reduction of this increase in ice volume occurs. Spatial maps (not shown) indicate that the RED method produces anomalous patches of thick ice due to the accumulation of errors (e.g. as described above) in both hemispheres, with the largest errors occurring in the Southern Hemisphere.

While comparison of the RED and RFT methods with a model free run should not be confused with a more meaningful

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

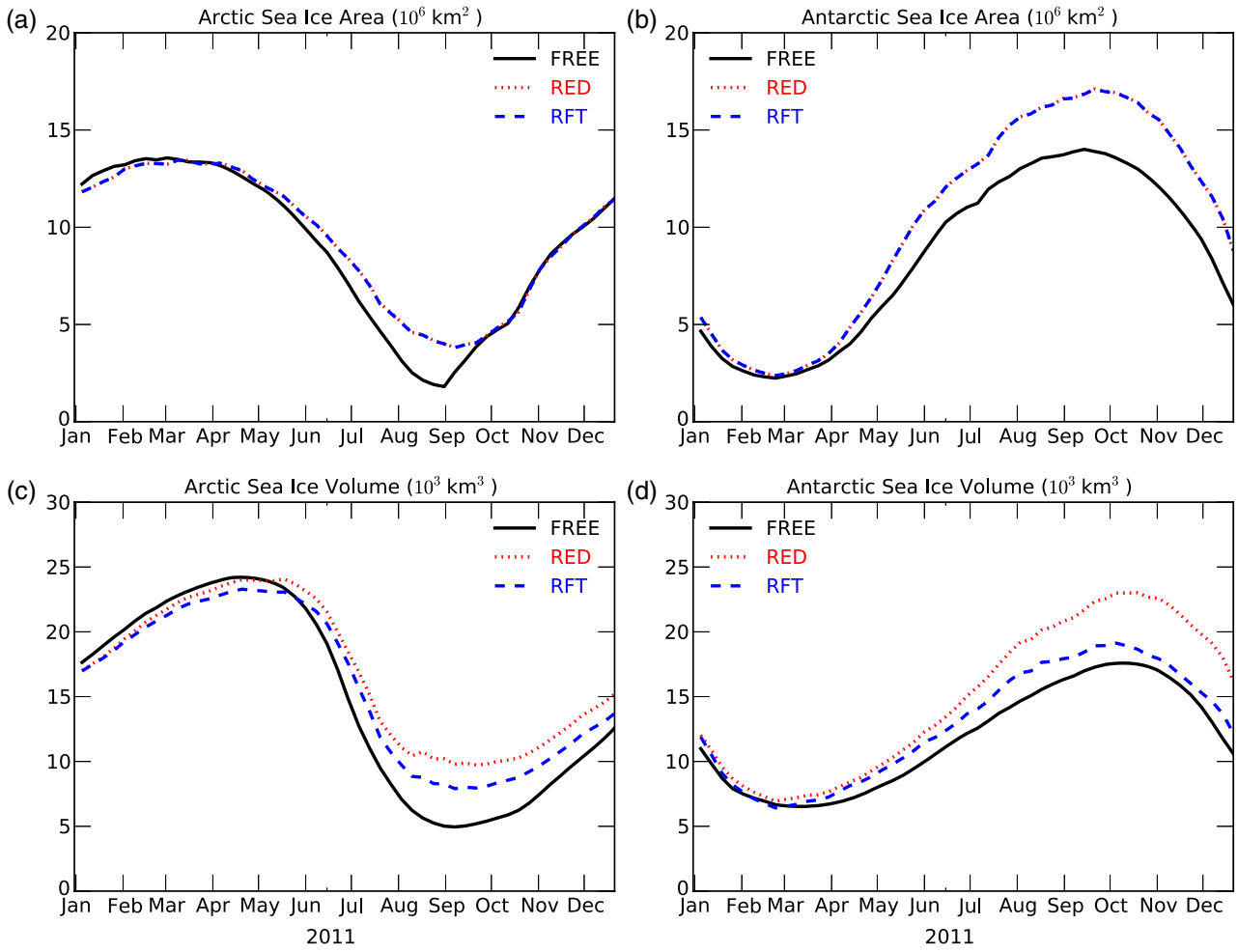

Figure 1. Impact of the sea ice blending method on total ice area ($\times 10^6$ km^2; top row), and volume ($\times 10^3$ km^3; bottom row) for the Northern (left column) and Southern (right column) Hemisphere for the year 2011.Three experiments are shown: a free run with no assimiliation (FREE; black curves) and assimilative runs using the RED (blue curves) and RFT (red curves) methods.

evaluation against observations, it does provide an indication of the impact these methods can have on ice thicknesses. Moreover, for both ice concentration verification methods presented here (sections 4.1 and 4.2) a small improvement in forecast skill is found for RFT over RED during the months of June and July over the Northern Hemisphere (not shown). However, further study is clearly required to better understand the sources of error and how to optimize these methods.

3. Ice verification methods

3.1. Spatially targeted analysis verification

The most straightforward means to evaluate sea ice concentration forecasts is simply to compare model forecasts at a specific lead time to the analysis valid at the same time. The difficulty with this method is that areas where little or no data have been assimilated will be included, leading to no change in the analysis over the lead time. This makes comparisons unreliable in these areas as the persistence of initial conditions will tend to be favoured. Verification of model forecasts with analyses over these data-poor areas could lead to a misinterpretation of differences as being errors, when in fact the model could be correct. For example, due to the contamination from land in the footprint of satellite retrievals (which are not assimilated), the 3D-Var analysis is expected to underrepresent the formation of coastal polynyas. Evaluating model forecasts over these areas could thus unfairly penalize the model for correctly forecasting these features. Similarly, the formation of leads in the pack ice could also be seen as errors when compared against 3D-Var analyses.

To account for these drawbacks, and to focus on areas of activity (i.e. along the ice edge), a filter is imposed such that we only include points in the evaluation where the concentration has changed by more than 10% in the analysis over the forecast lead time. This spatial targeting method provides a focused evaluation on the most relevant areas of the domain, resulting in a more reliable assessment of forecast skill and error. Note that this method is similar to that proposed by Van Woert *et al.* (2004) except that they included the areas where either the model or the analysis had changed over the forecast lead time. As this may result in the inclusion of areas of high analysis uncertainty, we prefer to sub-sample the domain using only areas of change in the analysis. The disadvantage of this method is that certain aspects of model error may remain unevaluated, such as incorrect coastal polynya formation and false alarms (ice formation). An additional drawback is that it does not permit an evaluation of areas of analysis error.

3.2. Contingency table analysis

To provide a complementary evaluation of forecast skill, GIOPS ice forecasts are verified against the Interactive Multisensor Snow and Ice Mapping System (IMS) ice extent product (Ramsay, 1998; Helfrich *et al.*, 2007) produced manually at the NIC. The IMS daily product provides binary values of ice/open water on a 4 km resolution grid. It is generated by using information from a wide variety of satellite imagery, mapped products and surface observations. As the IMS and GIOPS ice analyses are produced using different methods, quality control and observational sources (albeit they share the use of some passive microwave data), the

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Table 1. Contingency table entries for sea ice verification of GIOPS forecasts as compared to IMS binary analyses of ice or open water.

	IMS ice	IMS water
Forecast ice	Hit ice (a)	False alarm (b)
Forecast water	Miss (c)	Hit water (d)

verification of GIOPS forecasts against IMS analyses provides an independent measure of skill. However, errors in IMS analyses are not taken into account and thus the results need to be considered together with other methods.

As the IMS analyses provide only binary values of ice/open water, a 40% concentration threshold (the same threshold used subjectively in the IMS production) is applied to model forecasts. Contingency tables are then constructed (see Table 1) using model forecasts as well as the persistence of the ice analysis used to initialize the model. Of particular interest are the following derived quantities (for all these quantities a value of one is a perfect score):

- Proportion Correct Ice, $PCI = a/(a+c)$
- Proportion Correct Water, $PCW = d/(b+d)$
- Proportion Correct Total, $PCT = (a+d)/n$
- Frequency Bias, $BIAS = (a+b)/(a+c)$

where n is the total number of points on the IMS grid.

