Article

An Evaluation of the Performance of Sea Ice Thickness Forecasts to Support Arctic Marine Transport

Tarkan Aslan Bilge [1,*], Nicolas Fournier [2], Davi Mignac [2], Laura Hume-Wright [2], Laurent Bertino [3], Timothy Williams [3] and Steffen Tietsche [4]

1   Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, 5007 Bergen, Norway
2   Met Office, Exeter EX1 3PB, UK; nicolas.fournier@metoffice.gov.uk (N.F.);
    davi.carneiro@metoffice.gov.uk (D.M.); laura.hume-wright@metoffice.gov.uk (L.H.-W.)
3   Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research,
    5007 Bergen, Norway; laurent.bertino@nersc.no (L.B.); timothy.williams@nersc.no (T.W.)
4   European Centre for Medium-Range Weather Forecasts, Reading RG2 9AX, UK; steffen.tietsche@ecmwf.int
*   Correspondence: tarkan.bilge@uib.no

Abstract: In response to declining sea ice cover, human activity in the Arctic is increasing, with access to the Arctic Ocean becoming more important for socio-economic reasons. Accurate knowledge of sea ice conditions is therefore becoming increasingly important for reducing the risk and operational cost of human activities in the Arctic. Satellite-based sea ice charting is routinely used for tactical ice management, but the marine sector does not yet make optimal use of sea ice thickness (SIT) or sea ice concentration (SIC) forecasts on weekly timescales. This is because forecasts have not achieved sufficient accuracy, verification and resolution to be used in situations where maritime safety is paramount, and assessing the suitability of forecasts can be difficult because they are often not available in the appropriate format. In this paper, existing SIT forecasts currently available on the Copernicus Marine Service (CMS) or elsewhere in the public domain are evaluated for the first time. These include the seven-day forecasts from the UK Met Office, MET Norway, the Nansen Environmental and Remote Sensing Center (NERSC) and the European Centre for Medium-Range Weather Forecasts (ECMWF). Their forecast skills were assessed against unique in situ data from five moorings deployed between 2016 and 2019 by the Barents Sea Metocean and Ice Network (BASMIN) and Barents Sea Exploration Collaboration (BaSEC) Joint Industry Projects. Assessing these models highlights the importance of data assimilation in short-term forecasting of SIT and suggests that improved assimilation of sea ice data could increase the utility of forecasts for navigational purposes. This study also demonstrates that forecasts can achieve similar or improved correlation with observations when compared to a persistence model at a lead time of seven days, providing evidence that, when used in conjunction with sea ice charts, SIT forecasts could provide valuable information on future sea ice conditions.

Keywords: shipping; sea ice thickness; forecast; Copernicus; validation; Arctic; Barents Sea; mooring

1. Introduction

In response to declining sea ice cover, human activity in the Arctic is increasing, with access to the Arctic Ocean becoming more important for socio-economic purposes. These purposes include commercial ventures such as tourism, fishing, shipping, mineral and oil extraction, along with activities of importance to local communities such as subsistence hunting, search and rescue and community resupply.

Furthermore, with the International Maritime Organization (IMO) Polar Code [1] and the ongoing construction of new ice class ships, increased activity in all economic sectors is expected [2]. The Polar Code now mandates that 'ships shall have the ability to receive up-to-date information including ice information for safe navigation' and requires a risk assessment methodology, Polar Operational Limit Assessment Risk Indexing System

(POLARIS), to determine the limitations for ice operations [1]. Ships will thus require more detailed sea ice and weather information, but navigating ice-covered areas requires both broad and precise knowledge of ice conditions depending on the region and type of activity.

The main Arctic transportation routes include the Northern Sea Route (NSR), the Canadian Archipelago Waters, which include the Northwest Passage (NWP), and the coasts of Svalbard and Greenland. The Arctic Bridge links the European Arctic to Canada and NSR to the Pacific. Resource extraction occurs in the Barents and Beaufort Seas and the Russian Pechora and Kara Seas. The Nordic (Barents, Norwegian, Greenland and Icelandic) Seas and Bering Sea are key fisheries and routes for passenger vessels. Operations in these regions begin as seasonal ice retreats in the summer [3]. Most simulations in the Coupled Model Intercomparison Project 6 (CMIP6) suggest that September sea ice area will fall below 1,000,000 km^2 for the first time before 2050 under four different emissions scenarios [4]. As Arctic sea ice continues to decline, it is expected that a Transpolar Sea Route across the North Pole will become feasible for shipping, and the length of the shipping season will be extended for NSR and NWP [5].

Accurate forecasts of Arctic sea ice are thus becoming increasingly important for the safety of human activities in the Arctic, especially for shipping. Improved knowledge of sea ice conditions from short to seasonal timescales would allow for better planning, reducing the level of risk and operational costs of navigating in the Arctic. Whilst satellite-based sea ice charting is routinely used for tactical ice management, the marine sector does not yet make use of existing operational sea ice thickness (SIT) or sea ice concentration (SIC) forecasting products, such as those now freely available from the Copernicus Marine Service (CMS). Arctic asset managers and vessels' crews are generally not aware of such products, or these products have so far suffered from insufficient accuracy, verification, resolution or are not available in an adequate format to be well integrated within the existing decision-making processes and systems [3,6,7].

Tietsche et al. (2020) provide a good overview of existing forecast model types and limitations [7]. This summary illustrates that current-generation numerical forecasting systems are able to provide regional or global forecasts of basic sea ice parameters with a spatial resolution of up to 3 km, with up to hourly temporal resolution, and with an update frequency of up to several times a day. The systems broadly fall into three categories according to their lead time. Short- and medium-range (SMR) forecasts have a lead time of up to two weeks and are usually of higher resolution and are updated at least once a day. Extended-range (ER) forecasts try to predict sub-seasonal anomalies occurring a few weeks ahead, are usually updated every few days and have a coarser resolution. Long-range or seasonal (LR) forecasts aim at providing guidance on seasonal anomalies up to a year ahead but are generally low resolution and are usually only updated once a month. CMS is currently providing several short- and medium-range (SMR) forecasts (see Table 1). The Copernicus Climate Change Service (C3S) is providing a multi-model seasonal (LR) forecast. However, extended-range (ER) forecasts are not provided by either CMS or C3S.

Current physics-based forecast models provide at least the following basic sea ice properties: SIC, SIT, sea-ice drift and thickness of snow on sea ice. Forecast quality varies considerably depending on the geographical region, time of the year, forecast lead time, spatial and temporal resolution and the sea ice property of interest. Additionally, forecast quality for the same target can vary equally dramatically between forecasting systems, depending on numerous factors: appropriateness of resolution and coverage, fidelity of numerical representation of physical processes in the forecast model, quality and completeness of forecast initialisation and representation of forecast uncertainty. It is important to note that making statements on the quality of a forecasting system is only possible in a statistical sense. Each forecast case is unique and the forecast error can vary widely between different cases. Therefore, the forecast for each individual case must be used together with robust statistical information about the expected quality derived from many past forecasts that have been verified against the observed outcome [7].

Table 1. Sea-ice forecasting systems available from the Copernicus Marine Service (CMS), Copernicus Climate Change Service (C3S) and European Centre for Medium-Range Weather Forecasts (ECMWF). Forecasts are labelled as either long-range (LR), extended-range (ER) or short- and medium-range (SMR). Three of the models in this table are evaluated later in this study: the Coupled atmosphere–land–ocean–ice Data Assimilation (CPLDA) system, the Towards an Operational Prediction System for the North Atlantic and European Coastal Zones 4 (TOPAZ4) and the Nansen Environmental and Remote Sensing Center (NERSC) neXt generation Sea Ice Model (neXtSIM-F). Table adapted from Tietsche et al. (2020) [7].

Forecasting system	Seasonal multi-system	ENS-extended	CPLDA (GLO-CPL)	GLO-HR	TOPAZ4 (ARC-MFC TOPAZ)	neXtSIM-F (ARC-MFS neXtSIM-F)	BAL-MFC
Provider	C3S	ECMWF	CMS (UK Met Office)	CMS (MERCATOR)	CMS (MET Norway)	CMS (NERSC)	CMS (DMI/BSH) [2]
Forecast range	6 months (LR)	7 weeks (ER)	10 days (SMR)	10 days (SMR)	10 days (SMR)	7 days (SMR)	6 days (SMR)
Components	Sea ice, ocean and atmosphere	Sea ice, ocean and atmosphere	Sea ice, ocean and atmosphere	Sea ice and ocean	Sea ice and ocean	Sea ice	Sea ice and ocean
Spatial coverage	Global	Global	Global	Global	Arctic	Arctic	Baltic
Update frequency	Monthly	Twice-weekly	Daily	Daily	Daily	Daily	Twice-daily
Spatial resolution	1° [1]	36 km	1/4°	1/12°	12.5 km	3 km	2 km

[1] Higher resolution available depending on underlying forecasting system (up to 36 km). [2] Danish Meteorological Institute (DMI)/Federal Maritime and Hydrographic Agency of Germany (BSH).

There is ample research on the quality of numerical forecasts for large-scale sea ice presence at the seasonal time scale. This research is looking for potentially useful correspondence between the forecasts and observations. A consistent finding across this research is that currently operational physics-based forecast models provide skilful seasonal forecasts of large-scale sea ice presence for some regions and seasons [8–12]. For instance, Dirkson, Denis and Merryfield (2019) report that the C3S multi-model seasonal forecast for Arctic sea ice cover during the annual minimum in September is substantially better than simple statistical forecasts derived from past observations [13]. Forecast skill increases as lead time decreases, and a forecast issued by the C3S multi-model system at the beginning of August should be expected to be able to predict September SIC better than a statistical reference forecast throughout the Arctic Ocean domain. However, these studies often focus on targets such as the pan-Arctic sea ice volume or extent, which is of little relevance for most operational users. The predictability is often much lower for more user-relevant targets such as the seasonal time of the opening of the Northern Sea Route [14].

For sub-seasonal time scales (weeks ahead), the forecast skill for current numerical prediction systems extends to smaller scales of 10–100 km [15], although skill is highly dependent on the region and season. Moreover, verification of sea ice forecasts at these higher spatial resolutions and with higher accuracy requirements is problematic because the passive microwave observations of SIC that are widely used in the climate and forecast research community are not appropriate for such purposes.

On weather time scales (days ahead), which are the focus of this research, there are less available studies on the performance and comparison of existing operational sea ice forecast models. Nevertheless, there are some efforts to routinely monitor forecast performance of some of the CMS forecasts such as in Melsom, Palerme and Müller (2019) [16], which demonstrates explicitly and objectively how ARC-MFC forecast skill decreases for increased spatial resolutions. Some verification has also been performed for other operational ice forecasting products, including the Canadian Global Ice Ocean Prediction System (GIOPS) [17], the Regional Ice Prediction System (RIPS) from Environment Canada [18] and the US navy ice forecast system [19].

Based on the current research, the main limitations of existing systems concern the spatial resolution, data assimilation schemes, sea ice observations used for forecast initialisation and biases present in coupled atmosphere–ocean–ice systems. Probabilistic information about decametre-scale features could also provide useful information to improve such forecasts. Additionally, widespread provision of ensembles and retrospective forecasts would allow for improved forecast calibration and assessment.

The objective of this paper is to present, for the first time, the performance of the main existing SIT short-range forecasts currently on CMS or elsewhere in the public domain. These are seven-day forecasts from the UK Met Office, MET Norway, the Nansen Environmental and Remote Sensing Center (NERSC) and the European Centre for Medium-Range Weather Forecasts (ECMWF). Forecast performance is evaluated against unique in-situ data from five moorings deployed between 2016 and 2019 by the Barents Sea Metocean and Ice Network (BASMIN) and Barents Sea Exploration Collaboration (BaSEC) Joint Industry Projects. The observed SIT can vary considerably during a seven-day forecast, with the primary source of variability being ice advection. Significant thermodynamic growth of ice can also occur on these timescales, particularly in leads and areas of thin ice where growth can be of the order of 10 cm per day [20]. Assessing the viability of short-term SIT models for navigational purposes is therefore important in order to understand the extent to which this variability can be forecasted.

