



University of Bremen

Alfred Wegener Institute
Helmholtz Centre for Polar and Marine Research

Master's Thesis

Verification of Seasonal and Sub-seasonal Sea Ice Forecasts

Author:

Lorenzo Zampieri

Anagraphic data:

Venezia (IT), 01.09.1992

Matriculation Number:

3034011

First Supervisor:

Thomas Jung

Second Supervisor:

Christian Melsheimer

Tutor:

Helge Goessling

August 8, 2017

Urheberrechtliche Erklärung

Erklärung gem. § 10 (10) Allgemeiner Teils der MPO vom 27.10.2010

Hiermit versichere ich, dass ich meine Masterarbeit ohne fremde Hilfe angefertigt habe, und dass ich keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, habe ich unter Angabe der Quellen als solche kenntlich gemacht.

Die Masterarbeit darf nach Abgabe nicht mehr verändert werden.

Erklärung zur Veröffentlichung von Abschlussarbeiten

Ich bin damit einverstanden, dass meine Abschlussarbeit im Universitätsarchiv für wissenschaftliche Zwecke von Dritten eingesehen werden darf.

Bremen August 8, 2017

Lorenzo Zampieri



Abstract

Sea ice forecasts are becoming a demanding need since human activities in the Arctic are constantly increasing and this trend is expected to continue. Forecast system development needs to be guided by verification metrics that quantify skill in an appropriate way. Here we apply different verification metrics to real sea ice forecasts to study the behavior of the metrics and to quantify the predictive skills of the models, focusing on the sea ice edge position and on sub-seasonal to seasonal time scales. The employed metrics are the pan-Arctic sea ice extent (SIE) and area (SIA), the Integrated Ice Edge Error (IIEE), the Spatial Probability Score (SPS), and the Modified Hausdorff Distance (MHD). While the first two metrics evaluate a single integrated quantity, the latter three assess the spatial distribution of the ice cover. Forecasts are verified against the high resolution AMSR-E and AMSR2 89 GHz sea ice concentration products provided by the University of Bremen. Sea ice forecast products from various research institutes and operational centers are analyzed, in particular those collected within the Sub-Seasonal to Seasonal Prediction Project. The forecast systems are characterized by quite different features with regard to the spatial resolution and the complexity of the forecast model, the number of ensemble members and the forecast length. The broad pool of models allows a comprehensive analysis of the metrics' behavior in different situations, highlighting strengths and weaknesses of the models and of the metrics themselves.



Contents

- Contents** **3**

- 1 Introduction** **5**

- 2 Methods for sea ice forecast verification** **9**
 - 2.1 Pan-Arctic sea ice extent and area 10
 - 2.2 IIEE - Integrated Ice-Edge Error 10
 - 2.3 SPS - Spatial Probability Score 13
 - 2.4 Benchmark values for IIEE and SPS 15
 - 2.5 Hausdorff Distance and Modified Hausdorff Distance 16

- 3 Models and satellite data** **23**
 - 3.1 S2S forecasts models 23
 - 3.2 Prescriptive models 26
 - 3.3 Predictive Models 27
 - 3.4 Further sea ice forecast products 31
 - 3.5 Satellite measurement of sea ice concentration 32
 - 3.6 Contributions to model-observation disagreement 34

- 4 Verification of the S2S sea ice forecasts** **37**
 - 4.1 Verification against satellite observations 38
 - 4.1.1 Example verification of a single forecast 38

4.1.2	Extensive verification of the S2S database	40
4.1.3	Evaluation of the models' predictive skills	52
4.2	Comparison between predictive and prescriptive ECMWF model	56
4.3	Verification against models own analysis	58
5	Comparison with verification based on Modified Hausdorff Distance	67
6	Conclusions	73
6.1	Evaluation of the verification metrics	74
6.2	Predictive skills of the S2S forecast systems	76
6.3	Outlook	77
A	Climatological length of the sea ice edge	79
	Acronyms	81
	Bibliography	83

Introduction

The retreat of Arctic ice coverage is probably the most striking indicator of climate change observed in the last three decades. This statement is supported by model simulations and observations [IPCC (2013)]. The decrease of sea ice coverage and thickness in the Arctic region is particularly marked during summer. Climate models predict a continuation in this trend for the next century, mainly due to the amount of CO₂ released into the atmosphere and feedback mechanisms in the climate system. Several indicators even suggest the concrete possibility of having in some years a virtually ice-free Arctic Ocean in the month of September, during the second half of this century.

However, retreat of sea ice opens new scenarios for human activities in the Arctic region, mainly related to marine transport, offshore fuel, mineral extraction and in minor extent tourism. New shipping routes would be advantageous in particular for the commercial exchange between Europe and East Asia countries [Emmerson and Lahn (2012)]. As example, the navigation distance between Norway and China is half in case of Arctic route instead of Suez Canal route, with a consequent reduction in travel time of 16-20 days. For the previous reasons, an improvement of the forecast skills in the Arctic is a demanding need. This should not be limited to the atmospheric component but include a good description of sea ice properties, which is equally relevant for navigation. Emergency management and safety of humans and goods would certainly benefit from efficient and skillful forecast systems [Smith and Stephenson (2013)].

Despite the growing importance of sea ice forecasts, the research in this field is still in an early stage, concerning the initialization process, the forecast procedure itself and the final verifi-

cation of the products. In particular, the scientific community has still not tested enough possible verification metrics for sea ice forecasts. Most of the experiments conducted in this field are limited to idealized model scenarios and there is a lack of tests on real forecasts.

The research conducted in this thesis work aims to answer two main questions. Firstly, we investigate the behavior of new verification metrics when applied to real forecasts and observations. Special attention is given to those metrics able to describe the spatial distribution of Arctic sea ice, which is a relevant parameter for potential forecast users. Secondly, we will use those metrics to assess the prediction skills of sea ice in real forecast. The analysis is focused on sub-seasonal to seasonal time scales. In the last years, the scientific community made relevant progresses in seasonal atmospheric forecasts, which should have positive impact on sea ice forecasts for the same timescale. In this context, the recent availability of seasonal forecast databases such as the Sub-Seasonal to Seasonal Prediction Project has a particularly good timing and provides a solid base for the development of sea ice verification research.

The Sub-seasonal to Seasonal Prediction Project ([S2S](#)) is one of the three current World Water Research Program (WWRP) projects (one other is the Polar Prediction Project) that jointly replaced The Observing System Research and Predictability Experiment (THORPEX). The main goals of this project are improving the forecast skill and the understanding of the sub-seasonal to seasonal timescale, promoting the initiative's uptake by operational centers and exploitation by the applications community and finally capitalizing on the expertise of the weather and climate research communities to address issues of importance to the Global Framework for Climate Services [[Vitart et al. \(2012\)](#)].

The scientific objectives are mostly related to atmospheric phenomena, in particular in the tropics and sub-tropics. Some examples are the prediction of the Madden-Julian Oscillation and its impacts in numerical models, monsoon prediction, rainfall predictability and extreme events and stratospheric processes. However, polar predictability is also investigated and sea ice related variables are provided (sea ice concentration and possibly sea ice thickness in the future). Thus, the S2S project with its forecasts database represents a timely opportunity for our research, allowing tests of our metrics on real sea ice forecast. Further technical details about the S2S forecasts will be given in Section [3.1](#).

Another current international research initiative highly relevant to this thesis paper is the Year of Polar Prediction (YOPP) [[Jung et al. \(2014\)](#)], which is aimed at improved environmental prediction capabilities in polar regions and beyond, on time scales from hours to seasons. This