4. Verification of ice forecast skill

It is currently standard practice in NWP to persist the initial sea ice concentration throughout the forecast period. Thus, here the skill of GIOPS forecasts will be evaluated in comparison with those obtained for the persistence of the initial 3D-Var analyses (i.e. the GIOPS initial condition). This provides an indication of the potential gain for a coupled atmosphere–ice–ocean forecasting system. An alternative comparison would be to compare to the persistence of the sea ice anomaly with respect to climatology as proposed by Van Woert *et al.* (2004). As the latter method removes the effect of the seasonal cycle, it is expected to give somewhat smaller estimates of forecast skill. Here, we choose to use persistence of the initial ice analysis for comparison due to its relevance for coupled environmental prediction.

In addition to comparing forecasts to persistence, we also evaluate the skill of trial fields. Trial fields are produced as part of the data assimilation cycle and benefit from improved atmospheric forcing available at the time the analysis is produced. That is, trial fields are produced by forcing the ice–ocean model by atmospheric fields from seven successive atmospheric forecasts produced daily at 0000 UTC using lead times from 6 to 27 h (the first 6 h of each forecast are neglected as they may be negatively affected by initialization shock). In contrast, forecasts are produced using atmospheric fields from a single atmospheric forecast (i.e. 0–240 h). Use of successive daily atmospheric forecasts is expected to result in more accurate surface fluxes, as the atmospheric fields will have smaller forecast error. Moreover, the atmospheric fields will benefit from updated ice analyses, whereas use of 0–240 h from a single atmospheric forecast will result in errors due to the persistence of the initial ice concentration over the forecast period. As a result, important inconsistencies between GIOPS forecasts and atmospheric fields may develop in regions where the ice concentration evolves rapidly. Comparison of forecast and trial fields may thus be viewed as an estimate for the potential impact of coupling with NWP systems on sea ice forecast skill.

4.1. *Spatially targeted analysis verification*

Using the method outlined in section 3.1, 10-day forecasts performed every Wednesday from 12 January 2011 to 28 December 2011 were evaluated against the 3D-Var ice analyses produced at 0000 UTC. The forecast skill for both Northern and Southern Hemispheres as a function of lead time, seasonality and spatial distribution are calculated. For the latter two diagnostics, a lead time of 168 h (7 days) is used as it is fairly representative of the forecast error over the 5–10-day range for which GIOPS forecast products would be most useful for coupling with NWP and for marine applications. Since 168 h is also the length of the assimilation window, this provides a useful reference for comparing errors in the trial fields with actual forecasts (see below).

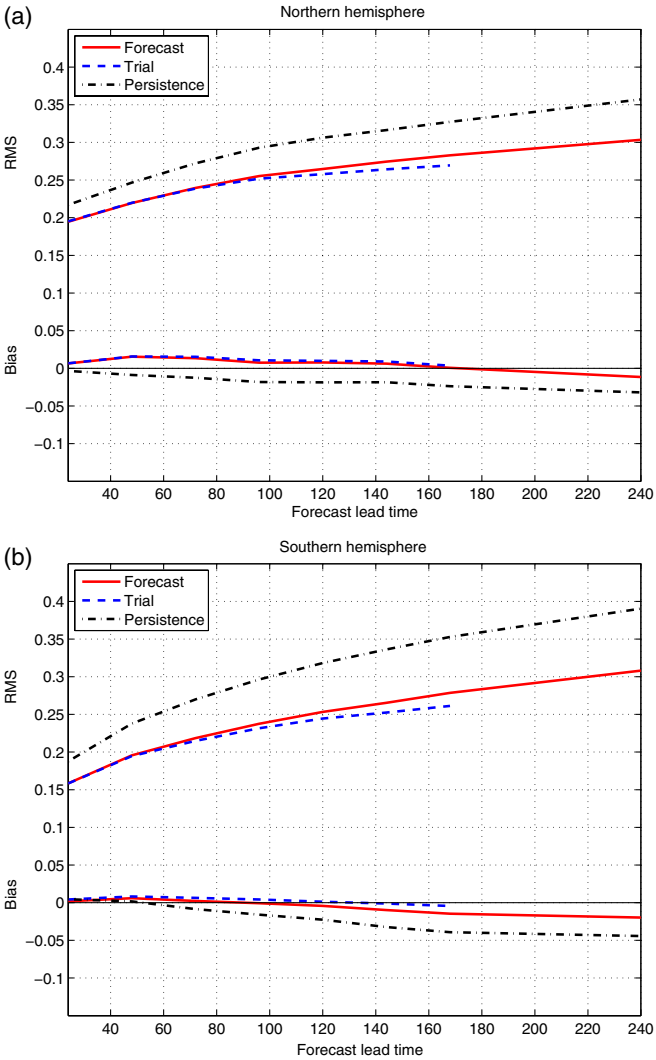

Figure 2. Evaluation of sea ice concentration as a function of lead time (h) using the spatially targeted analysis verification method for the Northern (a) and Southern (b) hemispheres. Root-mean-squared (RMS) and mean (bias) errors are shown as a function of forecast lead time based on weekly forecasts produced for 2011. GIOPS forecasts are shown as a solid red line, trial fields as a dashed blue line and the persistence of initial 3D-Var analyses as a black dot-dashed line.

GIOPS forecast error as a function of lead time for both hemispheres is shown in Figure 2 as compared to trial fields and the persistence of the initial 3D-Var analysis. Both GIOPS trial fields and forecasts show an important reduction in root-mean-squared error (RMSE) in ice concentration over the full 10-day forecast period for both hemispheres, even at short lead times. Error growth is faster over the first few days and gradually decreases with time. This transition may be associated with a change in the source of error from synoptic-scale variability (e.g. due to the passing of weather systems) to the evolution of the seasonal cycle on longer time-scales. Moreover, the steep error growth at lead times less than about 48 h highlights the large variability of the sea ice cover on hourly-to-daily time-scales and the importance of having short observation cut-off times in the assimilation of sea ice concentration data to provide the most