2. Materials and Methods

2.1. Observational Data

SIT observations were obtained from five moorings deployed between 2016 and 2019 by the BASMIN Joint Industry Project. The locations of these moorings are shown in Figure 1. Each mooring is equipped with an ASL Ice Profiler, an upward facing sonar that can sample sea ice drafts at a single location once per second. One-minute and ten-minute averages of sea ice draft were calculated from the raw observations in 2016 and 2017–2019, respectively, with an estimated accuracy of measurements being $+/-0.05$–0.10 m.

In this study, the sea ice draft was averaged to generate daily values and converted from sea ice draft measurements to SIT by dividing observations by 0.89, following the method of Rothrock, Zhang and Yu [21]. This conversion factor is applied because approximately 89% of sea ice thickness is below the surface and is measured by the ASL Ice Profiler, but there are uncertainties that surround this value. In particular, snow depth uncertainty is a large source of error in satellite and in situ measurements, both in the retrievals of freeboard and the subsequent conversion to SIT [22,23]. Due to the linear relationship between SIT and snow depth, an underestimation of snow depth can result in an underestimation of SIT. Additional uncertainty is also introduced through a lack of knowledge of snow and sea ice densities.

Mooring data were averaged to daily values so that a direct comparison could be made with forecast data that is available at hourly to daily resolutions depending on the model. This allows for a statistical comparison to be made, but the process of averaging masks the underlying spatial and temporal heterogeneity of SIT. If the probability distribution of observed SIT is skewed such that the mean and mode values differ, then the mean SIT could be unrepresentative of the typical ice thickness on a given day. This limitation to the method is exemplified in Appendix A by Figure A1a–c, which displays the distribution of sea ice draft for a day in April 2018 at three of the mooring sites.
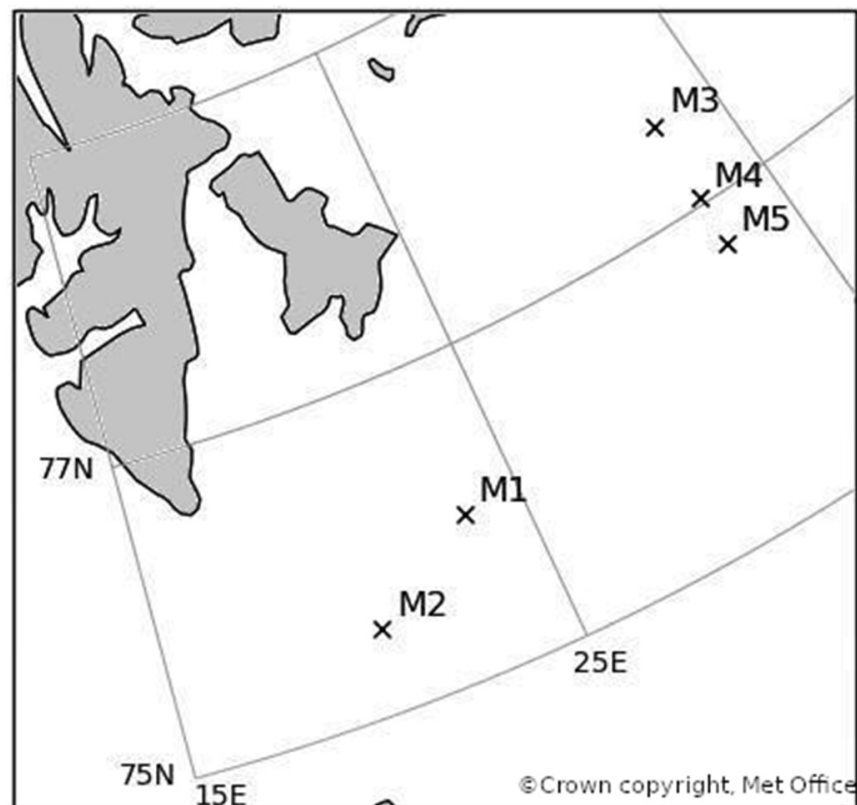
Figure 1. The locations of the five mooring sites M1–5 in the Barents Sea near Svalbard [24].

2.2. Model Data

The SIT forecasting systems evaluated in this study are the UK Met Office Coupled atmosphere–land–ocean–ice Data Assimilation (CPLDA) system, the ECMWF ensemble hindcast system, the MET Norway Towards an Operational Prediction System for the North Atlantic and European Coastal Zones 4 (TOPAZ4) and the NERSC neXt generation Sea Ice Model (neXtSIM-F). Table 2 presents information on the forcing, resolution, assimilation, thermodynamics and dynamic schemes for each product. Table 2 also contains values for h_0, the model parameter that determines the thickness of newly formed sea ice in a grid cell that previously was ice free.

Data from the different models were archived on different days, and unless otherwise stated, comparisons were only made where data were available for all models on the same days. This means that direct comparisons between all models across every lead time is not always possible. Model data are archived more frequently for analysis fields than forecast fields, and this has been used to produce multi-model comparisons of analysis performance in this study.

In order to analyse the forecasts at lead times up to seven days, a persistence model has been used as a performance baseline for comparison. This persistence model is produced by using SIT values at a mooring site as the predictor for SIT in the future at that same site. Versions of the dynamical models evaluated are available on CMS or otherwise in the public domain, as described in the data availability statement, although exact model versions may differ and generally only analysis fields are archived on CMS.

Table 2. Information about the forcing, resolution, assimilation, thermodynamics and dynamic schemes used by each model system. There are four models included: Coupled atmosphere–land–ocean–ice Data Assimilation (CPLDA); Towards an Operational Prediction System for the North Atlantic and European Coastal Zones 4 (TOPAZ4); ECMWF Ensemble hindcast; Nansen Environmental and Remote Sensing Center (NERSC); and neXt generation Sea Ice Model (neXtSIM-F).

Model system	CPLDA	Ensemble hindcast (Cycle 46R1)	TOPAZ4	neXtSIM-F
Institution	UK Met Office	ECMWF	Met Norway	NERSC
Atmosphere	Coupled to Unified Model (UM) atmosphere	Coupled to the Integrated Forecast System (IFS)	ECMWF atmospheric forcing	ECMWF atmospheric forcing
Ocean	Coupled to Nucleus for European Modelling of the Ocean (NEMO) 3.4	Coupled to NEMO 3.4	Coupled to Hybrid Coordinate Ocean Model (HYCOM) ocean	TOPAZ ocean forcing
Ice resolution	~13 km	~13 km	12.5 km	7.5 km [1]
Ice assimilation	Sea ice concentration (SIC) from Special Sensor Microwave Imager/Sounder (SSMIS)	Univariate three-dimensional variational data assimilation—first guess at appropriate time (3DVAR-FGAT) of SIC	SSMIS SIC and Soil Moisture and Ocean Salinity (SMOS) satellite derived sea ice thickness (SIT) assimilated using the ensemble Kalman filter	SIC from SSMIS and Advanced Microwave Scanning Radiometer 2 (AMSR2)
Ice dynamics	Elastic Viscous Plastic (EVP) [25]	Viscous Plastic (VP) [26]	EVP [25]	Brittle Bingham–Maxwell (BBM) [27]
Thermodynamics	Zero-layer model [28]	Three-layer [28]	Three-layer [28]	Three category model [29]
Time period of data used	December 2016–July 2019	March 2016–July 2019	March 2016–July 2019	November 2018–July 2019
Model parameter h_0	0.5 m	0.6 m	0.1 m	0.05 m

[1] Note that neXtSIM-F uses an adaptive triangular mesh; thus, the effective resolution is higher than square grid Eulerian models at equal resolution.

2.2.1. neXtSIM-F

neXtSIM-F is a fully Lagrangian sea ice model being developed at NERSC [30]. The model runs on an adaptive triangular mesh with an average side length of 10 km (with the distance from each point to the opposite side being about 7.5 km) and is interpolated to a regular 3 km grid for distribution [27]. The overall scheme of the model, including information on forcing and data assimilation, is depicted in Figure 2. SIC data are assimilated as a weighted average of the Advanced Microwave Scanning Radiometer 2 (AMSR2) and Special Sensor Microwave Imager/Sounder (SSMIS) products as described in Williams et al. (2021) [27]. The model was initialised in October 2018 using SIT from the merged CryoSat-2 and Soil Moisture and Ocean Salinity (CS2SMOS) product from AWI [31].
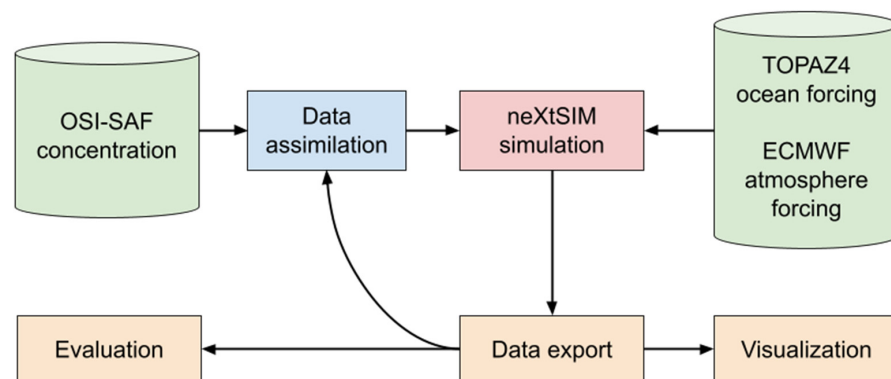
Figure 2. The neXtSIM-F model scheme [27]. Both SSMIS and AMSR2 concentration data from Ocean and Sea Ice Satellite Application Facility (OSI-SAF) are assimilated.

2.2.2. TOPAZ4

TOPAZ4 is a coupled ocean–sea ice data assimilation product which uses the Hybrid Coordinate Ocean Model (HYCOM) coupled to an early version of the Los Alamos Sea Ice model (CICE) with simple thermodynamics [28,32]. An ensemble Kalman Filter data assimilation scheme assimilates SIC and ocean observations [33]. Operationally, the model is run with ten ensemble members and produces ten days of forecast interpolated to a 12.5 km polar stereographic grid. Only the ensemble average is distributed through CMS, which causes some smoothing of model fields. TOPAZ4 assimilates the Ocean and Sea Ice Satellite Application Facility (OSI-SAF) SSMIS SIC hosted on CMS, with no additional corrections. In 2017, the assimilation of thin SIT from the Soil Moisture and Ocean Salinity (SMOS) product was introduced in the winter, and since November 2019, the merged CS2SMOS ice thickness product has replaced the SMOS product. Reconciliation between SIT and SIC measurements is part of the data assimilation procedure, and SIC and SIT observations and their model counterparts enter the same cost function using their respective uncertainty measurements. Where there is inconsistency between SIT and SIC, SIC measurements prevail because of their lower uncertainties. The system uses a seven-day assimilation cycle and assimilates the gridded Sea Surface Temperature (SST), SIC and ice drift fields for the day of the analysis; and along-track Sea Level Anomaly (SLA) and in-situ temperature and salinity for the week prior to the day of the analysis.

2.2.3. CPLDA

The Met Office CPLDA system [34] is run operationally at the UK Met Office with coupled data assimilation. The model assimilates SIC from OSI-SAF, which is based on SSMIS SIC measurements. The sea ice component of the system uses CICE v4.1 on the ORCA025 tripolar horizontal grid, which has a resolution of approximately 13 km in the Barents Sea. Forecasts and analyses from CPLDA have been provided to CMS since 2017 on a regular 0.25° grid.