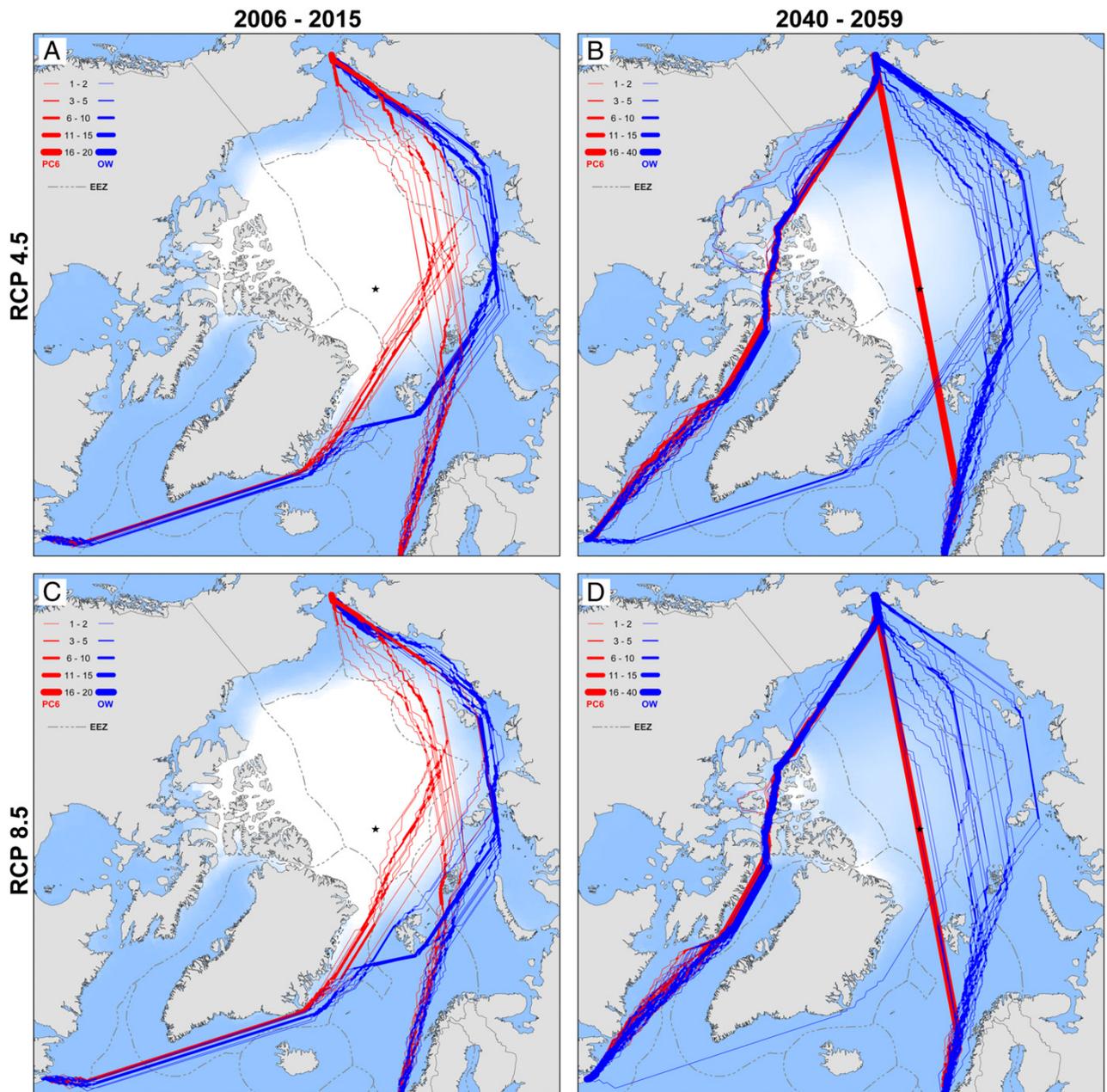


Figure 1.1: September navigation routes for hypothetical ships seeking to cross the Arctic Ocean between the North Atlantic and the Pacific during consecutive years 2006–2015 (A and C) and 2040–2059 (B and D). Predictions are driven assuming RCPs 4.5 (A and B; medium-low radiative forcing) and 8.5 (C and D; high radiative forcing) climate change scenarios. Red lines indicate fastest available trans-Arctic routes for PC6 (Polar Class) ships; blue lines indicate fastest available transits for common ships. Where overlap occurs, line weights indicate the number of successful transits using the same navigation route. White backdrops indicate period-average sea ice concentrations in 2006–2015 (A and C) and 2040–2059 (B and D). Figure from [Smith and Stephenson \(2013\)](#).

thesis is aligned with two of the objectives declared by YOPP, exploring the predictability of the atmosphere-cryosphere-ocean system and improving verification of polar weather and environmental predictions to obtain better quantitative knowledge on model performance [[Bauer et al.](#)

(2016)]. YOPP provides an excellent opportunity to deeply investigate sea ice forecasts in polar regions. One of the aspects tested in our research is understanding how sparseness and errors of observations in the polar regions affect sea ice verification results. This could serve as starting point during the core phase of YOPP, when enhanced observations will enable to better quantify observation uncertainties and possibly develop verification strategies to cope with such uncertainties.

Sea ice is one of the relevant components in the climate system. In particular, sea ice models play a key role in environmental prediction for polar regions, by providing ice products for polar marine users as well as a boundary forcing factor for atmospheric prediction. Thus, such a relevant actor needs an adequate verification system. The results of this our work could be taken into account when designing the metrics included in the verification environment, especially concerning the verification of spatial sea ice properties.

Methods for sea ice forecast verification

This section gives an overview of the metrics employed for the verification of sea ice forecasts. We will underline the advantages and disadvantages of each metric and their eventual suitability for probabilistic forecasts.

The main parameter investigated in this study is the sea ice concentration or quantities directly related to it. Furthermore, we will often refer to the sea ice edge. For this reason, we give here a brief definition of these parameters.

- **Sea Ice Concentration** is defined for each pixel or cell into which we divide our domain. Sea ice concentration is the percentage of the pixel area that is covered with sea ice.
- **The Sea Ice Edge** is generally defined as the 15% sea ice concentration contour. Even if a threshold of 15% is the most commonly used value in the scientific community, this is still arbitrary. Different thresholds can be defined in relation to the sea ice condition, particularly if 15% is found to be inappropriate to describe the edge.

One of the goals of this study is to estimate the agreement of forecasts and observations concerning the sea ice edge position. Thus, the metrics have been specifically chosen to meet this objective in the most accurate way, paying attention to the investigation of the spatial distribution of sea ice in the Arctic.

2.1 | Pan-Arctic sea ice extent and area

Pan-Arctic sea ice extent and area are currently the most used metrics to determine the goodness of a sea ice forecast. We report here the definitions we adopted in this study.

- **the Sea Ice Extent** is the area of sea with sea ice concentration larger than a threshold value. We fixed this value to the commonly used 15% threshold.
- **the Sea Ice Area** is the actual area covered by sea ice. To estimate ice area, the sea ice concentration of each pixel is multiplied by the pixel area, and then the values of all pixels are summed up to obtain the final value.

Sea ice extent and area have several advantages. They represent a physical quantity easy to read and to understand. They do not require difficult and time consuming calculations and they can be visualized in simple time series, instead of maps. Furthermore, their meaning can be easily explained to the non-scientific community. These two metrics have however some limitations, particularly when used for comparison and verification purposes. Firstly, they can not provide information about the spatial distribution of ice. Two totally different sea ice configurations can have the same sea ice extent or area and this can be misleading. Secondly, sea ice extent is a non conservative property when sea ice concentration is remapped to a new grid. Sea ice area is not sensitive to most of the remapping algorithms, and in general less sensitive than the extent to this procedure [Notz (2014)]. For the previous reasons, further verification metrics are needed to evaluate sea ice forecasts and to investigate the predictability of sea ice.

2.2 | IIEE - Integrated Ice-Edge Error

The Integrated Ice-Edge Error (IIEE) can be defined as the area where the forecast and the true state disagree on the ice concentration being above or below 15% (Eq. 2.1) [Goessling et al. (2016)]. Consequently, the IIEE is the sum of the area where the sea ice extent of the considered forecast is overestimated (O) or underestimated (U)

$$\text{IIEE} = U + O \quad (2.1)$$

with respect to the truth, which we assume to coincide with the observation.

As first step for the calculation, the sea ice concentration sic must be converted to binary sea ice concentration c . This is done by applying the usual 15% categorization

$$c = 0 \text{ if } sic < 15\%, \quad c = 1 \text{ if } sic \geq 15\%. \quad (2.2)$$

With c_o the binary sea ice concentration of the observation and c_f the binary sea ice concentration of the forecast, we can express O (Eq. 2.3) and U (Eq. 2.4) over the domain S (in our study S coincides with the northern hemisphere) as follows:

$$O = \int_A \max(c_f - c_o, 0) dS, \quad (2.3)$$

$$U = \int_A \max(c_o - c_f, 0) dS. \quad (2.4)$$

The map in Fig. 2.1 depicts the previous concepts.

The IIEE has some useful properties that make it a good metric for forecast verification.

1. The IIEE is conceptually simple and easy to calculate from modeled and observed gridded sea ice concentration data.
2. Sea ice concentration data are available since the late '70s and will be more and more accurate. This allows the verification of retrospective forecast as well as future predictions.
3. This parameter is more relevant for forecast users than the simple the pan-Arctic sea ice extent and area, already described in Section 2.1.
4. The decomposition into an absolute extent error component (AEE) and a misplacement error (ME) is immediate (see below).
5. An alternative formulation of the IIEE can be obtained by dividing the IIEE by the total length of the sea ice edge. More detail about the calculation of the total length of the climatological sea ice edge can be found in Appx. A. This new version of the IIEE can be interpreted as a sort of normalization of the metric. For this reason, we will refer to it as NIIEE. The NIIEE has the dimension of a length and provides an approximate average estimation of the distance between the observed and forecasted ice edges. This formulation of the IIEE is particularly easy to understand and thus to delivered to potential final users, for example shipping companies.

Concerning the absolute and misplacement error, we can express the IIEE as follows:

$$\text{IIEE} = \text{AEE} + \text{ME}. \quad (2.5)$$

Since $\text{AEE} = |O - U|$, an expression for ME can be easily derived from Eq. 2.1 and results

$$\text{ME} = 2 \min(O, U). \quad (2.6)$$

The misplacement error is a relevant parameter and can be useful for a potential forecast user, since it is adding relevant information compared to the simple total extent or area error. The users are reasonably not only interested in whether or not some ice is missing or overabundant, but also in an eventual error in the position of the ice itself. Furthermore, judging the goodness of a forecast only looking at the pan-Arctic sea ice extent and area is superficial and can lead to overestimate the value of the forecast.