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

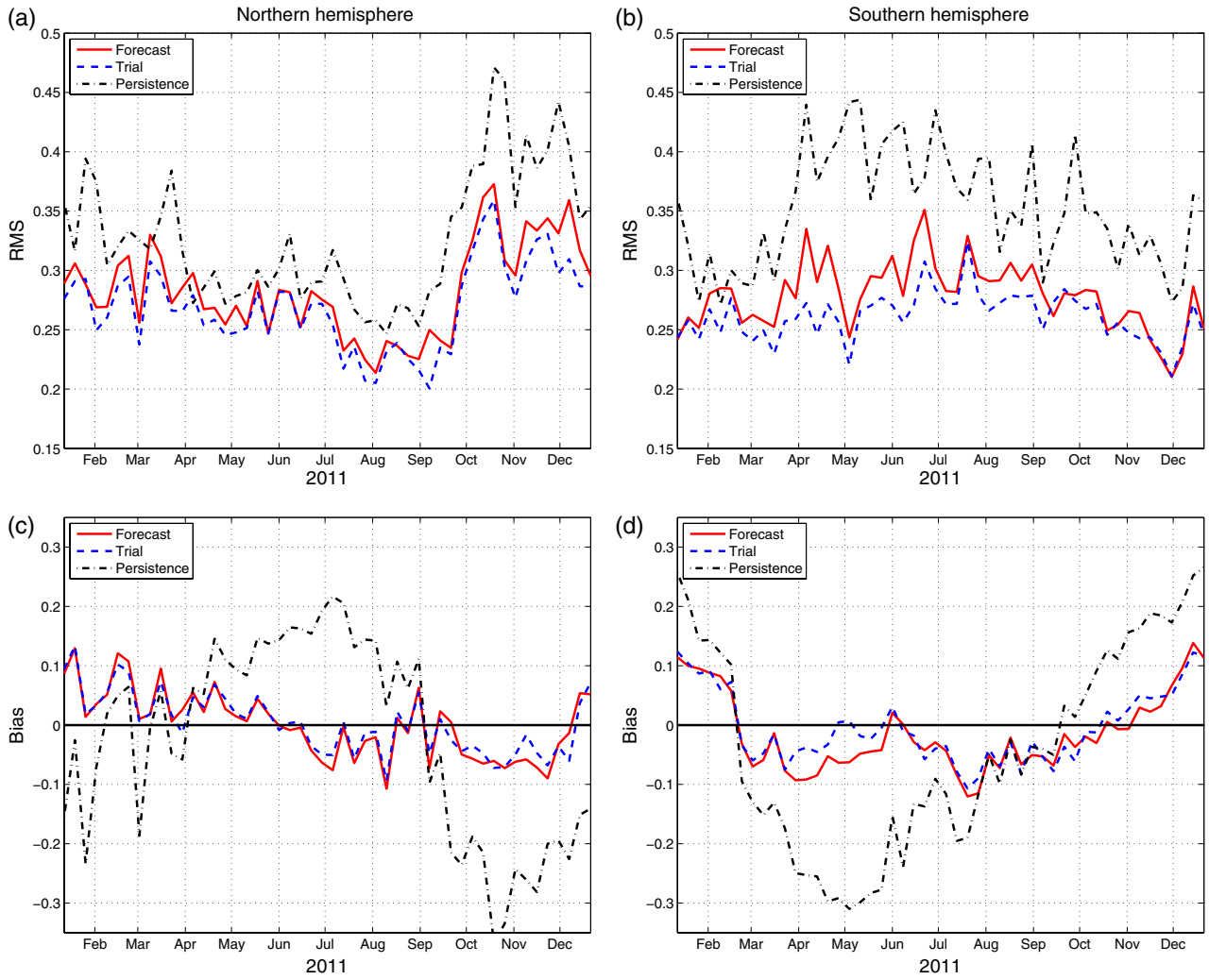

Figure 3. Evaluation of sea ice concentration at a lead time of 168 h using the spatially targeted analysis verification method. The seasonality of root-mean-squared (RMS; (a,b)) and mean (bias; (c,d)) errors are shown throughout the year based on weekly forecasts produced for 2011 for both the Northern (a,c) and Southern (b,d) Hemispheres. GIOPS forecasts are shown as a solid red line, trial fields as a dashed blue line and the persistence of initial 3D-Var analyses as a black dot-dashed line.

up-to-date information on the ice cover. Note that this initial rapid increase in error is not due to initialization shock in the model as it is also present for the persistence of the ice analysis.

Overall, model forecasts tend to exhibit smaller negative biases in ice concentration than persistence. However, the biases vary considerably by region and season (as discussed below).

Trial fields show somewhat smaller RMSE than forecasts for both hemispheres. Reduced RMSE can be seen in particular for days 4–7, with little difference in error growth apparent between forecasts and trial fields over the first 3 days. The percentage reduction in RMSE with respect to persistence at a lead time of 168 h for the Northern Hemisphere is 13% for forecasts and 17% for trial fields, with values for the Southern Hemisphere of 20 and 24% respectively. The larger errors in forecasts are due in part to the lack of an evolving ice cover in the atmospheric forecasts used to force GIOPS forecasts.

Figure 3 shows the seasonality of forecast skill for the Northern and Southern Hemispheres at a lead time of 168 h. The RMSEs of trial fields and forecasts in the Northern Hemisphere show a fairly stable improvement of about 0.05 and 0.04 respectively as compared to persistence over the year. In the Southern Hemisphere, the RMSEs in persistence show a strong seasonal cycle not seen in the forecasts. As a result, improvements with respect to persistence show a strong seasonal cycle, with the smallest gains being present in summer, while the largest values occur during the rapid growth period in austral autumn.

The advance and retreat of the ice cover through the seasons can be seen from the biases in persistence (Figure 3, bottom row), with a positive bias in spring–summer during the melt season and a negative bias in autumn–winter during the growth period. The extent to which GIOPS forecasts reduce these biases is thus an indication of seasonally dependent forecast errors. In both hemispheres, GIOPS biases are quite small compared to persistence, with ice formation in autumn underestimated somewhat. Summer melt in the Southern Hemisphere also appears underestimated. The most significant bias appears in the Northern Hemisphere in winter, where GIOPS overestimates ice concentrations in the Greenland Sea.

Spatial maps of RMSEs at a lead time of 168 h are shown for the Northern (Figure 4) and Southern (Figure 5) Hemispheres. Along all of the major marginal ice zones, GIOPS shows a reduction in forecast error as compared to persistence. For the Arctic, the greatest gains are found in the Bering Sea, the Labrador Sea/Baffin Bay and the Barents Sea. The most challenging region is in the Greenland Sea where GIOPS appears to have little skill. This error may be due to a lack of wave–ice interactions in GIOPS that are known to be important in this region (Williams *et al.*, 2013). For the Southern Hemisphere, GIOPS shows reduced errors all around Antarctica with the largest errors occurring in the South Pacific Ocean.

4.2. *Contingency table analysis*

A shortcoming of the spatially targeted analysis ice verification method is that it does not evaluate locations where the modelled ice concentration changed but the analysed concentration did not. As such, incorrect formation of coastal polynyas, or overextension

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

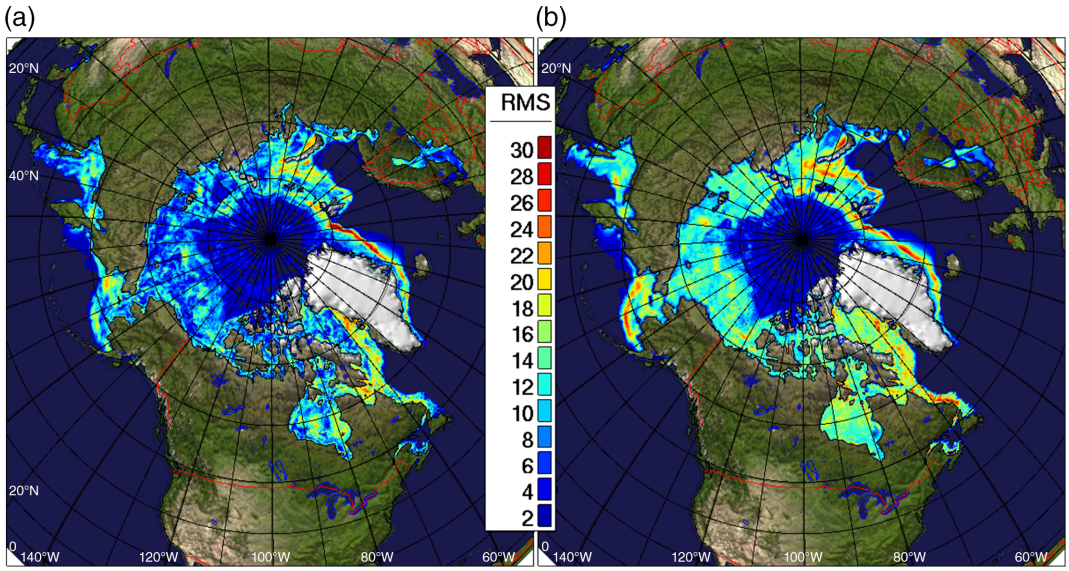

Figure 4. Same as Figure 3 except showing the spatial distribution of RMS forecast errors for (a) GIOPS and (b) persistence, over the Northern Hemisphere.