2.2.4. ECMWF

Archived ECMWF ensemble perturbed hindcasts of SIT (model Cycle 46R1) were used on the ORCA025 tripolar grid, which has a spatial resolution of approximately 13 km in the Barents Sea. These hindcasts have a control member and ten perturbed members, and they extend to a lead time of 46 days. For the purposes of this study, the ensemble mean of the ten perturbed members has been used up to seven days. The model assimilates SIC from the OSTIA product, which applies some corrections to the OSI-SAF product, and is based on SSMIS SIC measurements.

2.3. Performance Metrics

In order to compare the performance of different models with observational data, daily averages of the mooring data and of the model data were interpolated to each mooring location and used to calculate correlation coefficients, mean residuals and mean accuracy. This was repeated at different lead times of the models, up to seven days.

Modified Taylor diagrams were produced to display the correlation coefficients and mean residuals between the modelled and observed SIT. These are similar to those used by Johnson et al. (2012) [35] to compare modelled SIT against a variety of data sources including moored upward looking sonars. Positive correlation coefficients in these diagrams are represented by the radial difference from the origin. Negative coefficients cannot be displayed on this diagram; however, in this study, no forecast was negatively correlated with observations. The mean residuals are scaled to an angle of (–180, 180) from the positive x-axis such that the perfect forecast lies where the outer circle meets the positive x-axis (Figure 3). Pearson's correlation was used as the coefficient to represent correlation values in these plots, although similar results are found if this is repeated with Kendall's Tau. Forecast accuracy was calculated for all models as a percentage of days where the model SIT was within +/−0.2 m of the observed SIT, following the definition of a hit in Figure 4.
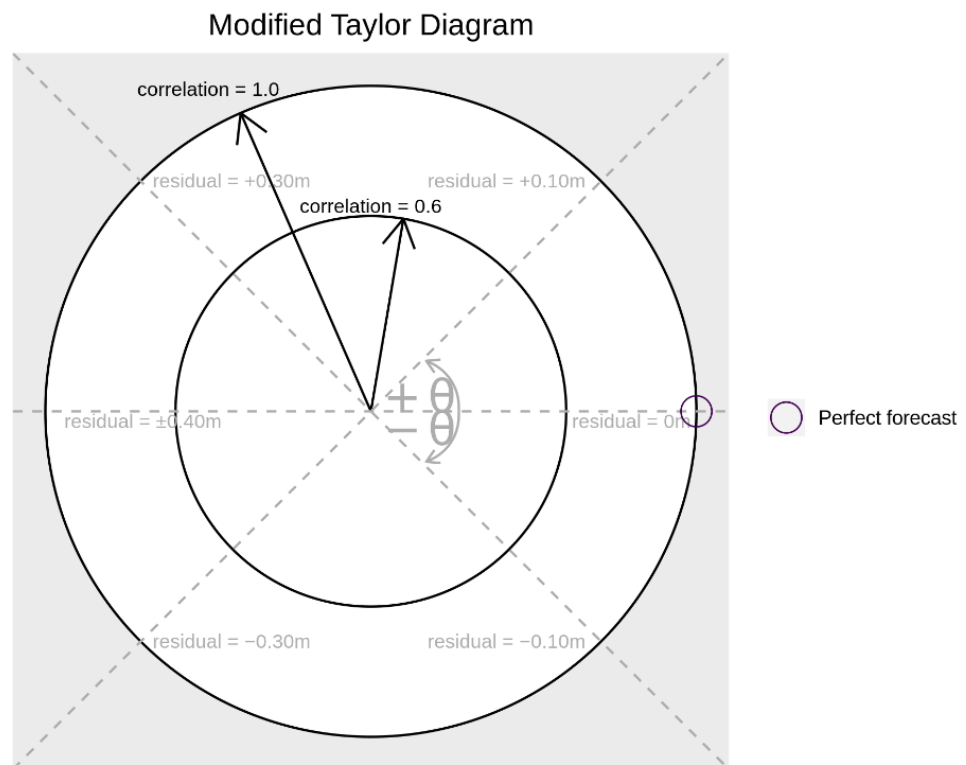
Figure 3. An example of a modified Taylor diagram, where the correlation coefficient is represented by the radial distance from the origin and the mean residual by the angle from the positive x-axis.
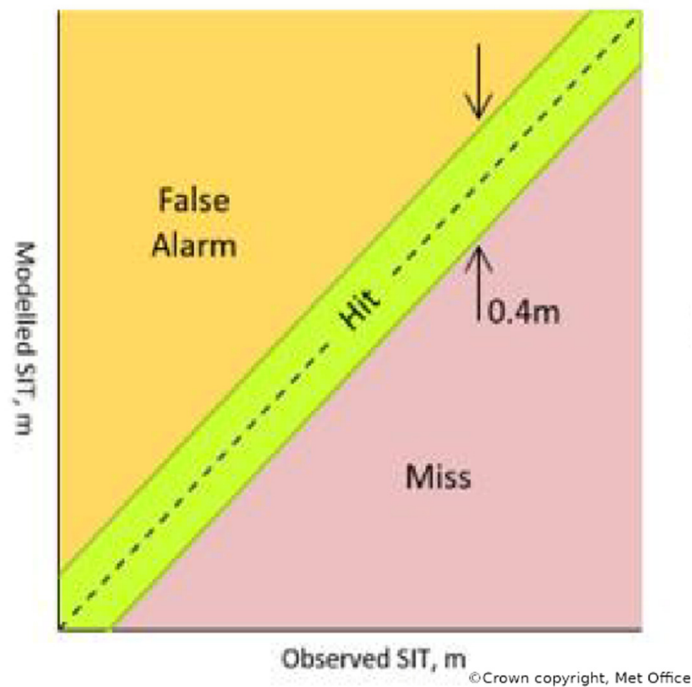
Figure 4. Accuracy is defined as the percentage of successful hits. Hits are defined as forecasts within +/−0.2 m of mooring measurements [24].

3. Results

3.1. Case Study at Mooring Location M3

The model systems appear to perform poorly in 2019 at the mooring location M3, with all four models having a large number of days where the analysis field modelled

SIT is high, but the observed values are zero. This overestimation of SIT is evident in the scatter plots of daily modelled and observed SIT (Figure 5) and can be observed for all four models. By plotting different winters in separate colours, it is apparent that the models do not generally overestimate the SIT in this manner.
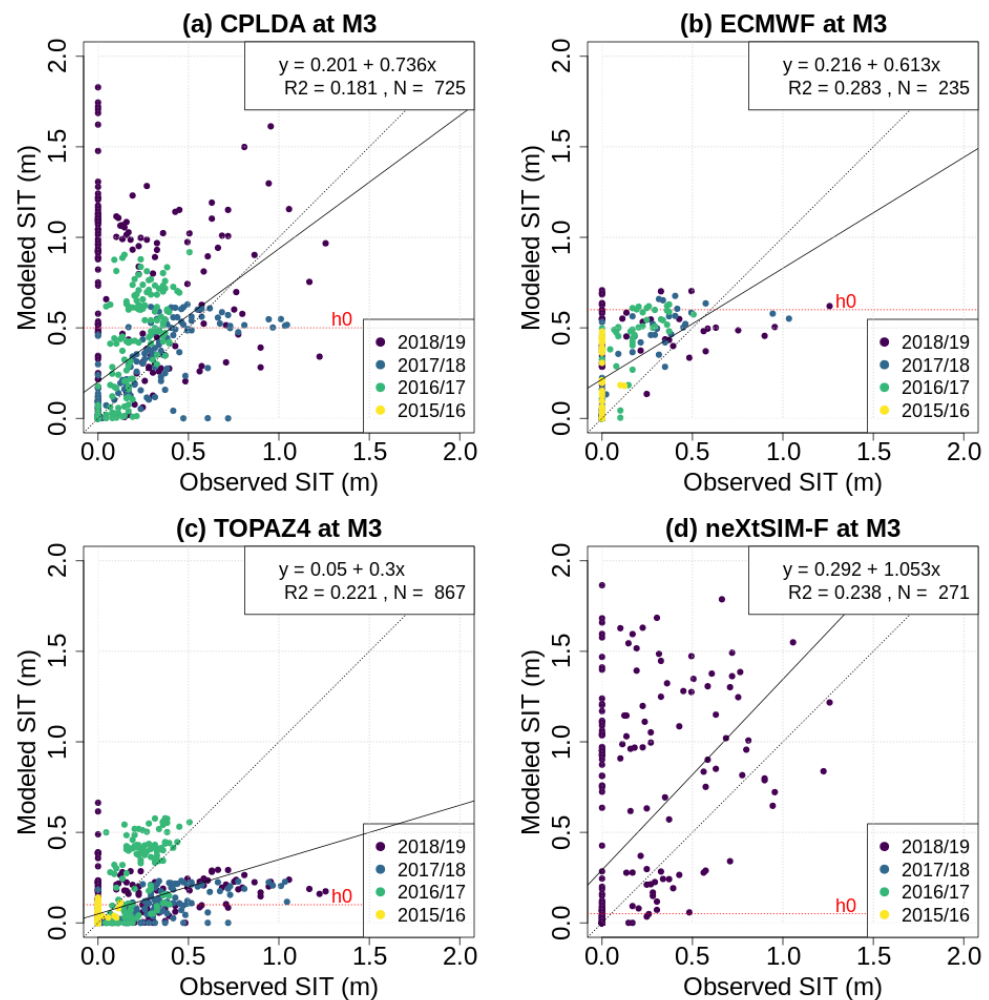
Figure 5. Scatter plot of daily average model analysis fields against mooring observations of SIT at mooring M3 for the following: (**a**) CPLDA; (**b**) ECMWF; (**c**) TOPAZ4; and (**d**) neXtSIM-F. Different colours represent the values from different winter observational campaigns. The line x = y (dashed) is depicted as well as the model regression line (solid) and the value for the model tuning parameter h_0 (red, dashed).

Comparing the seven-day running mean of modelled and observed SIT in 2019 suggests that poor performance could be related to an unusual temporal pattern in SIT at the mooring M3 during this period. The observed SIT increases rapidly in February and early March 2019 and then abruptly declines in late March, whilst the modelled SIT develops well but fails to decrease (Figure 6c). SIT over the same period in previous years displays a more typical evolution and is better represented by the models (Figure 6a,b). Repeating these plots at mooring locations M1 and M5 revealed that this event was localised in nature, occurred to a much lesser extent at M5 (Figure A2) and was not observed at M1 (Figure A3). SIT charts from April 2019 suggest that there was substantial ice in the region surrounding M3, but the presence of an area categorised as 'Close Drift Ice' rather than 'Very Close Drift Ice' on some days indicates that there may have been a localised break-up of sea ice over the mooring location (Figure 7). The relevance of this type of event is highlighted in the Section 4. Observations from M3 have been excluded from most of the remainder