Until now, we treated the IIEE only in the case of a single deterministic forecast, or, in case of a probabilistic forecast ensemble, a single ensemble member. However, the core of our analysis concerns predictive probabilistic forecasts with different ensemble members. The application of IIEE to this situation is not straightforward and requires particular attention.

A first intuitive approach would be to average the sea ice concentration of different ensembles, and afterwards calculate the IIEE. However, this method leads to some problems and has thus been discarded. Specifically, it would lead to a systematic overestimation of the sea ice edge position: the resulting average sea ice edge would be closer to the most external edge of the ensembles, and not in the middle of the edges interval as required.

A more sophisticated approach has been followed to overcome the overestimation problem encountered in the previous method. This method determines the actual median edge of the ensemble. Firstly, the probability of having a sea ice concentration larger than 15% has been calculated for each pixel based on the forecast ensemble (the so-called sea ice probability). Secondly, we defined the sea ice edge by considering as ice-covered the pixels with probability larger than 50%, and the rest as open ocean. The calculation of sea ice probability will be described in more details in Sec. 2.3. Additionally, if the number of ensemble members is even and relatively small ($N < 20$) the probability field needs to be corrected by applying a Gaussian filter, to generate a well-defined 50% contour avoiding again a systematic error of the edge.

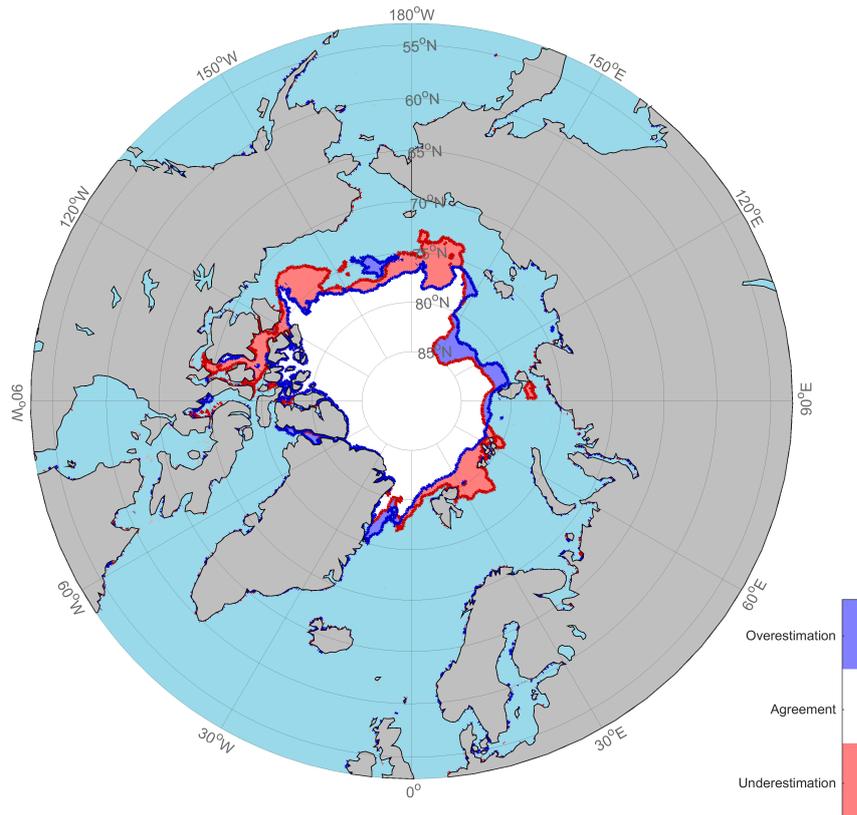


Figure 2.1: The map shows a sea ice forecast (blue contour) and the relative observation (red contour). The white region represent the agreement between the forecast and the observation, the blue region refers to the forecast overestimation area (O) and the red region to the underestimation area (U). The forecasted ice edge is indicated by the red solid line, while the observed edge by the blue solid line.

2.3 | SPS - Spatial Probability Score

The Spatial Probability Score (SPS) is a valuable metric for ensemble based forecasts of contours and thus particularly suitable for this study. Firstly, a formulation of SPS in its general definition is given. Afterwards, we will focus on the application of SPS to the sea ice edge in a two-dimensional domain.

Consider two spatial probability distributions, P_o describing the observation and P_f the forecast. Those are both defined on an N-dimensional domain V with $\vec{x} = (x_1 \dots x_N)$ a general element of V . Following Goessling and Jung (unpublished manuscript), we define the SPS as follows:

$$\text{SPS} = \int_V (P_o(\vec{x}) - P_f(\vec{x}))^2 dV. \quad (2.7)$$

We immediately note that SPS has m^N dimension (where m stands for meters, or in general a length). The argument of the integral in Eq. 2.7 is the local (Half) Brier Score [Wilks (2006)]. The previous definition is as general as possible. It can be applied to spatial variables in different dimensions, when these can be described by a probability distribution. SPS does not set particular constraints on the definition of the contour that we want to verify. This is defined only by choosing the statistical description of our quantity. Another relevant aspect is that SPS allows a probabilistic description of both the forecast and the observation. This means that we could include in our estimation eventual errors of the measurements.

To define the SPS in the frame of sea ice edge verification for probabilistic forecasts, we need first of all to describe the probability of our forecast. In our case this is the probability of finding a sea ice concentration above 15%. For the practical calculation, the sea ice concentration of each ensemble member must be firstly binarized as described in Eq. 2.2. Afterwards, the probability will be the average of the binarized sea ice concentrations within the ensembles. The domain is in our case the two-dimensional surface of the sea in the northern hemisphere, called S in this paper.

The observations considered in the first part of this thesis work are "deterministic" and thus their probability field simply consists of a two-dimensional scalar field with the following properties: $\tilde{x}_{ij} = 0$ if $sic < 15\%$ and $\tilde{x}_{ij} = 1$ if $sic \geq 15\%$.

The definition of the SPS for a 15% sea ice concentration contour line is

$$\text{SPS} = \int_S (P_o[sic \geq 15\%](\vec{x}) - P_f[sic \geq 15\%](\vec{x}))^2 dS. \quad (2.8)$$

Note that the SPS is also meaningful when applied to deterministic sea-ice edge forecasts. In this case, the SPS is reduced to the Integrated Ice Edge Error (IIEE) introduced in Section 2.2. This fact allows us to evaluate deterministic and probabilistic forecasts with these two verification metrics in the same frame.

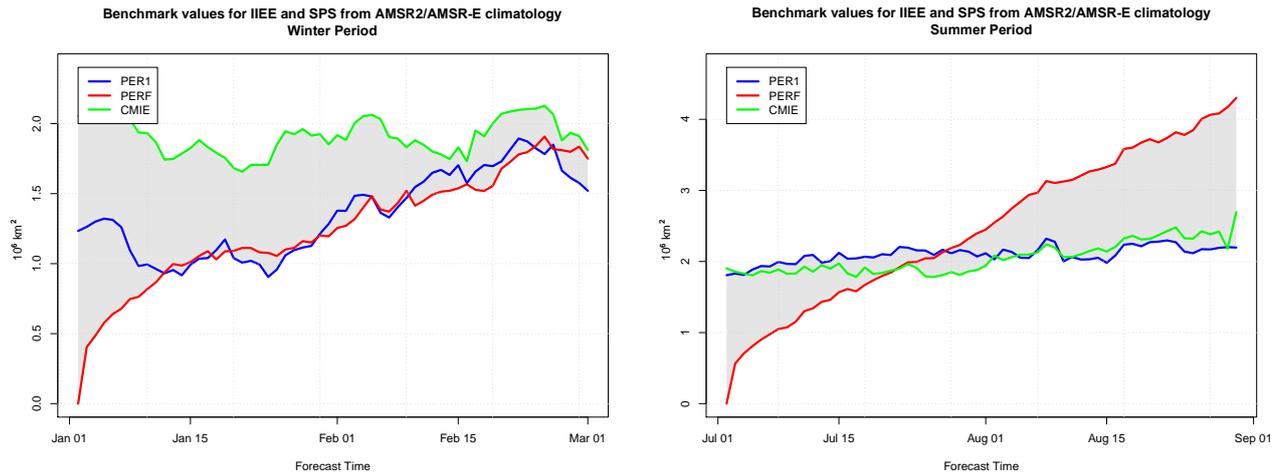
Similarly to the IIEE, the SPS can also be normalized and formulated as average distance between the observed and forecasted ice edge. This is done simply dividing the SPS by the total (climatological) length of the ice edge. The new formulation of the verification metrics, which will be called NSPS in the continuation of the thesis, is more easily understandable and therefore of interest for potential final users.