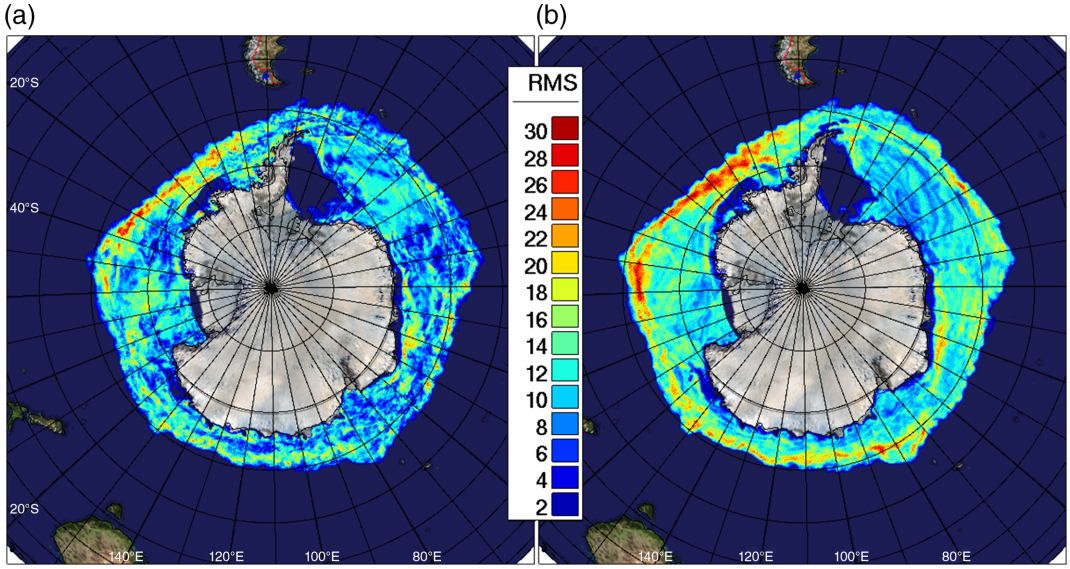

Figure 5. Same as Figure 4 except for the Southern Hemisphere.

of the ice edge (false alarms) are not fully accounted for. Evaluation against IMS analyses using a contingency table approach allows for a more complete examination of the ice edge position. The disadvantage of this method is that a 0.4 cutoff is applied to the ice concentration to determine the binary value of ice/water. Thus small changes in concentration can be missed or excessively penalized depending on where the 0.4 isoline lies. Additionally, since both methods verify against analyses, the analysis error (3D-Var or IMS) itself is not considered.

The misses and false alarms for both GIOPS 168 h forecasts and persistence are shown in Figures 6 and 7. The colours indicate the number of counts for a given $1^\circ \times 1^\circ$ grid box, with warmer colours indicating greater error. Overall, Figures 6 and 7 show similar errors for forecasts and persistence. The high number of misses (Figure 6) along the predominant marginal ice zones (Labrador Sea, east of Greenland, Bering Sea, Sea of Okhotsk) is associated with a bias toward open water in the 3D-Var analyses and forecasts in spring and summer. The greater number of misses than false alarms is suggestive of a bias of forecasts and persistence to underestimate the ice extent. It may also be related to systematic differences between 3D-Var and IMS ice analyses (discussed further below).

To more clearly highlight areas of GIOPS forecast skill, the differences in PCI and PCW of forecasts minus persistence are shown in Figure 8. Both PCI and PCW are in the range [0,1], where 0 and 1 correspond to no skill and perfect skill, respectively. Misses will result in lower PCI and false alarms reduce the value of PCW. Areas of both higher and lower scores for PCI and PCW can be seen in Figure 8. The significant increases in PCI for forecasts (Figure 8, left panel) are mostly related to the skill of the system in predicting ice formation in autumn and winter. Similarly, the broad areas of increased PCW (Figure 8(b)) reflect the skill of the system during the melt season (i.e. a greater number of false alarms for persistence). The areas of lower PCI (higher number of forecast misses represented as blue areas in Figure 8(a)) are quite sensitive to details of the SST assimilation along the ice edge. In a series of sensitivity experiments (not shown), increasing the rejection of SST data along the ice edge was found to result in an overall cooling of SSTs and an increase in ice formation, nearly eliminating these misses, albeit with much larger corresponding increases in false alarms.

Clearly, improving either one of the PCI or PCW scores is fairly straightforward, as one needs only to cover the entire domain with ice to achieve PCI = 1; however, this would result in low values of PCW and vice versa. For this reason, it is often more insightful to refer to the PCT. Figure 9 shows the difference in PCT between GIOPS forecasts and persistence at a lead time of 168 h. We can see that over many regions the improvements

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

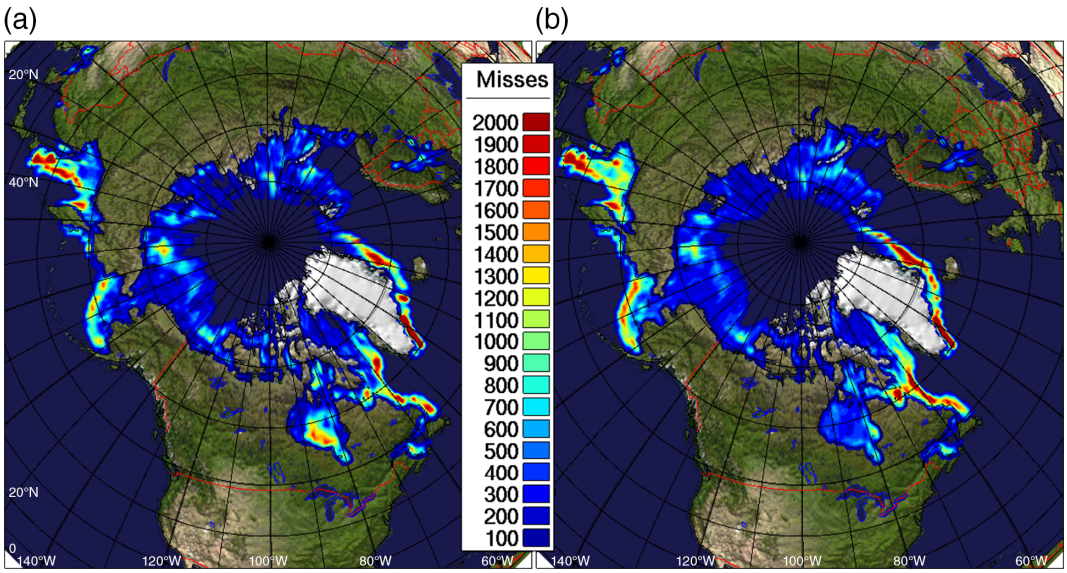

Figure 6. Spatial maps of sea ice concentration forecast error as compared to IMS analyses at a lead time of 168 h using a contingency table approach. Misses are shown for (a) GIOPS and (b) persistence, for weekly forecasts over 2011.