of the analysis in this study because the results at this location are dominated by model overestimation presented in this section.
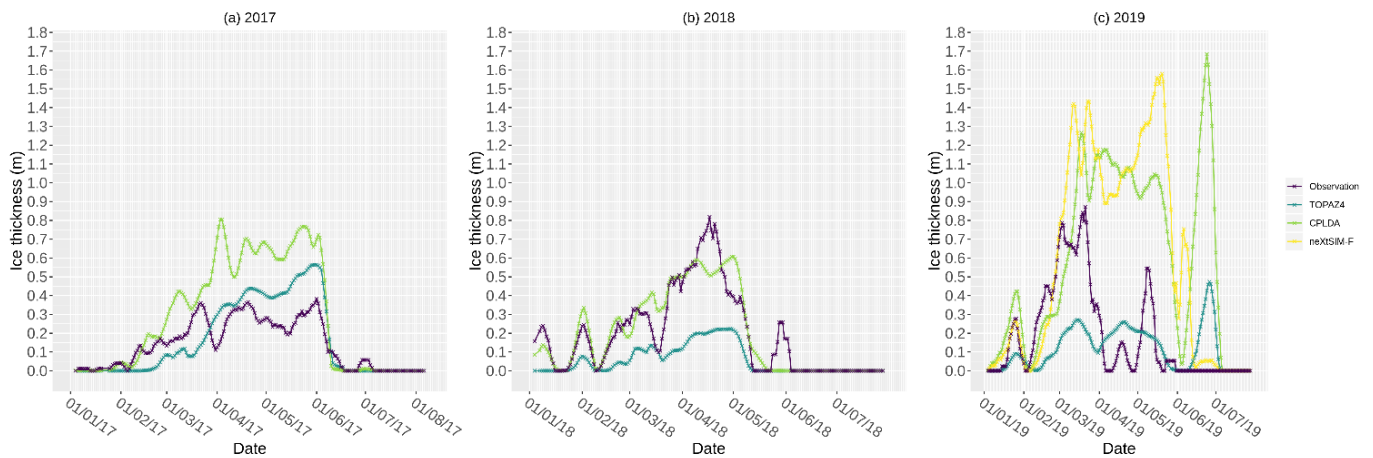
Figure 6. Timeseries plots of seven-day running mean SIT for Observations (purple), CPLDA (Green), TOPAZ4 (Blue) and neXtSIM-F (Yellow) at mooring M3 for the following periods: (**a**) January 2017–July 2017; (**b**) January 2018–July 2018; and (**c**) January 2019–July 2019. Note that neXtSIM-F data are only available for 2019.
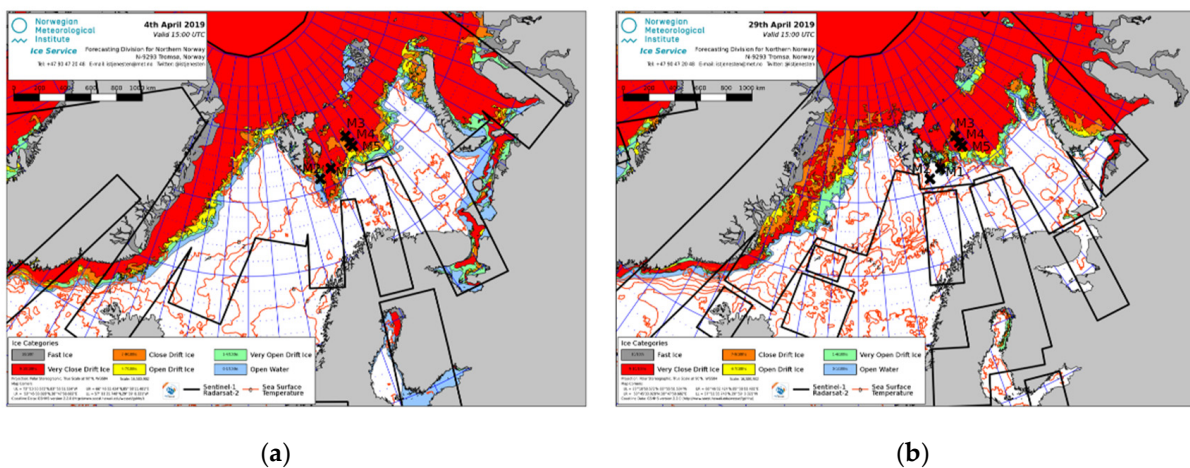

Figure 7. Ice charts for the following: (**a**) 4 April 2019; (**b**) 29 April 2019. These charts show localised areas of close drift ice near the mooring M3 [36].

3.2. Analysis of Correlation and Residuals

Correlation coefficients and mean residuals can be used to compare the performance of the model analysis and forecast fields interpolated to the mooring locations M1, M2, M4 and M5. These are displayed in modified Taylor Diagrams, which are described in Figure 3.

3.2.1. Model Analysis Fields

Forecast data from neXtSIM-F were only available from November 2018 onwards; thus, the residuals and correlations of the analysis fields have been plotted for all four model systems for the period November 2018–July 2019. At mooring locations M1 and M2, all models have correlation coefficients between 0.6 and 0.9 (Figure 8a), but the TOPAZ4 model has a negative bias, and the ECMWF model has a positive bias. Results are similar at moorings M4 and M5; however, at these locations, CPLDA and neXtSIM-F forecasts overestimate SIT (Figure 8b).
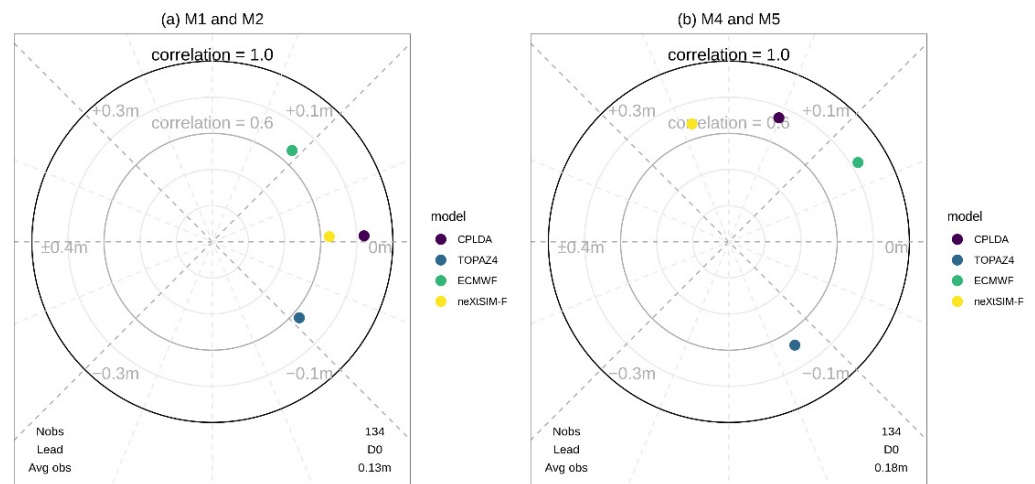
Figure 8. Modified Taylor diagrams for CPLDA, TOPAZ4, ECMWF and neXtSIM-F models for November 2018–July 2019 for the analysis fields at moorings: (**a**) M1 and M2; (**b**) M4 and M5. As well as correlation coefficients and mean residuals, the number of total observations and the average SIT across observations are displayed at the bottom of the plot.

This comparison is repeated for CPLDA, TOPAZ4 and ECMWF models for a longer period spanning December 2016–July 2019. Compared with the shorter period, results show similar performances at M1 and M2 but show a smaller overestimation of SIT by CPLDA at moorings M4 and M5 (Figure 9). CPLDA, TOPAZ4 and ECMWF model systems all had an average bias of less than $+/-0.1$ m at the mooring locations during this longer period of time.
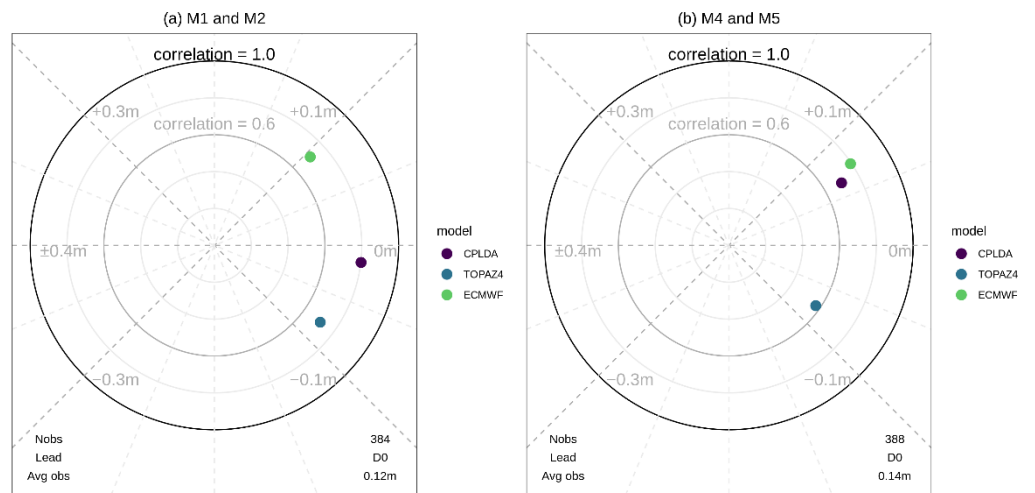
Figure 9. Modified Taylor diagrams for CPLDA, TOPAZ4 and ECMWF models for December 2016–July 2019 for the analysis fields at moorings: (**a**) M1 and M2; (**b**) M4 and M5.

3.2.2. Model Forecast Fields

Model data were more frequently archived for analysis fields than for forecast fields; thus, a full multi-model comparison of performance across all lead times was not possible. In order to evaluate the performance of the model forecasts at different lead times, correlations and residuals have been plotted for each lead day and compared to a persistence model.

The CPLDA system has a correlation coefficient that drops off gradually with lead time at the different mooring sites (Figure 10a,b). The bias varies depending on the mooring site and appears to increase with lead time, which could indicate that ice is growing too quickly in winter and spring months or is melting too slowly in the melt season. For the

ECMWF model system, the relationship between correlation and lead time is less clear, and bias does not vary greatly with lead time (Figure 10c,d). The results for the TOPAZ4 and neXtSIM-F models also reflect this, and there is generally only a weak relationship between model performance and lead time in the correlation and residual plots (Figure 11).

As one would expect, the persistence model has a very low bias, and its correlation coefficients decrease with increasing lead time. The correlation coefficients for the persistence model generally decay faster than that of the dynamical models, and at mooring location M1, all four models have a better correlation with observations at a lead time of seven days than the persistence model.
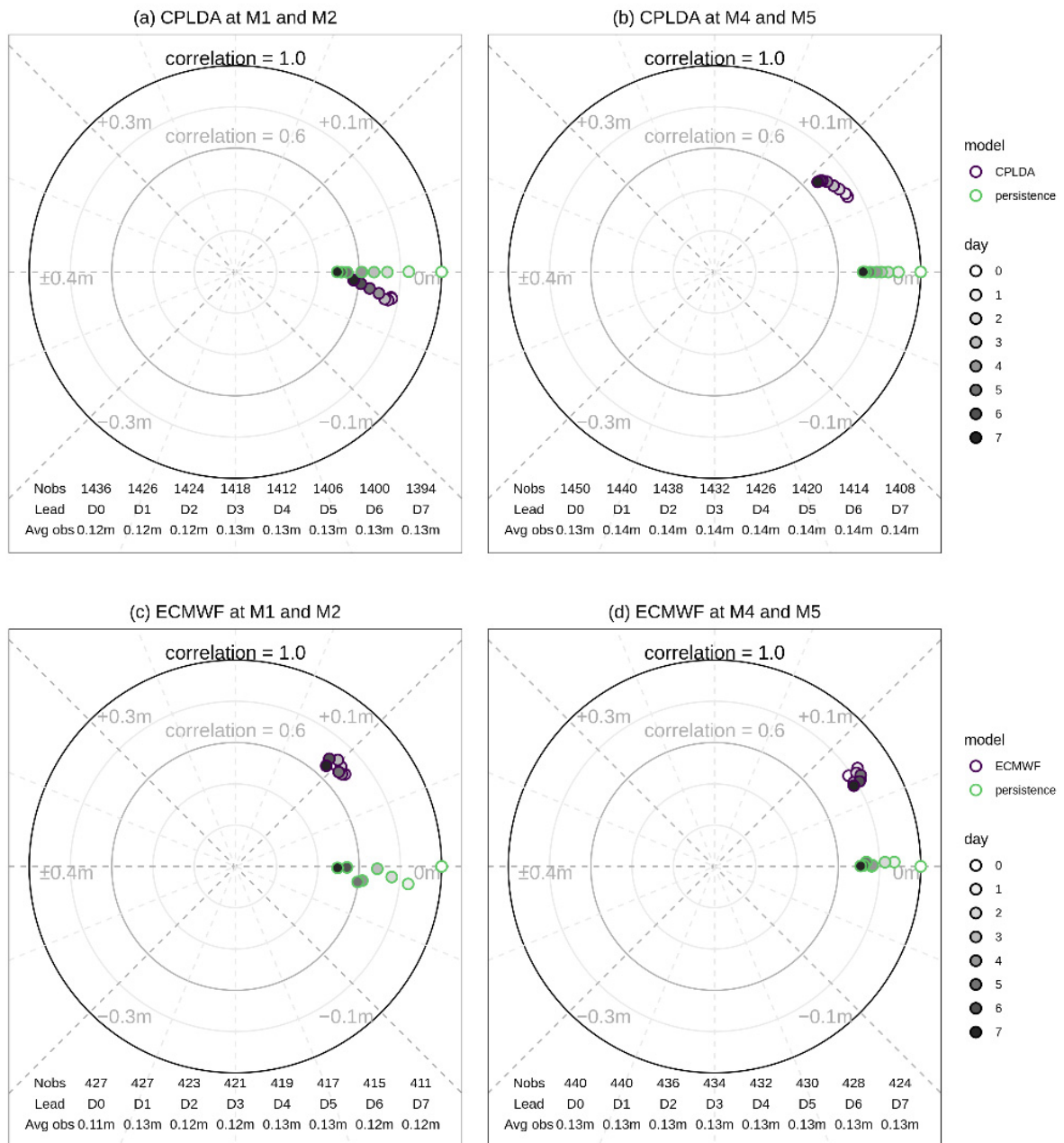
Figure 10. Mean residual and correlation coefficients for lead times of 0–7 days for forecast fields of: (**a**) CPLDA at M1 and M2; (**b**) CPLDA at M4 and M5; (**c**) ECMWF at M1 and M2; (**d**) ECMWF at M4 and M5.