2.4 | Benchmark values for IIEE and SPS

To assess the correctness of a forecast, the previously described metrics are certainly relevant, but not straightforward to interpret without a set of reference values. Basically, the main idea behind predictability is the possibility of obtaining additional information from the forecasts, compared to the information easily derivable from the climatological records. Therefore, a set of reference climatological values for IIEE and SPS must be derived. Different options have been investigated and four particular strategies appeared to be particularly suitable.

1. **Persistence from the previous year (PERI):** the IIEE is calculated between the observed sea ice concentration of a certain day and sea ice concentration of one year before. Since this approach is based on single observations and thus the state of the system is deterministic and not probabilistic, the IIEE coincides with SPS.
2. **Persistence from forecast initial time (PERF):** the IIEE is calculated comparing the observed sea ice concentration to the observed state at the beginning of the forecast, for the whole forecast period. We expect a perfect agreement at the beginning of the forecast and a consequent increase of IIEE in the next days. Again the IIEE coincides with the SPS, for the same reasons already mentioned.
3. **Climatological median ice edge (CMID):** this approach consists in defining a climatological median ice edge that is derived from observations of previous years, as already described in Sect. 2.2. Afterwards, the IIEE between the observed sea ice concentration and the climatological edge is calculated.
4. **Climatological spatial probability score (CSPS):** this is the equivalent of the previous point in the frame of spatial probability distributions. We basically calculate the probability field for each day, based on the climatology. Afterwards we apply the SPS, comparing a certain observation with the climatological sea ice probability valid at the observation date.

Two examples of benchmark values are shown in Fig. 2.2a and 2.2b. In this case, the benchmark values are calculated for two forecasts initialized on 01/01/2016 (winter period) and on 01/07/2016 (Summer period). The gray area in the plots represents the range between the maximum and minimum benchmark value on a daily base. A useful test to assess the quality of the forecasts consists in comparing the IIEE or SPS with the benchmark values as follows. If the IIEE



(a) Benchmark values for a forecast initialized on 01/01/16 (winter period). (b) Benchmark values for a forecast initialized on 01/07/16 (summer period).

Figure 2.2: Examples of PER1, PERF and CMIE benchmark values.

or SPS are lower than the benchmark values and therefore they lay below the gray area, this means that the forecast can predict the sea ice evolution better than the climatology or the persistence. If, on the other hand, the IIEE or SPS is higher than the benchmark values, intercepting the gray area or being above it, this means that a better forecast for the sea ice evolution can be obtained by considering the climatology or the persistence of the observations, which in this case are more skillful than the model results.

2.5 | Hausdorff Distance and Modified Hausdorff Distance

Hausdorff Distance (**HD**) is a topological method to describe shape matching [Dukhovskoy et al. (2015)]. Such methods use a metric distance between two objects, in our case the sea ice edge, as measure of shape similarity. A definition of HD is given in Eq. 2.9. We called our two sets of points A and B, which correspond to the observed ice edge and the forecasted ice edge in our study. The notations *inf* and *sup* stand for the topological operators infimum and supremum. The concepts of infimum and supremum are similar to minimum and maximum, but are more useful in this context from a formal mathematical viewpoint because they are defined also for special sets which may have no minimum or maximum. An exhaustive treatise of the topic can

be found in any analysis book.

$$\text{HD}(A, B) = \max \left\{ \sup_{a \in A} [d(a, B)], \sup_{b \in B} [d(A, b)] \right\} \quad (2.9)$$

$$\text{with } d(a, B) = \inf_{b \in B} [d(a, b)] \quad (2.10)$$

$$\text{and } d(A, b) = \inf_{a \in A} [d(a, b)] \quad (2.11)$$

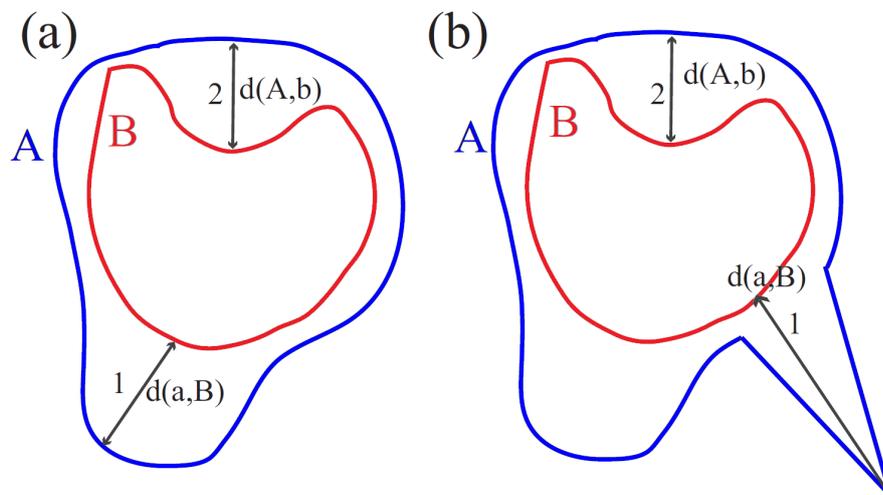


Figure 2.3: The figure shows two contours A and B. In the right side (b) an outlier point is added to A. This causes a relevant increase of $d(a, B)$ and consequently, and increase of HD. Figure from [Dukhovskoy et al. \(2015\)](#).

A relevant aspect is that this formulation does not set any constraint on the distance itself. This can be the most appropriate distance depending on the application (e.g. Euclidean distance or great circle distance). In our case we define d as the great circle distance. In general, a small HD indicates a good agreement between the two sets A and B. However, this method is sensitive to outliers as shown in Fig. 2.3. This is an undesirable property for the verification of the sea ice edge, since outliers are quite common in sea ice distributions. To overcome the problem of sensitivity to outliers, a more resistant metric is introduced following [Dukhovskoy et al. \(2015\)](#), the Modified Hausdorff Distance (MHD).

Generally, MHD is a valuable concept because it is very intuitive and allows comparison between different regions. However, this is much more demanding to compute compared to the

other metrics previously presented. MHD can be expressed as:

$$\text{MHD}(A, B) = \max \left\{ \frac{1}{|A|} \sum_{a \in A} d(a, B), \frac{1}{|B|} \sum_{b \in B} d(A, b) \right\} \quad (2.12)$$

$$\text{with } d(a, B) = \inf_{b \in B} [d(a, b)], \quad (2.13)$$

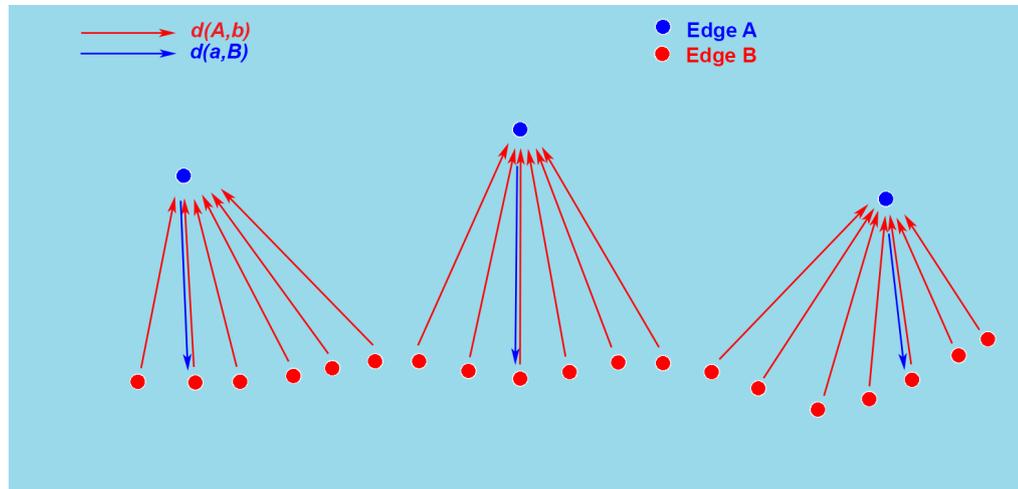
$$\text{and } d(A, b) = \inf_{a \in A} [d(a, b)]. \quad (2.14)$$

$|A|$ and $|B|$ are cardinalities of the sets A and B , in our case the number of points used to approximate the ice edges. A further advantage of MHD (and HD) is the ability to operate on contours or surfaces that have a different number of points and with no point-to-point correspondence required.