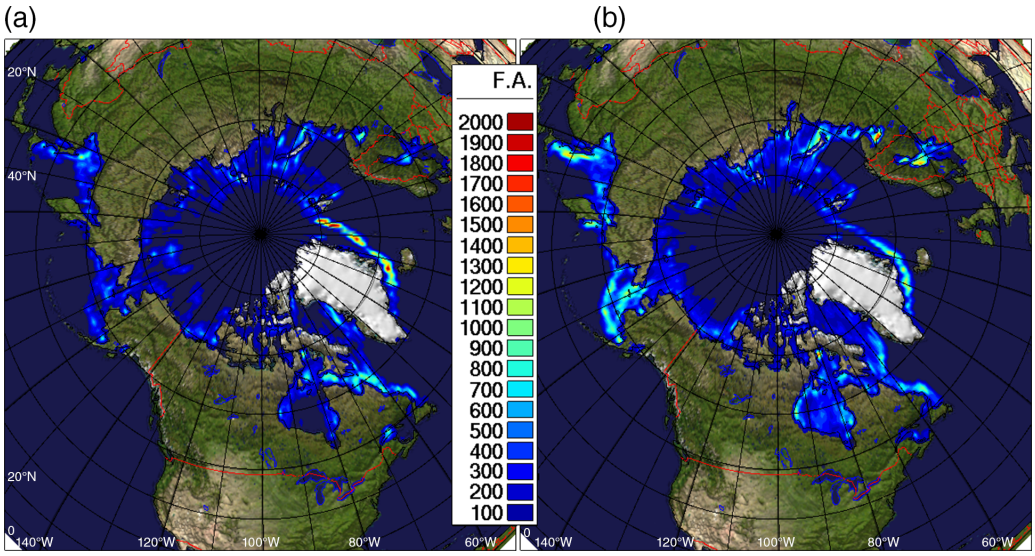

Figure 7. Same as Figure 6 except for false alarms.

in PCW or PCI outweigh the errors providing higher values of PCT. However, some localized areas of lower PCT are present, in particular in the Greenland Sea, north of the Chukchi Sea and along certain coastal areas, such as within the Canadian Arctic Archipelago and Hudson Bay. The lower PCT values are due mostly to an excessive melt in spring and summer.

Figure 10 shows the time series of PCT and Frequency Bias for 168 h GIOPS forecasts and persistence (left column). Here we can see that for most of the year GIOPS forecasts provide higher values of PCT than persistence. However, from July to September, forecasts exhibit a significant open-water bias resulting in poor forecast skill. This is due partially to a bias in the 3D-Var analyses (Buehner *et al.*, 2015) related to the difficulty of assimilating passive microwave observations when significant surface melt may be interpreted as open water.

It should also be noted that the errors shown here are quite sensitive to the threshold used in the determination of binary values of ice or open water. The right column in Figure 10 shows the PCT and Frequency Bias scores determined using a threshold of 0.2 (instead of the usual 0.4). This value was chosen as it is used in the assimilation scheme to obtain proxy values of freezing temperature from the ice analysis (section 2.2) and thus more closely corresponds to the ice/open-water boundary in the model. Use of the lower threshold nearly eliminates the bias toward open-water values in summer, resulting in equivalent or higher PCT values than persistence throughout the year. This sensitivity demonstrates that the areas of ice cover in summer with concentrations between 0.2 and 0.4 can have a large impact on skill scores. Indeed, it is during summer that these areas are largest (not shown). The analyses in these areas may also suffer from high error due to biases in passive microwave retrievals. Given that the threshold of 0.4 is applied subjectively in the production of IMS analyses, there is some uncertainty regarding the most appropriate threshold to use for forecast verification.

5. Discussion and conclusions

Here we provide an objective assessment of sea ice forecast skill in a new system, called GIOPS, implemented recently at the CMC. GIOPS provides daily global ice and ocean analyses and 10-day forecasts at 0000 UTC on a 1/4° resolution grid. GIOPS includes a multivariate ocean data assimilation system that combines satellite observations of SLA and SST together with *in situ* observations of temperature and salinity. Ice analyses are produced using a 3D-Var method that assimilates satellite observations from SSM/I and SSMIS, together with manual analyses from the Canadian Ice Service.

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

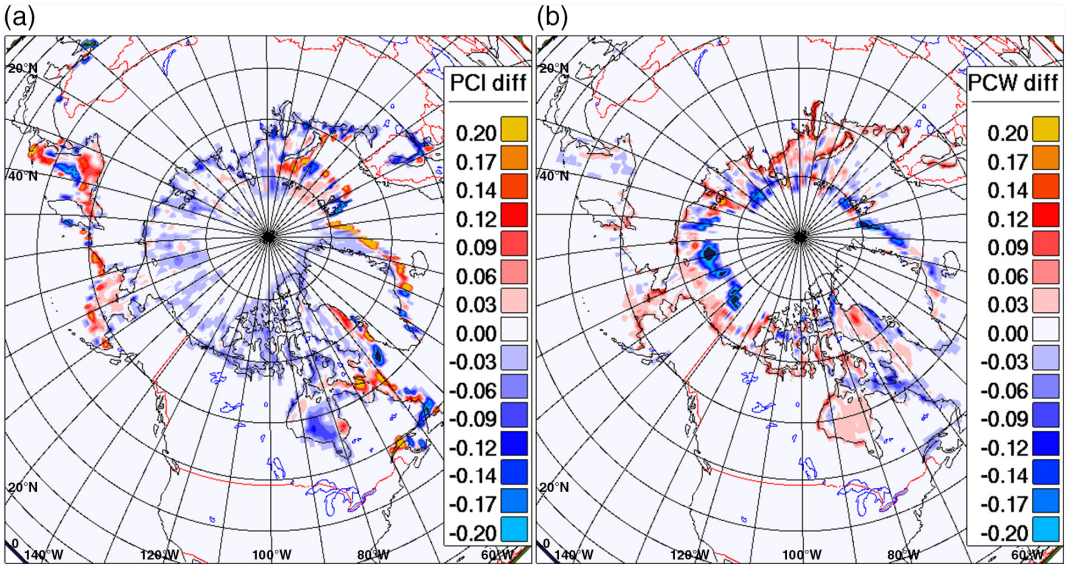

Figure 8. Spatial maps of sea ice concentration forecast error as compared to IMS analyses at a lead time of 168 h using a contingency table approach. Differences between GIOPS and persistence in (a) Proportion Correct Ice and (b) Proportion Correct Water are shown for weekly forecasts over 2011. Warm colours represent an improvement of GIOPS forecasts with respect to persistence.

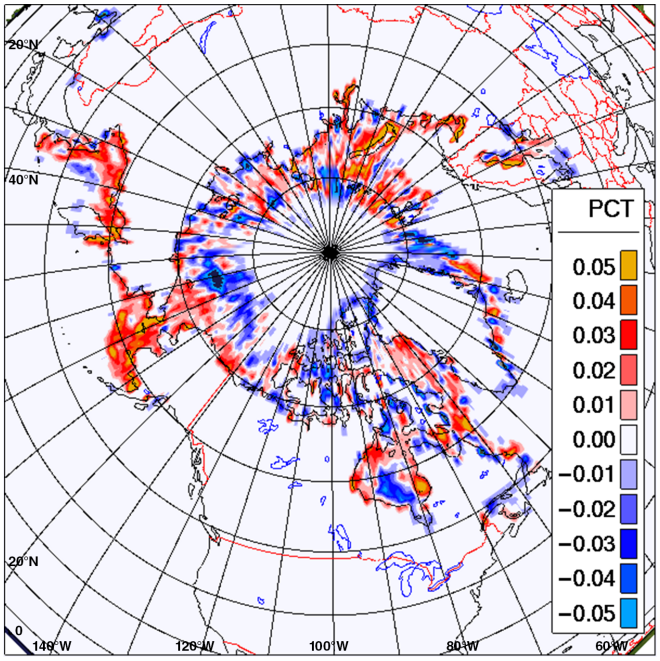

Figure 9. Same as Figure 8 except showing differences between GIOPS and persistence in Proportion Correct Total for weekly forecasts over 2011. Warm colours represent an improvement of GIOPS forecasts with respect to persistence.

Analyses of total ice concentration are projected onto the ten thickness categories used in the ice model using a new method whereby spatially and temporally varying increments to the thickness categories are derived by rescaling model forecast tendencies (RFT; Eq. (2)). The RFT method is compared to a simpler method (which simply rescales the existing distribution; Eq. (1)) as the choice of method is found to have a non-negligible effect on total ice volume for both hemispheres. The principal advantage of the RFT method is that it can directly modulate (and reverse) nonlinear processes such as ice formation, melt and deformation, although some residual impacts on total ice volume remain.