The ratio of the day seven to day zero forecast performance can provide information on how model performance varies with lead time and is provided as a ratio of correlations and residuals for all models in Figure 12. For the CPLDA system, the correlation coefficients at a lead time of seven days are all less than at day zero, but this ratio is a lot closer to 1 for other model systems. This indicates that model correlation coefficients at day seven are largely dependent on analysis field correlations.
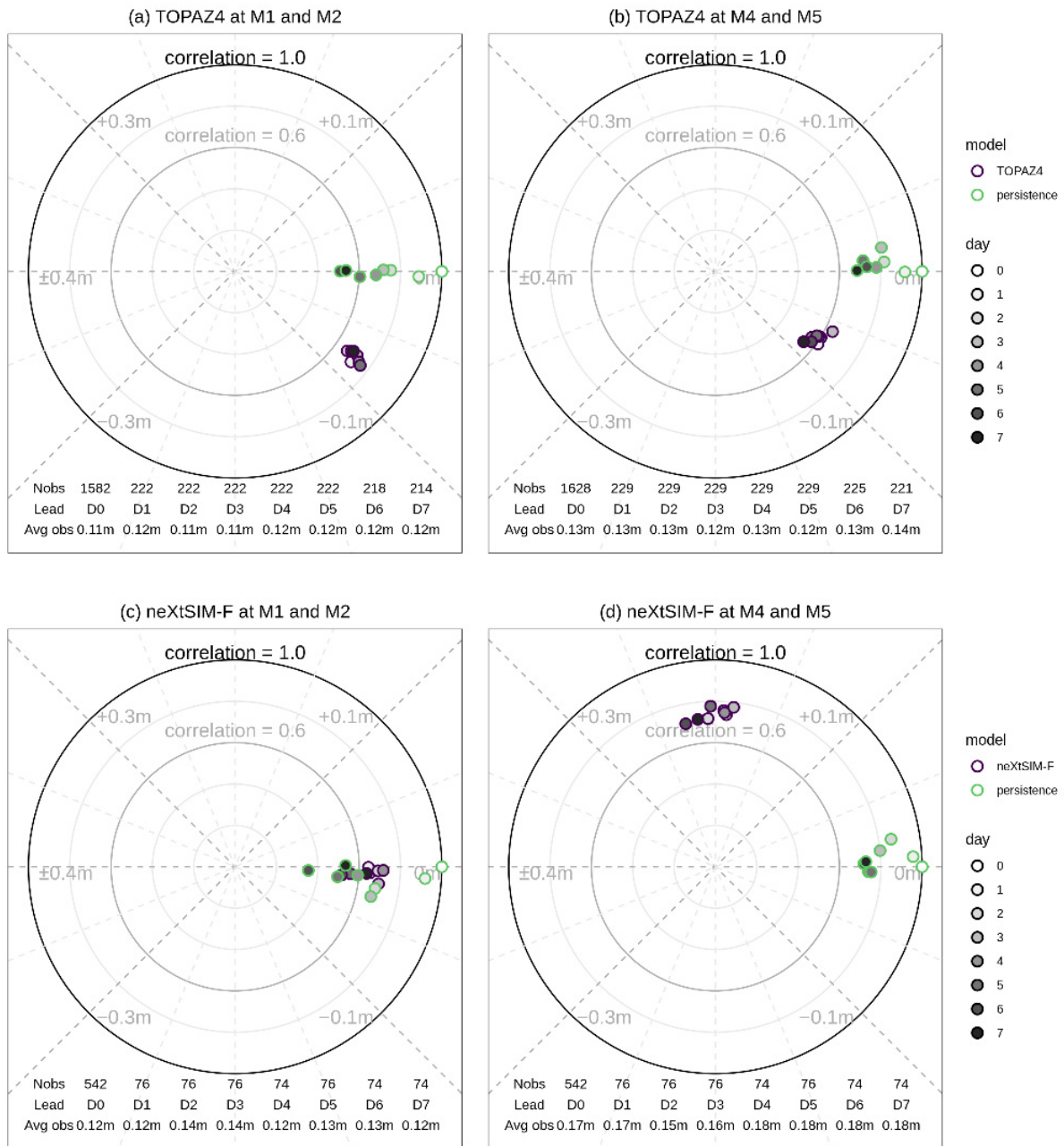
Figure 11. Mean residual and correlation coefficients for lead times of 0–7 days for forecast fields of: (**a**) TOPAZ4 at M1 and M2; (**b**) TOPAZ4 at M4 and M5; (**c**) neXtSIM-F at M1 and M2; (**d**) neXtSIM-F at M4 and M5.
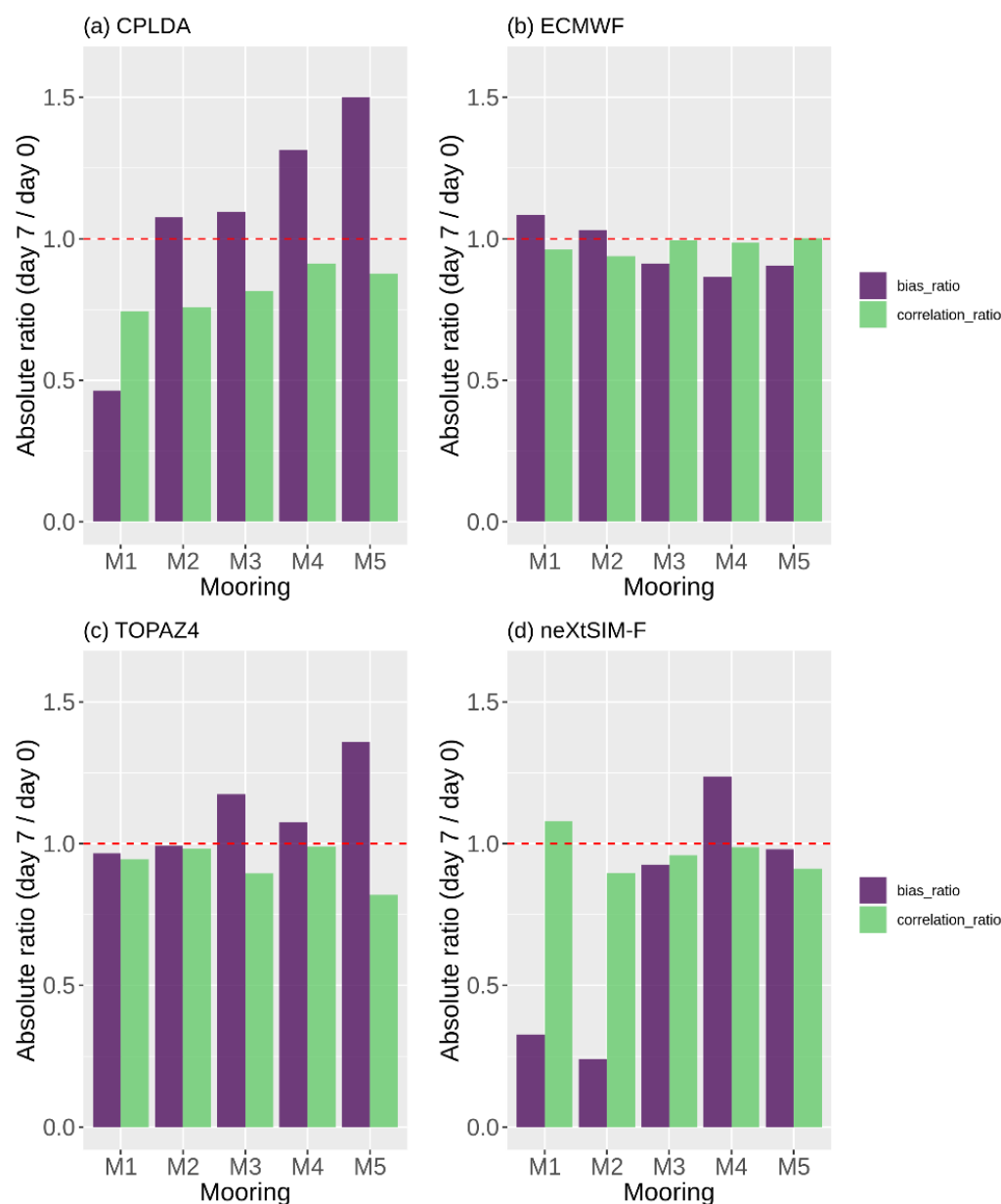
Figure 12. Plots displaying the bias and correlation ratios of day seven to day zero model fields for: (**a**) CPLDA; (**b**) ECMWF; (**c**) TOPAZ4; (**d**) neXtSIM-F.

3.3. Forecast Accuracy

Accuracy, defined as the percentage of days in which the modelled SIT is within +/−0.2 m of observed SIT, can be plotted as a function of lead time for the models at the different mooring sites. Figure 13 shows the accuracy of CPLDA, TOPAZ4 and ECMWF model systems averaged over mooring sites M1, M2, M4 and M5 between 2017 and 2019. Note that in order to produce this comparison across different lead times, data from different days have been used to produce accuracy values for each model. The period of data of January 2017–July 2019 has been selected because the three models compared all have data at a constant frequency and with no gaps during this time.