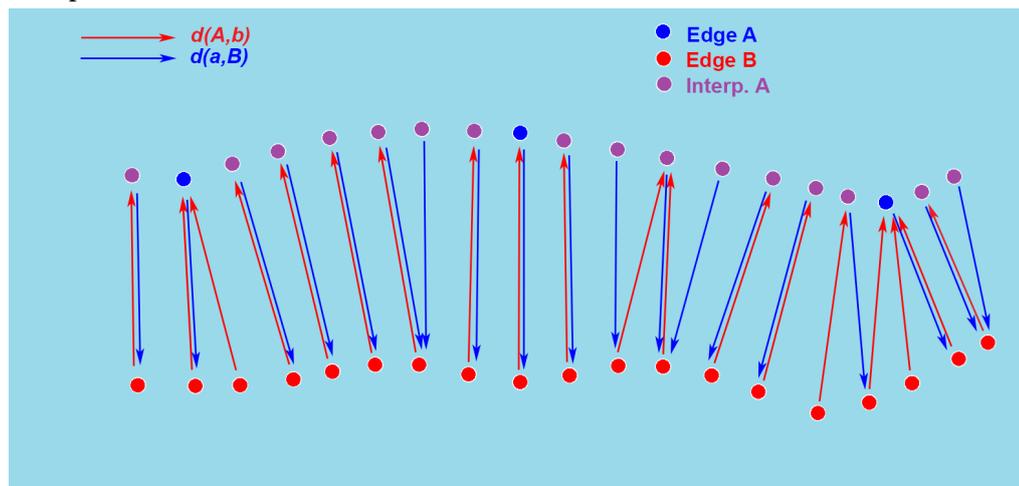
Sensitivity tests and realistic application to sea ice fields have shown that MHD is a meaningful metric for sea ice application. However, the MHD exhibits some relevant limitations when used as verification metric for sea ice forecasts. These limits are summarized in the following points.

1. The MHD computation is quite demanding, even with a proper parallelization of the algorithm. A superficial and certainly not elegant approach consists in neglecting some points in one or both the edges, especially for those edges with a higher spatial step than the opposite one or in case the ice edge length is much higher than the spatial step. However, neglecting some points affects most of the time the accuracy of the MHD calculations.
2. A different resolution for the forecasts and the observations implies a different spatial step of the two edges. This introduces an overestimation of the MHD. A possible solution consists in equalizing the spatial step of the two edges by adding new interpolated points in the edge with the lower resolution. This procedure is schematically illustrated in Fig. 5.3. In the example, the edge A (blue points) has a lower spatial step compared to the edge B (red points). Fig. 2.4a outlines the usual calculation of the MHD without any correction, while in Fig. 2.4b additional interpolated points (purple dots) are added to the edge A, equalizing the spatial step of B with the spatial step of A. In the second sketch, the MHD provides a more accurate estimation of the mean distance between the two ice edges, due to the higher perpendicularity between the distance vectors (arrows) and the edges themselves. However, this correction strategy has a further negative impact on the computation time of the met-

ric, because more points are considered in the calculations. Furthermore, the addition of interpolated points requires that the elements of the edge A are sequentially ordered, which is not a requirement for the simple calculation of the MHD.



(a) Representation of the direct MHD calculation, without the addition of interpolated points.



(b) Representation of the MHD calculation with a corrected edge A, through the addition of interpolated points in between its elements.

Figure 2.4: Illustration of a possible correction strategy in case the spatial steps of the two edges are considerably different.

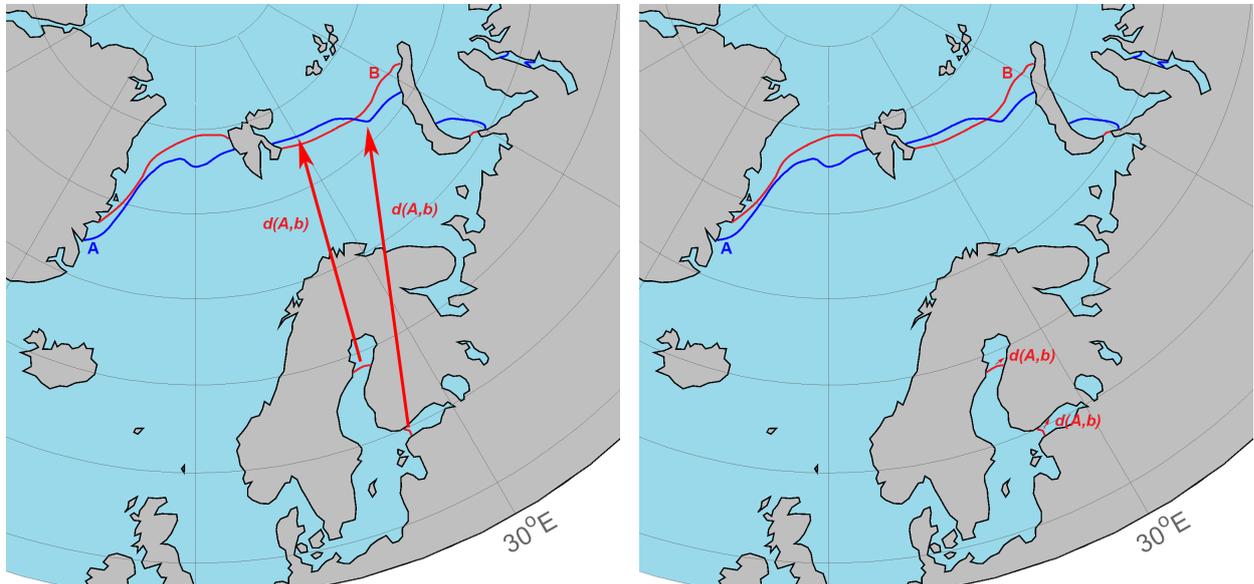
3. Even though the MHD is designed to be more resistant to biases induced by outliers, those still affects this metric in particular cases. As sketched in Fig. 2.5a, some sea ice can be hypothetically observed in the Baltic Sea (edge B) while not being correctly forecasted by the model (edge A). We will refer as b_i to an arbitrary point of the edge B located in the Baltic Sea. In this situation, the point a_j of the edge A against which $d(A, b_i)$ is calculated, will likely be either in the Barents Sea or in the Kara Sea, at a distance of thousands of kilometers

from b_i . This configuration has a non negligible impact on the final result, especially if several points of B present this specific behavior. An option to avoid this situation is the inclusion of the coastline as part of the edge A. In this case, the point a_j against which $d(A, b_j)$ is calculated is the closest point of the coast line, reducing substantially the "errors" induced by outliers. This second situation can be observed in Fig. 2.5b. However, including the coastline generates further problematic events, in particular when the sea ice is strongly underestimated by the models during summer. This situation is schematically represented in Fig. 2.5c. Since most of the distances are calculated in respect to the coastlines, the MHD is not a reliable estimation of the mean distance between the two edges but rather reflects the mean distance between one of the edges and the coast.

To conclude, obtaining meaningful results when the MHD is employed as verification metric for the sea ice edge position is not straightforward and requires a high level of critical evaluation of the results. These are easily affected by non-predictable and non-detectable issues that can alter the meaning of the MHD itself.

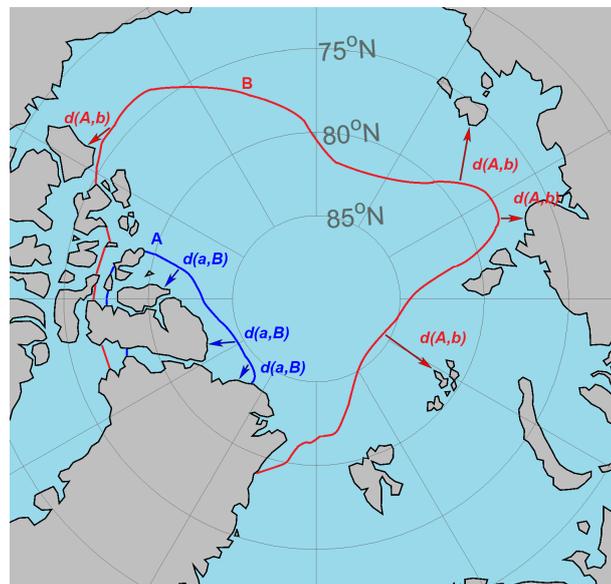
Some routines for the detection of the sea ice edge and for the calculation of the MHD have been developed as part of the R package `spherlab`¹ in the frame of the thesis work.

¹The package can be downloaded from the GIT repository <https://github.com/FESOM/spherlab>



(a) Schematic representation of the MHD calculation without the addition of the coastline.

(b) Schematic representation of the MHD calculation with the addition of the coastline.



(c) Schematic representation of a wrong MHD calculation induced by the introduction of the coastline.

Figure 2.5

Models and satellite data

In this section, we present the datasets analyzed in the thesis work. Firstly, we focus our attention on the sea ice forecasts and analyze the features of the models employed. In particular, we describe in detail the models contributing to the Sub-seasonal to Seasonal Prediction Project (S2S), which represent the central dataset of forecasts employed in our research. Secondly, the attention is moved to the satellite observations used to verify the forecasts. Different sea ice concentration products are compared, emphasizing particularly the features of [ASI](#) algorithm and its sea ice concentration products. An overview on the different possible reasons for model-observation disagreement concludes this section.

3.1 | S2S forecasts models

The sea ice forecasts analyzed in this paper are those collected in the database of the Sub-seasonal to Seasonal Prediction Project. As explained in the introduction, most of the focus of S2S is concentrated on atmospheric phenomena with a regional emphasis on lower latitudes. For this reason, only a part of the contributing institutes are coupling their atmospheric models with a dynamic sea ice model, or even with an ocean model. In [Tab. 3.1](#) the main features of the analyzed models are presented [[Vitart et al. \(2016\)](#)]. Note that the information reported in the table are valid only for the real time forecasts released between 01/2015 and 06/2017. A broader description of the forecasts and reforecasts will be given in [Sec. 3.3](#).