An alternative method proposed by Lindsay and Zhang (2006) is to modify the total ice concentration by adding or removing ice from the thinnest categories. This has the advantage that it minimizes changes to the total ice mass. However, it relies on a specified value for new ice thickness and neglects the nonlinear impact of many processes on the evolution of ice thickness distribution. While a detailed evaluation of different methods is beyond the scope of this article, it may provide valuable insights into how to treat this challenging issue.

Accurately representing the distribution of ice thicknesses at a grid cell is important both in terms of its effect on ice concentration forecasts and for forecasting possible areas of severe internal pressure, which are of prime concern for marine safety. Improvements in this area require the addition of the direct assimilation of ice thickness data. Despite detailed surveys from missions such as IceBridge (Kurtz *et al.*, 2013) that provide valuable information about the ice thickness and snow distribution in specific locations, more real-time data are required to meet operational needs. To this end, improved estimates are beginning to become available based on Cryosat2 data (Laxon *et al.*, 2013) and other satellite retrievals (e.g. thickness derived from Advanced Very High Resolution Radiometer (AVHRR): Wang *et al.*, 2010), although further sources are required in order to meet operational needs for the real-time delivery of high-quality observations with a small spatial footprint.

An objective verification of sea ice forecasts was presented using two methods: analysis-based error assessment using spatially targeted sub-sampling, and a contingency table approach to evaluate ice extent as compared to an independent analysis. Together these methods demonstrate a consistent picture of skilful medium-range forecasts in both the Northern and Southern Hemispheres as compared to persistence. Forecast error estimated using the analysis-based verification method is smaller than persistence error at all lead times with rapid growth over the first few days followed by slower growth thereafter. The rapid growth at short lead times highlights the importance of the observation cut-off time used in the sea ice data assimilation system. Here, a 6 h assimilation window is used centred at the analysis time. As such, at the expense of a greater delay in the production time, the analysis benefits from observations at and just following the validity time. While this compromise is well known in the NWP community, it is often overlooked in operational oceanographic systems due to the longer time-scales of variability and the relatively large delays present in ocean observing systems (especially *in situ* temperature and salinity profiles). Clearly, any evaluation of sea ice forecasts should take this into consideration.

For both verification methods, significant forecast skill is found along the predominant marginal ice zones. Areas of particular skill in the Northern Hemisphere are shown to be present in the Labrador Sea, Baffin Bay, the Bering Sea and the Barents Sea. These areas tend to have more direct *in situ* observations of

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

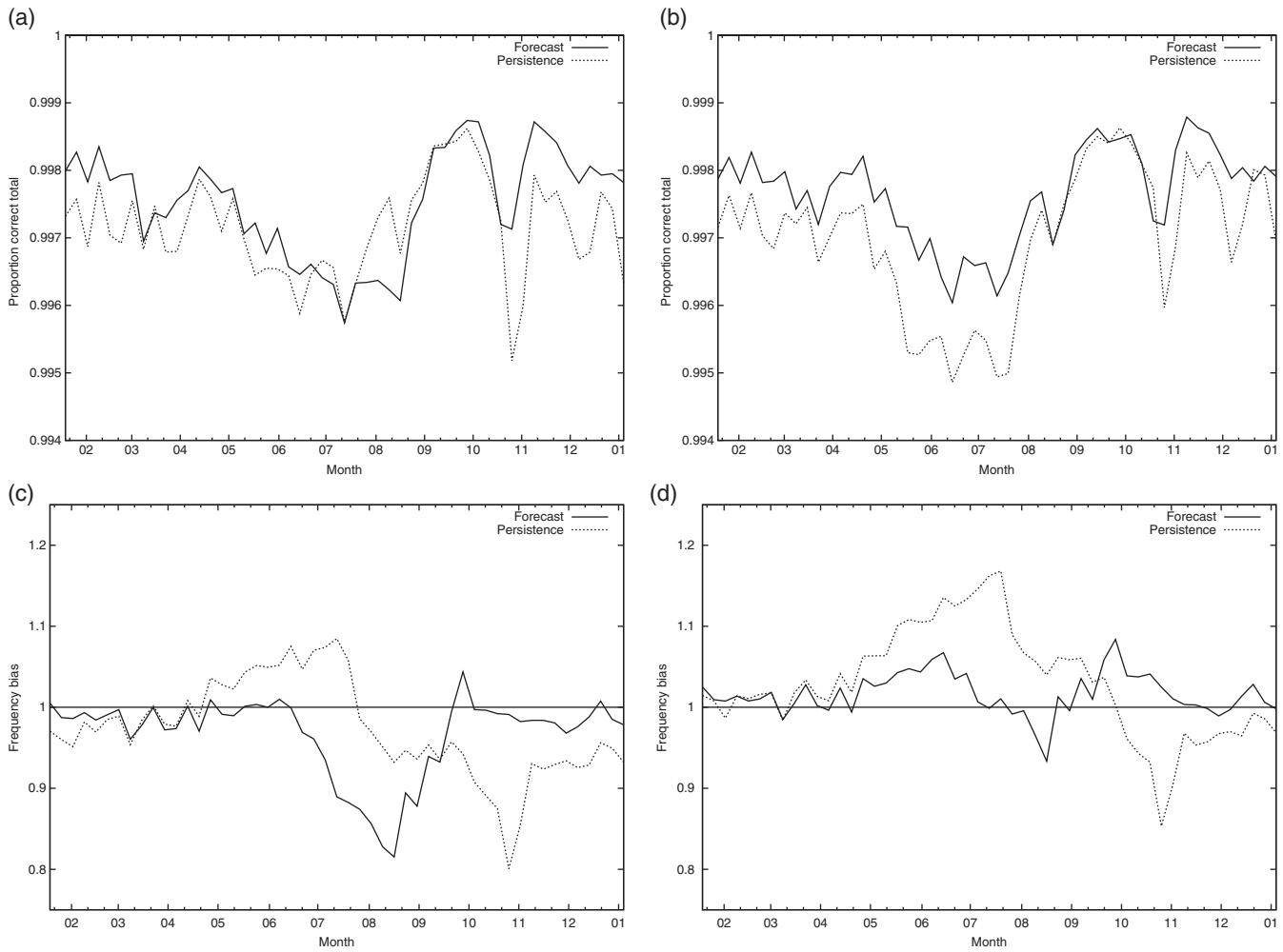

Figure 10. Evaluation of Northern Hemisphere sea ice concentration at a lead time of 168 h as compared to IMS analyses. The seasonality of Proportion Correct Total (a,b) and Frequency Bias (c,d) errors are shown throughout the year based on weekly forecasts produced for 2011. Values in the left column (a,c) were produced using an ice/open water threshold of 0.4, whereas values in the right column (b,d) used 0.2. GIOPS forecasts are shown as a solid line and the persistence of initial 3D-Var analyses as a dotted line.

water mass properties and thus are expected to benefit from more accurate mixed-layer depths. Conversely, in the central Arctic, the Russian Shelf Seas and in the Canadian Arctic Archipelago, fewer observations lead to less well constrained water mass properties and a greater likelihood of errors in the mixed-layer depth. As such, an extension of the oceanographic observing system (e.g. the Argo array) to include greater coverage of polar regions may lead to improved forecasting skill in these areas. The forecasting skill of GIOPS in the Russian Shelf Seas and the Canadian Arctic Archipelago is also limited by the lack of a landfast ice parametrization in the ice model.

As compared to the IMS analyses, GIOPS forecasts exhibit a significant open water bias in summer. These errors may be partially attributed to an equivalent bias in the GIOPS ice analyses (not shown) caused by surface melt in the passive microwave observations. However, lowering the threshold for determining binary values of ice or open water from 0.4 to 0.2 nearly eliminates the open-water bias in both forecasts and analyses in summer. Given the challenges of interpreting passive microwave observations during periods of melt, and the use of the 0.4 threshold subjectively in the production of IMS analyses, caution should be employed when interpreting forecast verification scores as compared to IMS analyses in summer.