Generally, the CPLDA model displays a slight decrease in accuracy from 85% to 80% as the lead time is increased to seven days. The TOPAZ4 and ECWMF models have accuracies of 77–83% and 74–78%, respectively, and for both systems, the accuracy does not have an obvious dependency on lead time.
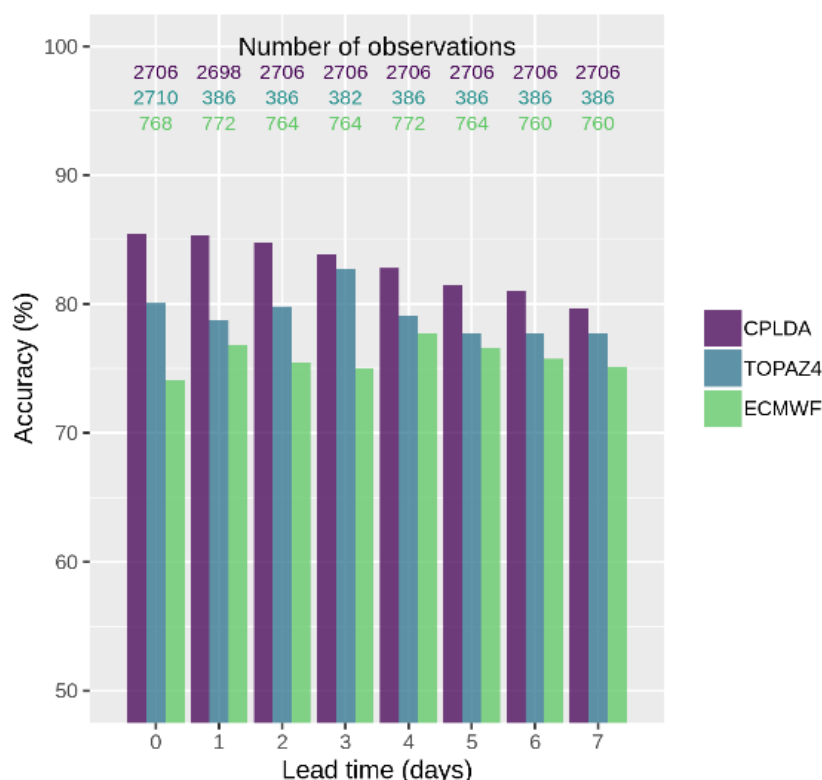
Figure 13. Accuracy for CPLDA, TOPAZ4 and ECMWF models for the period January 2017–July 2019 averaged across moorings M1, M2, M4 and M5. Accuracy is defined as the percentage of days that the forecast is within +/−0.2 m of the observed SIT.

3.4. Analysis Field Variability

Scatter plots can be used to further explore the extent to which models accurately represent SIT compared to observational data. Splitting the data by year can also inform us about how much the annual variability of SIT impacts the models' predictive capability. Data have been shown at mooring locations M1 and M5 because M1 is representative of the M1/M2 region, and M5 is representative of the M4/M5 region.

Figure 14a,b show that CPLDA slightly underestimates SIT at mooring M1 and slightly overestimates at mooring M5, with SIT being overestimated the most in 2019. On average, ECMWF hindcasts have very little bias in the scatter plots, but the distribution of points highlights some overestimation on days where the observed SIT was low (Figure 14c,d).

TOPAZ4 generally underestimates SIT at both mooring sites (Figure 15a,b) and nextSIM-F performs well at M1 but overestimates SIT at M5 (Figure 15c,d). The model seems to perform well on average at M1 but overestimates SIT at the M5 mooring. For the neXtSIM-F system there is a large variability in the modelled values compared to the observational ones, which could be a result of the model resolving high resolution structures in the sea ice; these structures can be observed in the example model field (Figure 16a) when compared to examples of fields from TOPAZ4 and CPLDA (Figure 16b,c).

The value of h_0 has also been plotted for each model, where h_0 is the model parameter that provides the initial thickness of ice that has formed in a previously ice-free grid cell. Previous work has found that there is a relationship between the value of h_0 and ice thickness in the LIM2 model, with annual mean model SIT increasing with the value of h_0 [37]. The ECMWF product appears to display an overestimation of SIT for thin ice between 0 and 0.5 m, with many of the modelled values for SIT being near the h_0 value of 0.6 m. It could be the case that, during ice formation, new grid cell values of 0.6 m contribute to the overestimation that we have observed. In a similar manner, the underestimation seen for the TOPAZ4 product could be in part due to the lower model h_0 parameter value of 0.1 m.
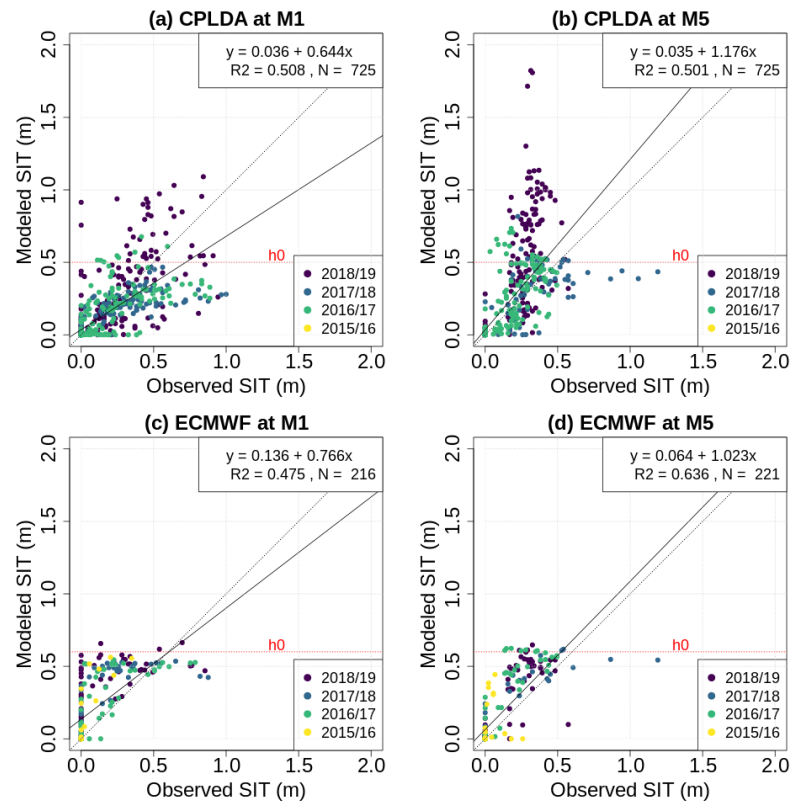
Figure 14. Scatter plots comparing daily analysis and observation SIT for the following: (**a**) CPLDA at M1; (**b**) CPLDA at M5; (**c**) ECMWF at M1; (**d**) ECMWF at M5. The line x = y (dashed) is depicted as well as the model regression line (solid) and the value for the model tuning parameter h_0 (red, dashed).
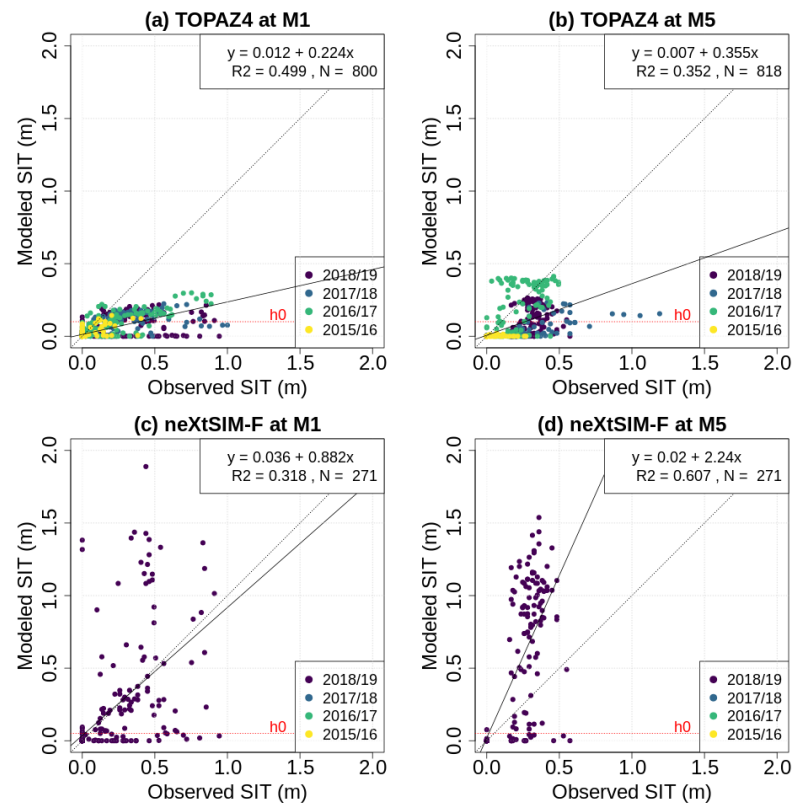
Figure 15. Scatter plots comparing daily analysis and observation SIT for the following: (**a**) TOPAZ4 at M1; (**b**) TOPAZ4 at M5; (**c**) neXtSIM-F at M1; (**d**) neXtSIM-F at M5.
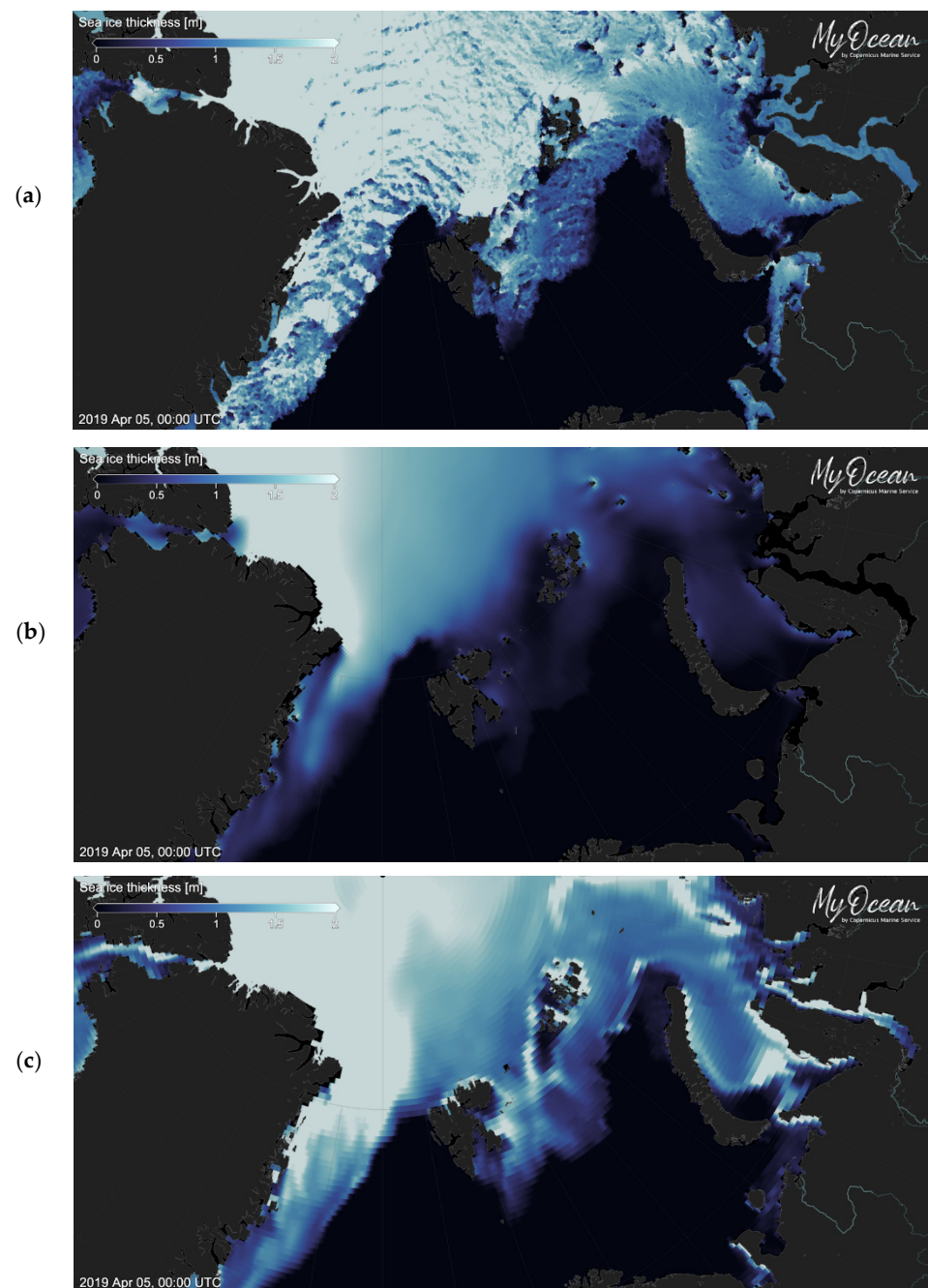
Figure 16. Example model SIT fields for 5 April 2019 from the following: (**a**) neXtSIM-F; (**b**) TOPAZ4; (**c**) CPLDA. These are obtained from the CMS viewing system.