We clearly observe that the contributing models have quite different features. This lack of

homogeneity is due to the fact that the S2S database is a database of "opportunities", meaning that the forecasts included are not following a coherent protocol and are not designed specifically for this project. The key disagreement elements are summarized in the following points.

- The forecasts lead time varies between the models from 32 to 60 days.
- The initialization frequency and the ensemble size vary considerably. Most models are running just once or twice a week producing a robust ensemble size. This is the case of BoM, ECCC, ECMWF, HMCR, ISAC-CNR, JMA and Meteo-France. Some models are running on a daily basis with a reduced ensemble size. Examples are CMA, KMA, NCEP, and UKMO, which however ensure a better temporal coverage.
- The complexity of the models is different, both regarding the spatial resolution and the coupled components.

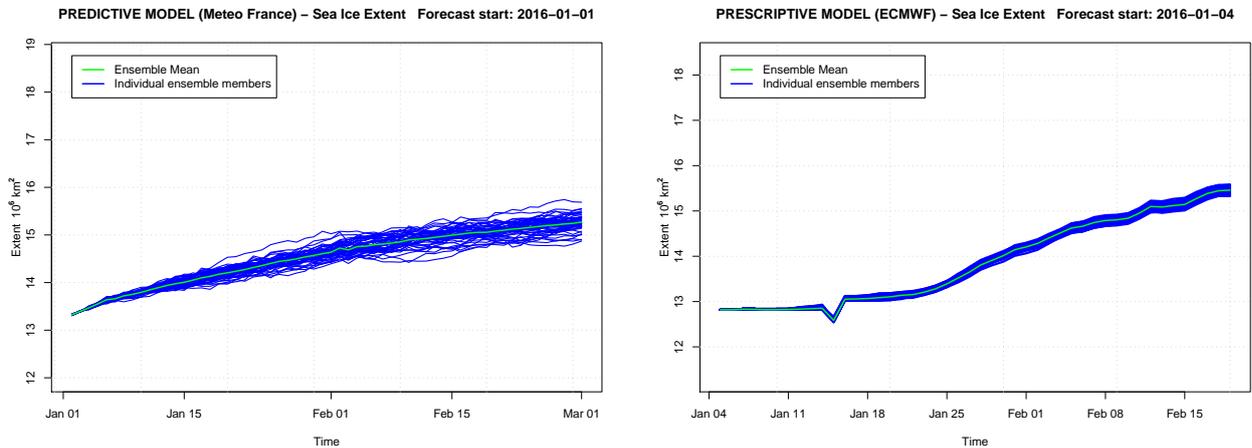
Table 3.1: Summary of the S2S models features.

Name	Atmosphere	Ocean Coupl.	Sea Ice Coupl.	Frequency	Ens. Size	Length
BoM	✓	✓		twice a week	32+1	62 days
ECCC	✓			weekly	20+1	32 days
HMCR	✓			weekly	40+1	61 days
ISAC-CNR	✓			weekly	40+1	31 days
JMA	✓			twice a week	49+1	33 days
CMA	✓	✓	✓	daily	3+1	60 days
KMA	✓	✓	✓	daily	3+1	60 days
ECMWF	✓	✓	✓ ¹	twice a week	50+1	46 days
MetFrance	✓	✓	✓	monthly/weekly	50+1	61/32 days
NCEP	✓	✓	✓	daily	15+1	44 days
UKMO	✓	✓	✓	daily	3+1	60 days

The last aspect has a particular relevance for our research. Indeed we expect more interesting results from those forecasts which includes dynamic sea ice models as detailed explained in the following paragraphs.

¹ECMWF switched from prescribed sea ice to a dynamic sea ice model on 11/24/2016. Thus, both the versions will be considered in our analysis.

As basic distinction, we define predictive sea ice forecasts as those that include a coupled sea ice model that is predicting the evolution in time of sea ice variables, whereas we define prescriptive sea ice forecast as those where the sea ice evolution is simply based on persistence and/or climatological records. This difference is illustrated in Fig. 3.1a and 3.1b.



(a) An example of predicted sea ice produced by the Meteo France model.

(b) An example of prescribed sea ice produced by the ECMWF model

Figure 3.1: Evolution of pan-Arctic sea ice extent in predictive and prescriptive forecasts.

The Meteo France model and the prescriptive version of the ECMWF model have been chosen on purpose because they both feature 51 ensemble members. Sea ice extent is in this case just an example variable to illustrate the concept.

On the left hand side, the predictive sea ice extent forecast shows an independent evolution for each ensemble member: a spreading of the ensembles in time can clearly be observed. This behavior is consistent with the strong dependence of the system state on the initial conditions typical of the chaotic systems. The same behavior can not be observed in the second case. Indeed, when the sea ice is prescribed, this operation is done in the same way for all the ensemble members, with a possible artificial spreading of the ensemble which however is not as chaotic as for the probabilistic forecasts. In this situation, having a large ensemble is not relevant for the sea ice, the information carried from the different ensemble members is the same. This makes the probabilistic analysis described in Section 2.3 meaningless. In this case, the SPS and the IIEE are the same.

To better understand the models' behavior, an investigation of the strategies applied by the different modelling centers is needed. We start analyzing the prescriptive models listed in the first part of Tab. 3.1, following the work of [Mladek et al.].

3.2 | Prescriptive models

BoM

The sea ice is prescribed to vary with a climatological seasonal cycle. We can not see an example of the prescription strategy, because the model output relative to the sea ice is not correctly saved, as reported in the S2S website. Fig. 3.2a clearly shows the error. The sea ice extent seems to follow a square wave function constant during the whole year. The extent varies between $1.85 \times 10^7 \text{ km}^2$ and $2.35 \times 10^7 \text{ km}^2$ with a frequency of approximately 10 days.

ECCC

The sea ice concentration is inferred from the sea surface temperature (SST) in the polar oceans. An exemplary evolution of the sea ice extent trend is plotted in Fig. 3.2b. Since the SSTs are prescribed identically for all ensemble members, the sea ice is identical in all ensemble members, too. Further details about the relaxation strategy of the SST are not available.

ECMWF - 1st Version

In this thesis, we refer to the ECMWF - 1st Version meaning the forecast system that has been run prior to 11/24/2016, with a prescribed sea ice. The sea ice initial conditions are persisted up to day 15 and then the state is relaxed to climatology up to day 45. Furthermore, in an updated version of the model (data from January 2016 to November 2016) the initial conditions are relaxed towards different sea ice conditions. These behaviors are clearly visible in Fig. 3.2c.

HMCR

According to the S2S documentation, sea ice initial conditions are persisted up to day 15 and then relaxed to climatology up to day 45. Looking at Fig. 3.2d, we observe a different behavior of the forecast. Sea ice seems to be immediately relaxed to the climatology during the first 3-5 days. The research group will be contacted to clarify this discrepancy.

ISAC-CNR

Sea ice is held fixed if sea ice fraction is above the climatological value in the fall-winter season that is, during the sea ice growth season, or below it during spring-summer season, that

is, during the sea ice melt season. Sea ice is relaxed to climatology otherwise. We can observe an example for the sea ice extent evolution in Fig. 3.2e. The ensemble members in this case are not completely overlapping because the sea ice is initialized with slightly different initial conditions and relaxed towards the climatology. It is not clear which relaxation strategy is applied, thus, we will contact the group to clarify this aspect.

JMA

The prescription strategy is based on SSTs and thus similar to the ECCO model. An example is reported in Fig. 3.2f.

The main focus of our research is the verification of the predictive models. Thus, the prescriptive results provided by the previous systems will not be extensively analyzed in this thesis work. An exception is the sea ice prescribed by ECMWF. This will be compared with the updated version of the same model, with the dynamic sea ice component, to reveal possible strengths and weaknesses of the two versions.