Comparison of error in GIOPS forecast and trial fields show that an important improvement can be obtained by forcing the ice–ocean model with more accurate atmospheric fields. In the Northern Hemisphere, a reduction of RMSE with respect to persistence was found for forecasts and trial fields to be 13 and 17% respectively at a lead time of 168 h. Similarly, the RMSE was reduced by 20 and 24% respectively in the Southern Hemisphere. This additional reduction in error for trial fields highlights the sensitivity of ice forecast skill to the surface fluxes applied to the ice–ocean model. In particular, the trial fields benefit from a better consistency between the sea ice concentrations used to calculate fluxes in the atmospheric model. These results provide a first estimate for the potential impact of coupling with NWP systems on sea ice forecast skill. Additional improvements beyond this may also be possible by coupling directly to an atmospheric model, as feedbacks between the atmosphere, ice and ocean may improve the representation of many boundary-layer processes.

Overall, the GIOPS sea ice verifications presented here demonstrate a consistent picture of skilful medium-range forecasts in both the Northern and Southern Hemispheres as compared to persistence. While these results help to support the use of GIOPS forecasts for coupled environmental prediction and marine applications, they should nonetheless be viewed as a baseline for further improvement. A particular weakness of GIOPS is the use of three separate analysis systems for the ocean (SAM2), ice (3D-Var) and SST (Brasnett, 2008). Coupling these systems together should lead to better consistency and the potential to further exploit ice–ocean covariances to improve GIOPS initial conditions, and as a result, sea ice forecasts. The use of additional data sources to constrain GIOPS analyses should also improve forecasts, in particular, sea ice thickness, drift and water mass properties near and under the ice. However, the development of these methods, as well as further model improvements, would benefit from more sophisticated verification techniques.

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Sea ice verification is hindered by a number of key challenges. These include limitations in the resolution and accuracy of satellite retrievals of ice concentration, together with the inherently small scale of most sea ice features. Clearly, this is a challenging area requiring further development of more refined and advanced techniques. As there is currently no panacea for the issues involved in sea ice verification, our philosophy is to combine complementary methods to provide robustness. The use of an analysis-based error assessment using spatially targeted sub-sampling allows a detailed evaluation of forecast skill along the ice edge and permits the identification of sources of error. A comparison with IMS analyses using a contingency table approach provides a general assessment with respect to an independent data source, but neglects errors in the IMS analysis itself and shows a significant sensitivity to the choice of ice concentration threshold.

This diversified approach could be further improved by the addition of other complementary verification methods. Comparison with ice drift, either from ice buoys or satellite estimates, can be quite valuable in improving ice deformations as well as the transfer of momentum between the simulated atmosphere, ice and ocean. Spatial verification methods (Gilleland *et al.*, 2009) may also be quite useful, given the strong role of advection in ice forecasting. Given the importance of ice forecasts for marine transportation, an improvement in the usefulness of forecasting products may be possible if more user-based metrics were used, such as the ice edge position and displacement, and internal ice pressure.

Acknowledgements

The development and implementation of GIOPS was made possible by the excellent operational and technical support by the staff at Environment Canada. In particular, the authors would like to recognize the important contributions of B. Brasnett and the CMC implementations group. The authors would like to thank R. Bourdallé-Badie, O. Le Galloudec and C. Régnier for helpful discussions as well as E. Hunke and an anonymous reviewer for their helpful comments. We would also like to thank Collecte Localisation Spatiale for access to the AVISO products, with special thanks to G. Larnicol and S. Guinehut for access to the *in situ* profile data. GIOPS also benefits from the contributions of M. Ouellet and D. Gilbert for their tireless efforts to maintain the Canadian *in situ* data feed and Argo Program. Observations from the Argo array were collected and made freely available by the International Argo Program and the national programmes that contribute to it (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System.

References

Adcroft A, Hill C, Marshall J. 1997. Representation of topography by shaved cells in a height coordinate ocean model. *Mon. Weather Rev.* **125**: 2293–2315.

Barnier B, the DRAKKAR Group. 2007. Eddy-permitting ocean circulation hindcasts of past decades. *Clivar Exchanges* **12**: 8–10.

Barnier B, Madec G, Penduff T, Molines J-M, Treguier A-M, Le Sommer J, Beckmann A, Biastoch A, Böning C, Dengg J, Derval C, Durand E, Gulev S, Remy E, Talandier C, Theeten S, Maltrud M, McClean J, De Cuevas B. 2006. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dyn.* **56**: 543–567, doi: 10.1007/s10236-006-0082-1.

Bertino L, Lisæter KA. 2008. The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans. *J. Oper. Oceanogr.* **1**: 15–18.

Blanke B, Delecluse P. 1993. Variability of the tropical Atlantic Ocean simulated by a general circulation model with two different mixed-layer physics. *J. Phys. Oceanogr.* **23**: 1363–1388.

Brasnett B. 2008. The impact of satellite retrievals in a global sea-surface-temperature analysis. *Q. J. R. Meteorol. Soc.* **134**: 1745–1760.

Buehner M, Caya A, Pogson L, Carrieres T, Pestieau P. 2013a. A new Environment Canada regional ice analysis system. *Atmos. Ocean* **51**: 18–34.

Buehner M, Caya A, Carrieres T, Pogson L, Lajoie M. 2013b. Overview of sea ice data assimilation activities at Environment Canada. *Proceedings of the ECMWF-WWRP/THORPEX Workshop on Polar Prediction*, 24-27 June 2013. ECMWF: Reading, UK, 10 pp.

Buehner M, Caya A, Carrieres T, Pogson L. 2015. Assimilation of SSMIS and ASCAT data and the replacement of highly uncertain estimates in the Environment Canada Regional Ice Prediction System. *Q. J. R. Meteorol. Soc.*, doi: 10.1002/qj.2408.

Charron M, Polavarapu S, Buehner M, Vaillancourt PA, Charette C, Roch M, Morneau J, Garand L, Aparicio JM, MacPherson S, Pellerin S, St-James J, Heilliette S. 2012. The stratospheric extension of the Canadian global deterministic medium-range weather forecasting system and its impact on tropospheric forecasts. *Mon. Weather Rev.* **140**: 1924–1944.

Cummings J, Bertino L, Brasseur P, Fukumori I, Kamachi M, Martin MJ, Mogensen K, Oke P, Testut CE, Verron J, Weaver A. 2009. Ocean data assimilation systems for GODAE. *Oceanography* **22**: 96–109.

Gilleland E, Ahijevych D, Brown BG, Casati B, Ebert EE. 2009. Intercomparison of spatial forecast verification methods. *Weather and Forecasting* **24**: 1416–1430.

Gould WJ. 2005. From swallow floats to Argo – The development of neutrally buoyant floats. *Deep Sea Res. Part II* **52**: 529–543.

Helfrich SR, McNamara D, Ramsay BH, Baldwin T, Kasheta T. 2007. Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). *Hydrol. Process.* **21**: 1576–1586.

Hernandez F, Bertino L, Brassington G, Chassignet E, Cummings J, Davidson F, Drévillon M, Garric G, Kamachi M, Lellouche J-M, Mahdon R, Martin MJ, Ratsimandresy A, Regnier C. 2009. Validation and intercomparison studies within GODAE. *Oceanography* **22**: 128–143.

Hibler WD. 1979. A dynamic thermodynamic sea ice model. *J. Phys. Oceanogr.* **9**: 815–846.

Hunke EC. 2001. Viscous-plastic sea ice dynamics with the EVP model: linearization issues. *J. Comput. Phys.* **170**: 18–38.

Hunke EC, Lipscomb WH. 2010. 'CICE: the Los Alamos sea ice model documentation and software user's manual version 4.1'. Technical Report LA-CC-06-012. Los Alamos National Laboratory: Los Alamos, NM.

Kurtz NT, Farrell SL, Studinger M, Galin N, Harbeck JP, Lindsay R, Onana VD, Panzer B, Sonntag JG. 2013. Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data. *Cryosphere* **7**: 1035–1056.