4. Discussion

In this study, we have compared four publicly available SIT forecasts against each other and against a persistence model at five mooring sites in the Barents Sea.

In Section 3.1, we present a case study at mooring location M3 which describes an event in spring 2019 in which observed SIT at M3 decreased rapidly, and the four model systems failed to replicate this, resulting in poor performances in both analysis and forecast fields. SIT models display thickness values as cell-averaged means and are unable to explicitly resolve subgrid-scale processes such as leads that form in the ice. This presents a challenge for usability in shipping because, from a navigational perspective, the sub-kilometre SIT is important for safety and route optimisation. Validation of SIT models against mooring data particularly highlights this challenge, as localised breakup of sea ice

over a mooring site can cause models to perform poorly when compared to cell-averaged means. The evidence presented in our case study suggests that, at their current resolution, these forecast models cannot provide reliable information for tactical ice management during operations where maritime safety is of importance. Developing forecast systems with improved resolution will enhance the ability to model localised ice conditions, but it is also important that high-resolution data are assimilated into the forecasts so that modelled SIT does not drift from observed conditions. Assimilating SIT could help prevent large discrepancies between modelled and observed SIT in cases where ice is present (e.g., Figure A2c), but in situations in which the modelled ice edge is inaccurate, assimilation of higher resolution SIC data would provide more substantial improvements for forecasting because of the better sensitivity of SIC measurements.

By analysing the model performance more generally, we found that the models correctly forecasted SIT out to seven days for approximately 75–85% of days. We found that correlation coefficients between models and observations at different lead times vary in the range of 0.5–0.9 depending on the mooring site and model. Comparing the model performances to that of a persistence model provided us with a benchmark performance to compare the models against. Model performance decreased only gradually with lead time, which resulted in the seven-day forecasts of the model systems performing as well as or better than that of the persistence model. This indicates that these dynamical model systems can be useful for predicting mean sea ice conditions in regions where ice is variable on daily to weekly timescales. Such information could be useful for short-term planning purposes to monitor how the large-scale ice thickness in a region is predicted to change, but in practice the low resolution of the models remains a limiting factor in usability.

A previous evaluation of the CPLDA product found that its sea ice has a tendency to melt too much in summer months [34], which could result in a negative bias at larger lead times in the summer. We did not observe such a negative bias in our results, and in fact found that bias generally increased slightly with lead time; this difference could have arisen because performance metrics in this study were calculated as an average across different seasons.

The TOPAZ4 product appears to underestimate SIT, which is consistent with the results of an evaluation of the corresponding reanalysis which found that there is a SIT bias of approximately -1.1 m when compared to data from ICESat and IceBridge, as well as sea ice draft data [38,39]. In this study, we have found that bias appears to be of the order of -0.1 m, and the TOPAZ4 system correctly predicts SIT to within $+/-0.2$ m on 75–85% of days analysed, across lead times of up to seven days. The neXtSIM-F product, which uses TOPAZ4 ocean forcing but with a more advanced and higher resolution representation of sea ice, appears to no longer underestimate SIT, although it subsequently overestimated SIT at mooring locations M4 and M5 in 2019.

We have considered the possible influence of h_0 on performance, where h_0 is the model parameter that defines the thickness of newly formed sea ice. It has been demonstrated that the tuning of h_0 can have a significant impact on modelled SIT [37], and in this study, we have seen evidence that the choice of h_0 could be causing a bias in SIT forecasts at the mooring sites.

In order to understand the relationship between performance and model lead time, we plotted the bias for each model system at different lead times up to and including the seven-day forecasts. These plots indicate that, rather than being a function of lead time, the bias of the forecasts is largely dependent on the initial SIT bias of the analysis field, which could be caused by errors in the SIT or SIC. On these same plots, the correlation coefficients between modelled and observed SIT seem to decrease with lead time, but the magnitude of decrease appears to be small compared to the variation in analysis field correlation coefficients across mooring sites. This highlights the importance of having an accurate analysis field, which can be achieved by assimilating SIT data into the models. Regular assimilation of SIT into forecasts could therefore increase the performance of the

models and should be considered an important step in the development of SIT forecasting products for arctic navigation.

It is important to note that because this study only covers up to four years of observational data, individual events can have a large impact on the overall performance metrics for a model system. In particular, it is hard to draw conclusions about the performance of the neXtSIM-F product because only one year of model data has been used for this study. We suggest that similar observational comparisons are made using longer periods of data and using the newly released version of neXtSIM-F.

To extend this study, we would recommend that future studies make use of the ensemble nature of the ECMWF SIT product. Ensemble members provide numerous realisations of ice conditions, the spread of which can be used as a measure of the forecast uncertainty, as currently applied in seasonal ice forecasting [12]. Whilst this does not improve the overall performance of the model, providing a measure of uncertainty alongside short-term forecasts could be useful for navigational planning purposes, although further work is necessary to explore the extent to which observations fall within the ensemble range of uncertainty. We would also recommend evaluating the performance of the ECMWF product on longer timescales because forecast data are available for lead times up to 46 days, although we note that our study has found that model error does not seem to increase appreciably within lead times up to seven days. Future work could also explore the possibility of evaluating the performance of multi-model ensemble forecasts. Batté et al. (2020) [12] have provided evidence that multi-model ensembles can be effective in predicting sea ice edge when compared to forecasts from single model systems, where sea ice edge was defined as the edge containing cells with 15% or more SIC.

All models evaluated in this study assimilate the SSMIS product, which due to the nature of passive microwave data, has uncertainties which are unsuitable for many maritime activities. Further studies are also therefore required to investigate the benefits of improving SIC assimilation into forecasting products to improve their usability for navigational purposes.
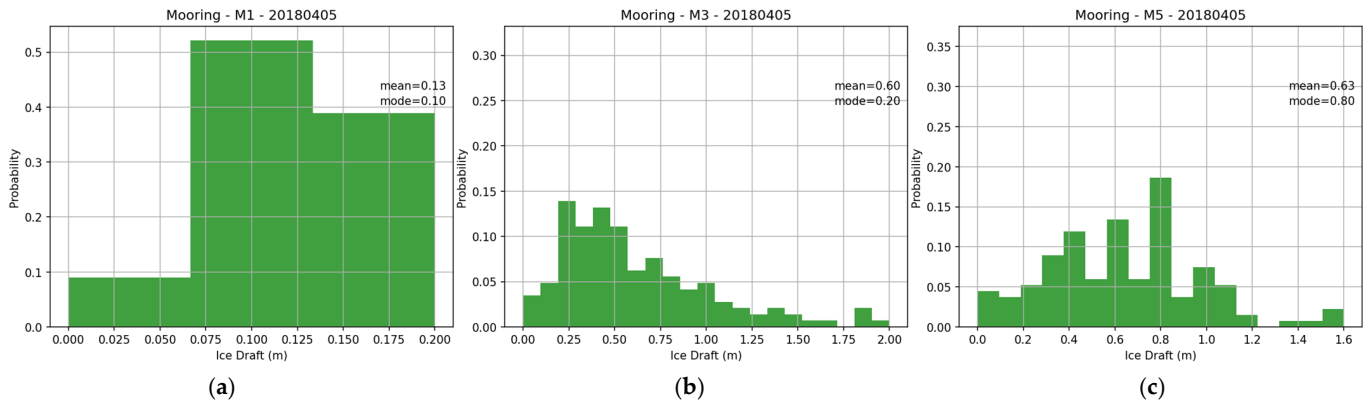
Appendix A

Figure A1. Histograms of ice draft on 5 April 2018 at mooring locations: (**a**) M1; (**b**) M3; (**c**) M5.

Figure A2. Timeseries plots of seven-day running mean SIT for Observations (purple), CPLDA (Green), TOPAZ4 (Blue) and neXtSIM-F (Yellow) at mooring M5 for the period: (**a**) January 2017–July 2017; (**b**) January 2018–July 2018; (**c**) January 2019–July 2019. Note that neXtSIM-F data are only available for 2019.
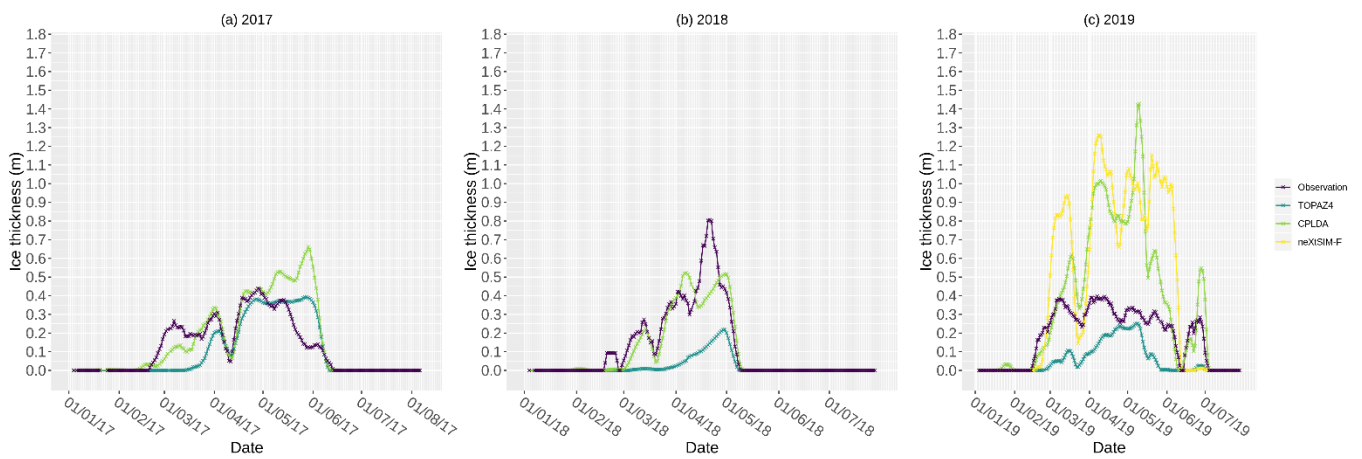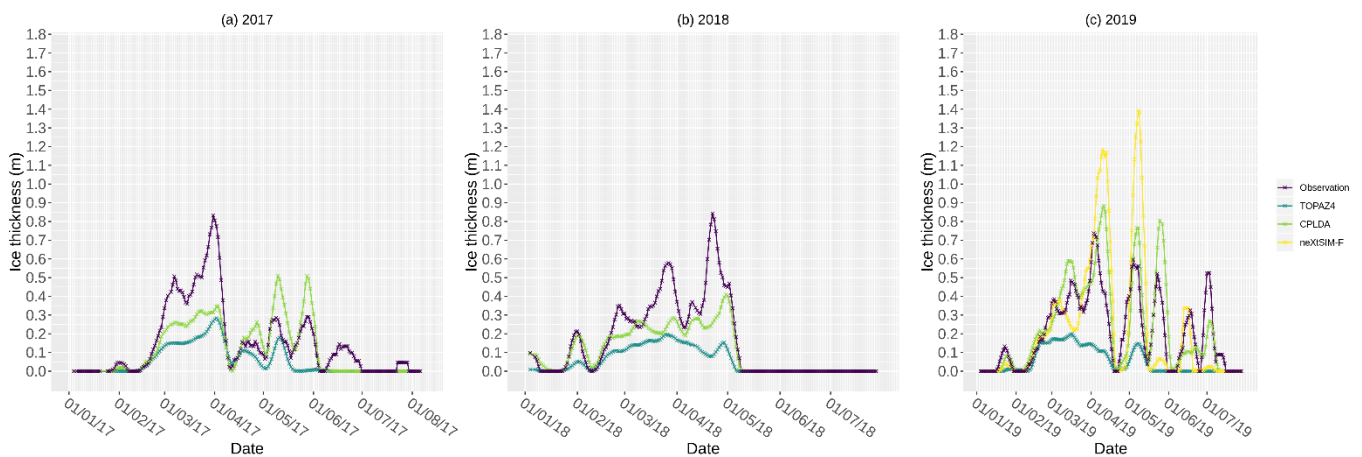

Figure A3. Timeseries plots of seven-day running mean SIT for Observations (purple), CPLDA (Green), TOPAZ4 (Blue) and neXtSIM-F (Yellow) at mooring M1 for the period: (**a**) January 2017–July 2017; (**b**) January 2018–July 2018; (**c**) January 2019–July 2019.