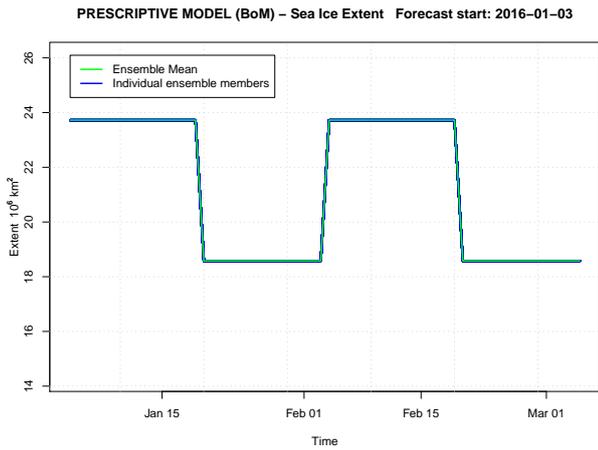
3.3 | Predictive Models

CMA

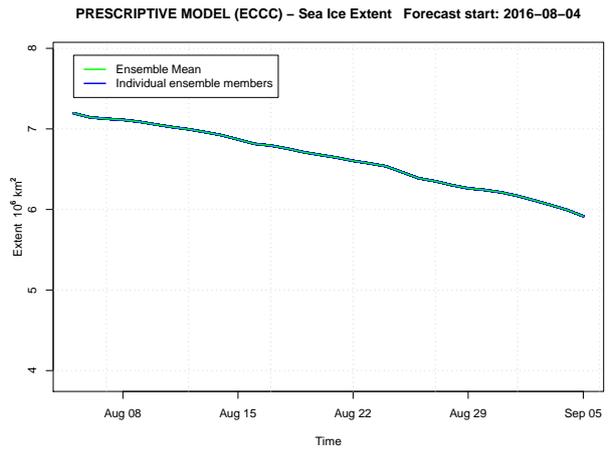
The CMA model is coupled with a sea ice model from day 0. The sea ice model is the GFDL Sea Ice Simulator (SIS) [Winton (2000)] with the same horizontal resolution as the ocean model (MOM4 [Griffies et al. (2008)] with 0.33°-1° horizontal resolution and 40 vertical levels). Sea ice initial conditions come from a Coordinated Initialization System (CIS).

ECMWF - 2nd Version

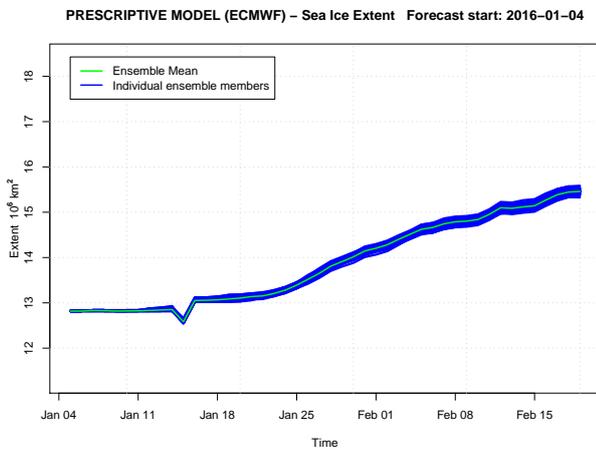
The ECMWF model - 2nd Version, which has been run since 11/24/2016, includes an interactive sea ice model (the Louvain-la-Neuve Sea Ice Model - LIM2 [Bouillon et al. (2009)]). The initial perturbations of sea ice are applied to five ensemble members of ensemble-based analyses (and reanalyses). There are no stochastic perturbations of sea ice included. The ocean model is NEMO3.4.1 [Madec (2008)] with a 0,25 degree horizontal resolution, 75 vertical levels, initialized from ECMWF Ocean Analysis. Four perturbed analyses are produced by perturbing the wind field during the ocean analysis. The frequency of coupling is hourly.



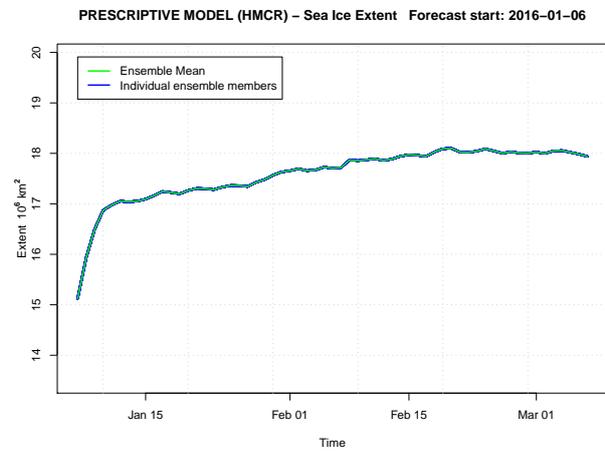
(a) Prescription strategy of the BoM model.



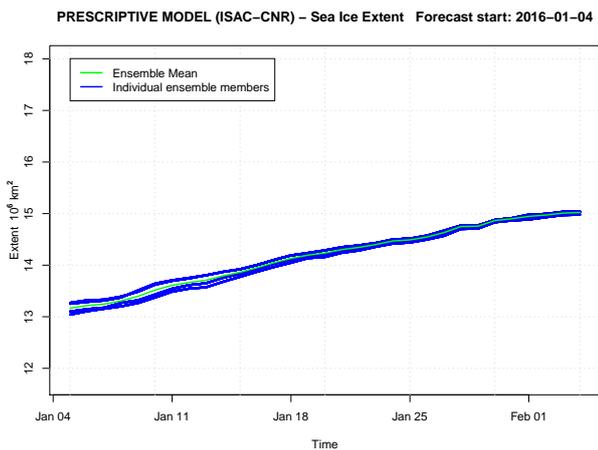
(b) Prescription strategy of the ECCC model.



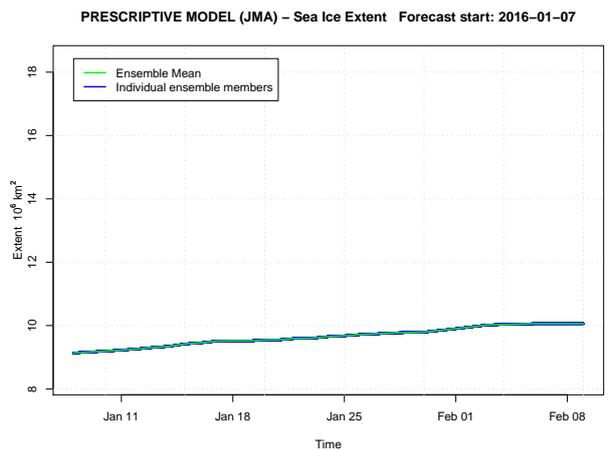
(c) Prescription strategy of the ECMWF model.



(d) Prescription strategy of the HMCR model.



(e) Prescription strategy of the ISAC-CNR model.



(f) Prescription strategy of the JMA model.

Figure 3.2: Note that all the forecasts reported in this panel, except the ECCC model, are initialized at the beginning of January 2016. The reason is that ECCC model started providing the data since July 2016.

KMA

The KMA model is coupled with a sea ice model from day 0. The sea ice model is CICE4.1 [Hunke and Lipscomb (2010)], initialized from Met Office Analysis (NEMOVAR) [Mogensen et al. (2012)]. The ocean model is NEMO3.4 with 0.25° degree horizontal resolution and 75 vertical levels. This is also initialized using NEMOVAR, without perturbations. The frequency of coupling is 3-hourly.

Météo France

The Meteo France model is coupled with a sea ice model from day 0. The ocean and sea ice model is NEMO3.2 with a 1° horizontal resolution and 42 vertical levels. The model is initialized from an unperturbed MERCATOR-OCEAN Ocean and Sea-ice Analysis. The frequency of coupling is 24-hourly.

NCEP

The NCEP model is coupled with a sea ice model from day 0. The sea ice model is part of the ocean model GFDL MOM4 which has a spatial resolution in the zonal direction of 0.5° and in the meridional direction, 0.25° from 10° S to 10° N, progressively decreasing to 0.5° from 10° to 30°, and is fixed at 0.5° beyond 30° in both hemispheres. There are 40 levels in vertical direction.

UKMO

The UKMO model is coupled with a sea ice model from day 0. The sea ice model is Global Sea Ice 6.0 (CICE4.1), initialized from NEMOVAR without perturbations. The ocean model is Global Ocean 5.0, based on NEMO3.4 with 0.25° degree horizontal resolution and 75 vertical levels. This is also initialized using NEMOVAR, without perturbations. The frequency of coupling is 3-hourly.

All the models are regridded to a common 1.5°×1.5° grid. For this reason, the original resolution of the models reported in the previous description is not available. This choice, made for reducing the data storage required for the numerous global fields that are part of the S2S database, causes an inevitable loss of information and has a negative impact on our analysis.

The S2S database collects real-time forecasts since the beginning of 2015, making them

public after four weeks from the forecast initialization day. The operational centers make available several probabilistic reforecasts, also called hindcasts, which are forecasts run using the same model version as the real-time forecast for a number of past dates. Different operational centers use different configurations for the implementation of their reforecast system. At the moment, the following two configurations are used: fixed configuration (FIX) and on-the-fly configuration (FLY). For the fixed configuration, the reforecasts for all past dates are produced once during the lifetime of a given model version and for when a new version of the model is released, a new set of reforecasts is produced. For the on-the-fly configuration, the reforecasts are produced at the same time as the real-time forecasts and every reforecast refers to a real-time forecast.

A summary of the predictive models features is reported in Tab. 3.2 (real time forecasts) and Tab. 3.3 (reforecasts).