Large WG, Yeager SG. 2004. 'Diurnal to decadal global forcing for ocean and seaice models', NCAR Technical Note NCAR/TN-4601STR. NCAR: Boulder, CO.

Laxon SW, Giles KA, Ridout AL, Wingham DJ, Willatt R, Cullen R, Kwok R, Schweiger A, Zhang JL, Haas C, Hendricks S, Krishfield R, Kurtz N, Farrell S, Davidson M. 2013. CryoSat-2 estimates of Arctic sea ice thickness and volume. *Geophys. Res. Lett.* **40**: 732–737, doi: 10.1002/grl.50193.

Lellouche J-M, Le Galloudec O, Drévillon M, Régnier C, Greiner E, Garric G, Ferry N, Desportes C, Testut C-E, Bricaud C, Bourdallé-Badie R, Tranchant B, Benkiran M, Drillet Y, Daudin A, De Nicola C. 2013. Evaluation of global monitoring and forecasting systems at Mercator Océan. *Ocean Sci.* **9**: 57–81, doi: 10.5194/os-9-57-2013.

Lemieux J-F, Beaudoin C, Dupont F, Roy F, Smith GC, Shlyaeva A, Buehner M, Caya A, Chen J, Carrieres T, Pogson L, DeRepentigny P, Plante A, Pestieau P, Pellerin P, Ritchie H, Garric G, Ferry N. 2014. The Regional Ice Prediction System (RIPS): Verification of forecast sea ice concentration. *Q. J. R. Meteorol. Soc.* , doi: 10.1002/qj.2526.

Lindsay RW, Zhang J. 2006. Assimilation of ice concentration in an ice–ocean model. *J. Atmos. Oceanic Technol.* **23**: 742–749, doi: 10.1175/JTECH1871.1.

Lipscomb WH, Hunke EC, Maslowski W, Jakacki J. 2007. Ridging, strength, and stability in high-resolution sea ice models. *J. Geophys. Res.* **112**: C03S91, doi: 10.1029/2005JC003355.

Madec G, Delecluse P, Imbard M, Levy C. 1998. 'OPA 8.1 general circulation model reference manual'. Notes de l'IPSL, University P. et M. Curie, B102 T15-E5, Paris, No. 11.

Madec G. 2008. 'NEMO reference manual, ocean dynamics component : NEMO-OPA. Preliminary version'. Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No. 27, pp. 1288–1619.

Markus T, Cavalieri DJ. 2000. An enhancement of the NASA team sea ice algorithm. *IEEE Trans. Geosci. Remote Sensing* **38**: 1387–1398.

Martin M, Dash P, Ignatov A, Banzon V, Beggs H, Brasnett B, Cayula J-F, Cummings J, Donlon C, Gentemann C, Grumbine R, Ishizaki S, Maturi E, Reynolds RW, Roberts-Jones J. 2012. Group for High Resolution Sea Surface Temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE). *Deep-Sea Res. II*: **77–80**: 21–30, doi: 10.1016/j.dsr2.2012.04.013.

Melsom A, Simonsen M, Bertino L. 2011. 'MyOcean Project Scientific Validation Report (ScVR) for V1 real-time forecasts'. MyOcean Technical Report MYO-WP05-ScCV-RT-ARC-METNO-OSLO-NO. Met.no, Oslo.

Pellerin P, Ritchie H, Saucier SJ, Roy F, Desjardins S, Valin M, Lee V. 2004. Impact of a two-way coupling between an atmospheric and an ocean–ice model over the Gulf of St. Lawrence. *Mon. Weather Rev.* **132**: 1379–1398.

Penduff T, Le Sommer J, Barnier B, Treguier A-M, Molines J-M, Madec G. 2007. Influence of numerical schemes on current–topography interactions in 1/4° global ocean simulations. *Ocean Sci.* **3**: 509–524.

Penduff T, Juza M, Brodeau L, Smith GC, Barnier B, Molines JM, Treguier AM, Madec G. 2010. Impact of global ocean model resolution on sea-level variability with emphasis on interannual time scales. *Ocean Sci.* **6**: 269–284.

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Pham DT, Verron J, Roubaud MC. 1998. A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.* **16**: 323–340.

Posey PG, Metzger EJ, Wallcraft AJ, Smedstad OM, Phelps MW. 2010. 'Validation of the 1/12° Arctic Cap Nowcast/Forecast System (ACNFS).' Naval Report NRL/MR/7320-10-9287. Stennis Space Center: Hancock County, MS.

Ramsay BH. 1998. The interactive multisensory snow and ice mapping system. *Hydrol. Processes* **12**: 1537–1546.

Rio M-H, Hernandez F. 2004. A mean dynamic topography computed over the world ocean from altimetry, *in situ* measurements and a geoid model. *J. Geophys. Res.* **109**: C12032, doi: 10.1029/2003JC002226.

Rothrock DA. 1975. The energetics of the plastic deformation of pack ice by ridging. *J. Geophys. Res.* **80**: 4514–4519, doi: 10.1029/JC080i033p04501.

Roullet G, Madec G. 2000. Salt conservation, free surface, and varying levels: a new formulation for ocean general circulation models. *J. Geophys. Res.* **105**: 23927–23942, doi: 10.1029/2000JC900089.

Sakov P, Counillon F, Bertino L, Lisæter KA, Oke PR, Korablev A. 2012. TOPAZ4: an ocean–sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci.* **8**: 633–656.

Smith GC, Roy F, Brasnett B. 2013a. Evaluation of an operational ice–ocean analysis and forecasting system for the Gulf of St Lawrence. *Q. J. R. Meteorol. Soc.* **139**: 419–433, doi: 10.1002/qj.1982.

Smith GC, Roy F, Belanger J-M, Dupont F, Lemieux J-F, Beaudoin C, Pellerin P, Lu Y, Davidson F, Ritchie H. 2013b. Small-scale ice–ocean–wave processes and their impact on coupled environmental polar prediction, *Proceedings of the ECMWF-WWRP/THORPEX Polar Prediction Workshop*, 24-27 June 2013. ECMWF: Reading, UK.

Smith GC, Roy F, Mann P, Dupont F, Brasnett B, Lemieux J-F, Laroche S, Bélair S. 2014. A new atmospheric dataset for forcing ice–ocean models: evaluation of reforecasts using the Canadian global deterministic prediction system. *Q. J. R. Meteorol. Soc.* **140**: 881–894, doi: 10.1002/qj.2194.

Steele M. 1992. Sea ice melting and floe geometry in a simple ice–ocean model. *J. Geophys. Res.* **97**: 17729–17738, doi: 10.1029/92JC01755.

Tranchant B, Testut CE, Renault L, Ferry N, Birol F, Brasseur P. 2008. Expected impact of the future SMOS and Aquarius Ocean surface salinity missions in the Mercator Océan operational systems: New perspectives to monitor ocean circulation. *Remote Sensing Environ.* **112**: 1476–1487.

Van Woert M, Zou C, Meier W, Hovey P, Preller R, Posey P. 2004. Forecast verification of the Polar Ice Prediction System (PIPS) sea ice concentration fields. *J. Atmos. Oceanic Technol.* **21**: 944–957.

Wang XJ, Key JR, Liu YH. 2010. A thermodynamic model for estimating sea and lake ice thickness with optical satellite data. *J. Geophys. Res.* **115**: C12035, doi: 10.1029/2009JC005857.

Williams TD, Bennetts LG, Squire VA, Dumont D, Bertino L. 2013. Wave–ice interactions in the marginal ice zone. Part 2: Numerical implementation and sensitivity studies along 1D transects of the ocean surface. *Ocean Modell.* **71**: 92–101, doi: 10.1016/j.ocemod.2013.05.011.

© 2015 Her Majesty the Queen in Right of Canada. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.