References

1. IMO Polar Code Operational Requirements. Available online: https://rules.dnv.com/docs/pdf/DNV/CG/2020-07/DNVGL-CG-0308.pdf (accessed on 1 September 2021).
2. Steidl, C.; Daniel, L.; Yildiran, C. Shipbuilding market developments, Q2 2018. *Organ. Econ. Co-Oper. Dev.* **2018**, *15*, 23.
3. Wagner, P.M.; Hughes, N.; Bourbonnais, P.; Stroeve, J.; Rabenstein, L.; Bhatt, U.; Little, J.; Wiggins, H.; Fleming, A. Sea-Ice Information and Forecast Needs for Industry Maritime Stakeholders. *Polar Geogr.* **2020**, *43*, 2–3. [CrossRef]
4. Notz, D.; Community, S.I.M.I.P. Arctic Sea Ice in CMIP6. *Geophys. Res. Lett.* **2020**, *47*, e2019GL086749. [CrossRef]
5. Wei, T.; Yan, Q.; Qi, W.; Ding, M.; Wang, C. Projections of Arctic Sea Ice Conditions and Shipping Routes in the Twenty-First Century Using CMIP6 Forcing Scenarios. *Environ. Res. Lett.* **2020**, *15*, 104079. [CrossRef]
6. Jung, T.; Gordon, N.D.; Bauer, P.; Bromwich, D.H.; Chevallier, M.; Day, J.J.; Dawson, J.; Doblas-Reyes, F.; Fairall, C.; Goessling, H.F.; et al. Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales. *Bull. Am. Meteorol. Soc.* **2016**, *97*, 1631–1647. [CrossRef]
7. Tietsche, S.; Mueller, M.; Bertino, L.; Garric, G.; Goessling, H.; Pedersen, L.T.; Wagner, P.; Lavergne, T. Recommendations for more user-relevant sea-ice forecasts (1.0). *Zenodo* **2020**, 33. [CrossRef]
8. Koenigk, T.; Mikolajewicz, U. Seasonal to Interannual Climate Predictability in Mid and High Northern Latitudes in a Global Coupled Model. *Clim. Dyn.* **2009**, *32*, 783–798. [CrossRef]
9. Blanchard-Wrigglesworth, E.; Armour, K.C.; Bitz, C.M.; Deweaver, E. Persistence and Inherent Predictability of Arctic Sea Ice in a GCM Ensemble and Observations. *J. Clim.* **2011**, *24*, 231–250. [CrossRef]
10. Tietsche, S.; Day, J.J.; Guemas, V.; Hurlin, W.J.; Keeley, S.P.E.; Matei, D.; Msadek, R.; Collins, M.; Hawkins, E. Seasonal to Interannual Arctic Sea Ice Predictability in Current Global Climate Models. *Geophys. Res.* **2014**, *41*, 1035–1043. [CrossRef]
11. Andersson, T.R.; Hosking, J.S.; Pérez-Ortiz, M.; Paige, B.; Elliott, A.; Russell, C.; Law, S.; Jones, D.C.; Wilkinson, J.; Phillips, T.; et al. Seasonal Arctic Sea Ice Forecasting with Probabilistic Deep Learning. *Nat. Commun.* **2021**, *12*, 5124. [CrossRef]
12. Batté, L.; Välisuo, I.; Chevallier, M.; Acosta Navarro, J.C.; Ortega, P.; Smith, D. Summer Predictions of Arctic Sea Ice Edge in Multi-Model Seasonal Re-Forecasts. *Clim. Dyn.* **2020**, *54*, 5013–5029. [CrossRef]
13. Dirkson, A.; Denis, B.; Merryfield, W.J. A Multimodel Approach for Improving Seasonal Probabilistic Forecasts of Regional Arctic Sea Ice. *Geophys. Res. Lett.* **2019**, *46*, 10844–10853. [CrossRef]
14. Melia, N.; Haines, K.; Hawkins, E.; Day, J.J. Towards Seasonal Arctic Shipping Route Predictions. *Environ. Res. Lett.* **2017**, *12*, 084005. [CrossRef]
15. Zampieri, L.; Goessling, H.F.; Jung, T. Bright Prospects for Arctic Sea Ice Prediction on Subseasonal Time Scales. *Geophys. Res. Lett.* **2018**, *45*, 9731–9738. [CrossRef]
16. Melsom, A.; Palerme, C.; Müller, M. Validation Metrics for Ice Edge Position Forecasts. *Ocean. Sci.* **2019**, *15*, 615–630. [CrossRef]
17. Smith, G.C.; Roy, F.; Reszka, M.; Surcel Colan, D.; He, Z.; Deacu, D.; Belanger, J.M.; Skachko, S.; Liu, Y.; Dupont, F.; et al. Sea Ice Forecast Verification in the Canadian Global Ice Ocean Prediction System. *Q. J. R. Meteorol. Soc.* **2016**, *142*, 659–671. [CrossRef]
18. Lemieux, J.F.; Beaudoin, C.; Dupont, F.; Roy, F.; Smith, G.C.; Shlyaeva, A.; Buehner, M.; Caya, A.; Chen, J.; Carrieres, T.; et al. The Regional Ice Prediction System (RIPS): Verification of Forecast Sea Ice Concentration. *Q. J. R. Meteorol. Soc.* **2016**, *142*, 632–643. [CrossRef]
19. Posey, P.G.; Metzger, E.J.; Wallcraft, A.J.; Hebert, D.A.; Allard, R.A.; Smedstad, O.M.; Phelps, M.W.; Fetterer, F.; Stewart, J.S.; Meier, W.N.; et al. Improving Arctic Sea Ice Edge Forecasts by Assimilating High Horizontal Resolution Sea Ice Concentration Data into the US Navy's Ice Forecast Systems. *Cryosphere* **2015**, *9*, 1735–1745. [CrossRef]
20. Thorndike, A.S.; Rothrock, D.A.; Maykut, G.A.; Colony, R. The Thickness Distribution of Sea Ice. *J. Geophys. Res.* **1975**, *80*, 4501–4513. [CrossRef]
21. Rothrock, D.A.; Zhang, J.; Yu, Y. The Arctic Ice Thickness Anomaly of the 1990s: A Consistent View from Observations and Models. *J. Geophys. Res. Ocean.* **2003**, *108*, 3083. [CrossRef]
22. Giles, K.A.; Laxon, S.W.; Wingham, D.J.; Wallis, D.W.; Krabill, W.B.; Leuschen, C.J.; McAdoo, D.; Manizade, S.S.; Raney, R.K. Combined Airborne Laser and Radar Altimeter Measurements over the Fram Strait in May 2002. *Remote Sens. Environ.* **2007**, *111*, 182–194. [CrossRef]
23. Ricker, R.; Hendricks, S.; Perovich, D.K.; Helm, V.; Gerdes, R. Impact of Snow Accumulation on CryoSat-2 Range Retrievals over Arctic Sea Ice: An Observational Approach with Buoy Data. *Geophys. Res. Lett.* **2015**, *42*, 4447–4455. [CrossRef]
24. Hume-Wright, L.; Fiedler, E.; Fournier, N.; Mendes, J.; Blockley, E.; Martin, M.; Eik, K. Sea Ice Thickness Forecast Performance in the Barents Sea. In Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering, Virtual, Online, 3–7 August 2020. [CrossRef]
25. Hunke, E.C.; Dukowicz, J.K. The Elastic-Viscous-Plastic Sea Ice Dynamics Model in General Orthogonal Curvilinear Coordinates on a Sphere-Incorporation of Metric Terms. *Mon. Weather Rev.* **2002**, *130*, 1848–1865. [CrossRef]
26. Hibler, W.D. A Dynamic Thermodynamic Sea Ice Model. *J. Phys. Oceanogr.* **1979**, *9*, 815–846. [CrossRef]
27. Williams, T.; Korosov, A.; Rampal, P.; Ólason, E. Presentation and Evaluation of the Arctic Sea Ice Forecasting System NeXtSIM-F. *Cryosphere* **2021**, *15*, 3207–3227. [CrossRef]
28. Semtner, A.J. A Model for the Thermodynamic Growth of Sea Ice in Numerical Investigations of Climate. *J. Phys. Oceanogr.* **1976**, *6*, 379–389. [CrossRef]

29. Rampal, P.; Dansereau, V.; Olason, E.; Bouillon, S.; Williams, T.; Korosov, A.; Samaké, A. On the Multi-Fractal Scaling Properties of Sea Ice Deformation. *Cryosphere* **2019**, *13*, 2457–2474. [CrossRef]

30. Rampal, P.; Bouillon, S.; Ólason, E.; Morlighem, M. NeXtSIM: A New Lagrangian Sea Ice Model. *Cryosphere* **2016**, *10*, 1055–1073. [CrossRef]

31. Ricker, R.; Hendricks, S.; Kaleschke, L.; Tian-Kunze, X.; King, J.; Haas, C. A Weekly Arctic Sea-Ice Thickness Data Record from Merged CryoSat-2 and SMOS Satellite Data. *Cryosphere* **2017**, *11*, 1607–1623. [CrossRef]

32. Drange, H.; Simonsen, K. *Formulation of Air-Sea Fluxes in the ESOP2 Version of MICOM*; NERSC Tech. Rep. 125; Nansen Environmental and Remote Sensing Center: Bergen, Norway, 1996; pp. 1–23.

33. Sakov, P.; Counillon, F.; Bertino, L.; Lister, K.A.; Oke, P.R.; Korablev, A. TOPAZ4: An Ocean-Sea Ice Data Assimilation System for the North Atlantic and Arctic. *Ocean. Sci.* **2012**, *8*, 633–656. [CrossRef]

34. Guiavarc'H, C.; Roberts-Jones, J.; Harris, C.; Lea, D.J.; Ryan, A.; Ascione, I. Assessment of Ocean Analysis and Forecast from an Atmosphere-Ocean Coupled Data Assimilation Operational System. *Ocean. Sci.* **2019**, *15*, 1307–1326. [CrossRef]

35. Johnson, M.; Proshutinsky, A.; Aksenov, Y.; Nguyen, A.T.; Lindsay, R.; Haas, C.; Zhang, J.; Diansky, N.; Kwok, R.; Maslowski, W.; et al. Evaluation of Arctic Sea Ice Thickness Simulated by Arctic Ocean Model Intercomparison Project Models. *J. Geophys. Res. Ocean.* **2012**, *117*, C00D13. [CrossRef]

36. Ice Service Charts. Available online: https://cryo.met.no/en/latest-ice-charts (accessed on 11 July 2021).

37. Wang, Z.; Lu, Y.; Wright, D.G.; Dupont, F. Sea Ice Sensitivity to the Parameterisation of Open Water Area. *J. Oper. Oceanogr.* **2010**, *3*, 3–9. [CrossRef]

38. Xie, J.; Bertino, L.; Knut, L.; Sakov, P. Quality Assessment of the TOPAZ4 Reanalysis in the Arctic over the Period 1991–2013. *Ocean. Sci.* **2017**, *13*, 123–144. [CrossRef]

39. Uotila, P.; Goosse, H.; Haines, K.; Chevallier, M.; Barthélemy, A.; Bricaud, C.; Carton, J.; Fučkar, N.; Garric, G.; Iovino, D.; et al. An Assessment of Ten Ocean Reanalyses in the Polar Regions. *Clim. Dyn.* **2019**, *52*, 1613–1650. [CrossRef]