Table 3.2: S2S predictive systems - Real-time Forecasts

Model Name	Forecasts Period	Frequency	Ens. Size	Forec. Length
CMA	01/01/2015 - present	daily	3+1	60 days
KMA	01/11/2016 - present	daily	3+1	60 days
ECMWF	24/11/2016 - present	twice a week	50+1	46 days
MetFrance	01/05/2015 - present	monthly/weekly ²	50+1	61/32 days
NCEP	01/01/2015 - present	daily	15+1	44 days
UKMO	01/12/2015 - present	daily	3+1	60 days

Table 3.3: S2S predictive systems - Reforecasts

Model Name	Hindcasts Period	Frequency	Ens. Size	Forec. Length	Config.
CMA	01/01/1994 - 31/12/2014	daily	3+1	60 days	FIX
KMA	01/01/1991 - 25/12/2010	four times a month	2+1	60 days	FLY
ECMWF	24/11/1996 - 01/05/2016	twice a week	10+1	46 days	FLY
MetFrance	01/01/1993 - 15/12/2014	biweekly	14+1	61 days	FIX
NCEP	01/01/1999 - 31/12/2010	daily	3+1	44 days	FIX
UKMO	01/01/1996 - 01/07/2015	four times a month	6+1	60 days	FLY

Overall, the period covered by forecasts and reforecasts goes from 1991 to the present. In this research, the analysis is limited to the 15 years period between June 2002 and June 2017,

²Before 03/2016: monthly frequency with 60 days length. Since 03/2016: weekly frequency with 32 days length.

because the observational dataset chosen to verify the forecasts was available since June 2002, as described in Sec 3.5.

3.4 | Further sea ice forecast products

In addition to S2S database, a number of other suitable forecasts products is available to test the sea ice forecast skill and predictability. Even if an extensive analysis of those systems is not part of our research, a brief description of those systems is included to underline the differences between those forecasts and the S2S products.

Sea Ice Outlook Predictions

The Sea Ice Outlook (SIO) is an initiative organized by the Sea Ice Prediction Network (SIPN). Since 2008, this has solicited predictions of the mean September sea ice extent from the Arctic research community. The purposes of this initiative are mostly linked to the academic research on Arctic sea ice prediction and thus does not aim to produce forecasts for potential final users. The predictions collected by this panel are based on different approaches: modeling, statistical, and heuristic and are provided in three cycles each year, around the first of June, July and August [Stroeve et al. (2014)]. The Sea Ice Outlook community is constantly growing and in the near future will probably focus its attention not only on the September minimum prediction but also on other periods in the year, especially on the autumn freezing season. The results of the predictions show a bimodal pattern of success. In those years when observed ice extent is following the climatological trend, the predictions tend to be on average accurate. However, when the observed extent is anomalous, most of the individual predictions are less accurate.

SIO originally focused on pan-Arctic extent, and that only some of the groups (using dynamic models) also provide spatial probability fields. One relevant difference between the SIO spatial probability fields and the S2S forecasts is the fact that SIO forecasts are limited to a single probability field for the month of September, without providing any additional information about the evolution of the sea ice during the summer. A further limitation of the SIO is that the verification of the predictions has been limited in the past to sea ice extent and area, mostly not considering the spatial distribution of the sea ice. An interesting aspect would be to apply the metrics described in the first part of the thesis paper to the model outputs contributing to the SIO. This would broaden the work of Frank Kauker (AWI) for the the last post-season report

where the SPS has in fact already been applied to the SIO forecasts. This could reveal potential sources of (spatial) sea ice predictability on different time scales. Furthermore, such a verification would certainly underline strength and weakness aspects of the individual models, which could be eventually improved and corrected.

GODAE Oceanview Sea ice Intercomparison

The Global Data Assimilation Experiment [GODAE](#) gathers the international ocean modeling and data assimilation communities around global ocean high resolution forecast systems. The project defined the "Class 4" metrics for global ocean, that refers to forecast verification which takes place in observation space. The institutes participating with their model to this projects are UK Met Office (UK), Bureau of Meteorology (AUS), [NOAA/NWS/NCEP/EMC/MMAB](#) (U.S.), Canadian Meteorological Center (CA) and REMO (BR).

An example for the models contributing to the GODAE Ocean View is the forecast system recently introduced by the Canadian Meteorological Centre, called the Global Ice Ocean Prediction System ([GIOPS](#)) [[Smith et al. \(2016\)](#)]. GIOPS provides daily global ice and ocean analyses and 10-day forecasts on a $1/4^\circ$ -resolution grid. GIOPS includes a multivariate ocean data assimilation system that combines satellite observations of sealevel anomaly and sea-surface temperature (SST) together with in-situ observations of temperature and salinity.

GIOPS model output that contributed to the GOV intercomparison project are not provided on a regular grid as the usual model data. Thus, a complex verification analysis of these forecasts is practically difficult. Furthermore, the timescale of GIOPS forecast can not be considered as part of the sub-seasonal and seasonal range, which is the focus of our research.

3.5 | Satellite measurement of sea ice concentration

A relevant aspect of every verification study is the observation of the forecasted variable. Choosing an appropriate observation system is crucial for an effective and accurate verification. Since we are dealing with sea ice concentration forecasts on an hemispherical domain, the observations that suit our verification purposes are satellite measurements of sea ice concentration based on passive microwave sensors. These guarantee a daily global coverage of both hemispheres with a good horizontal resolution, also under cloud coverage conditions and during polar night that make impossible active measurements of sea ice. In the following paragraphs we analyze main

features of satellite based sea ice concentration measurements, emphasising the effects that this choice entails for the forecast verification.

The main advantage of satellite measurements compared to in-situ measurements is the regularity and density of the observations. The basic concept behind this measuring technique is to convert satellite measurements of radiance to geophysical quantities, which in our case is sea ice concentration. However, in-situ observations are still needed to calibrate the algorithms that perform the conversion. The most widely used passive microwave instruments for deriving sea ice concentration variables are SMMR, SSM/I, SSMIS, AMSR-E and AMSR-2, which are operated on various satellite platforms located on quasi polar orbits.

The algorithms applied to the microwave brightness temperatures use different combinations of frequency channels, making different corrections for weather, satellite drift and other disturbances. The two satellite algorithms for sea ice concentration that are most widely used for model-data intercomparison studies are the Bootstrap algorithm [Comiso (1986)] and the NASA Team algorithm [Cavalieri et al. (1984)]. Comiso et al. (1997) analyze the major differences between the two algorithms. During the melting season sea ice concentration derived from the Bootstrap algorithm is probably closer to the real sea ice concentration conditions than that from the NASA Team algorithm. The latter has been shown to underestimate the Arctic sea ice extent when compared to independent observations [Partington et al. (2003)]. In contrast, the Bootstrap algorithm provides estimations of sea ice concentration that are very close to the “Climate Data Record of Passive Microwave Sea Ice Concentration”, a product merging products from different algorithms. All sea ice concentration algorithms are generally affected by the presence of melting ponds over ice during summer. These are frequently confused as open water and thus cause an underestimation of the sea ice concentration.

A third algorithm for sea ice concentration estimation is the ASI algorithm [Spreen et al. (2008)], which has been considered for the verification analysis of this thesis paper. This algorithm provides sea ice concentration based on SSM/I data from 1991 onwards, on AMSR-E data from June 2002 and AMSR-2 data from 2012. The period considered to assess the climatology is of 11 years only, from 2003 to 2014, in accordance with the availability of the AMSR-E and AMSR-2 data. This is shorter than the 30 years usually applied by the scientific community to define a climatology. However, considering the variability and trend in the state of the ice during this period, the sample is enough and should in fact not reach too far into the past to build a climatology that meets our needs. The channel used by the algorithm is at higher frequency (86 GHz) com-

pared to Bootstrap and NASA Team. The measurements are thus more sensitive to atmospheric disturbances and require sophisticated weather filters. However, this is a reasonable price to pay since the ASI algorithm is able to increase the spatial resolution to 6.25 km instead of 25 km for the Bootstrap and NASA Team algorithms. A high resolution of the observations leads to a better estimation of the "true" sea ice edge and consequently to an effective verification of the forecasts. While this is of subordinate importance for the verification of the coarse-resolution S2S forecasts, it will be an advantage for the verification of higher-resolution forecasts such as the GODAE Ocean View forecasts.

The three algorithms generally agree in describing features of sea ice coverage in the Arctic. A particularly good agreement is observed between Bootstrap and ASI during the melting season, see Fig. 3.3. A question that arises is the opportunity of combining different algorithm products to estimate the sea ice edge. However, considered the low resolution of our forecasts and the good agreement between algorithms, the choice of using only one algorithm product for verification purposes seems reasonable.

3.6 | Contributions to model-observation disagreement

Beside the inevitable forecast error growth brought about by the chaotic nature of the atmosphere-ocean-sea ice system, a number of different factors can cause disagreement between observations and forecasts. In the following points, we list the critical aspects that could have a negative influence on a forecast process:

- Errors or low complexity of the forecast model.
- Low spatial resolution of the model.
- Errors in the initialization of atmospheric, oceanic and sea ice conditions. Particularly, the eventual initialization of the sea ice by assimilation of a different satellite product leads to an initial bias in the forecast.
- Inaccuracies in the model land-sea mask, for example in complex coastal regions such as the Canadian Archipelago. This can be partially overcome by applying a specific mask to both observations and forecasts.
- Further resolution decrease caused by regridding for data distribution purposes.

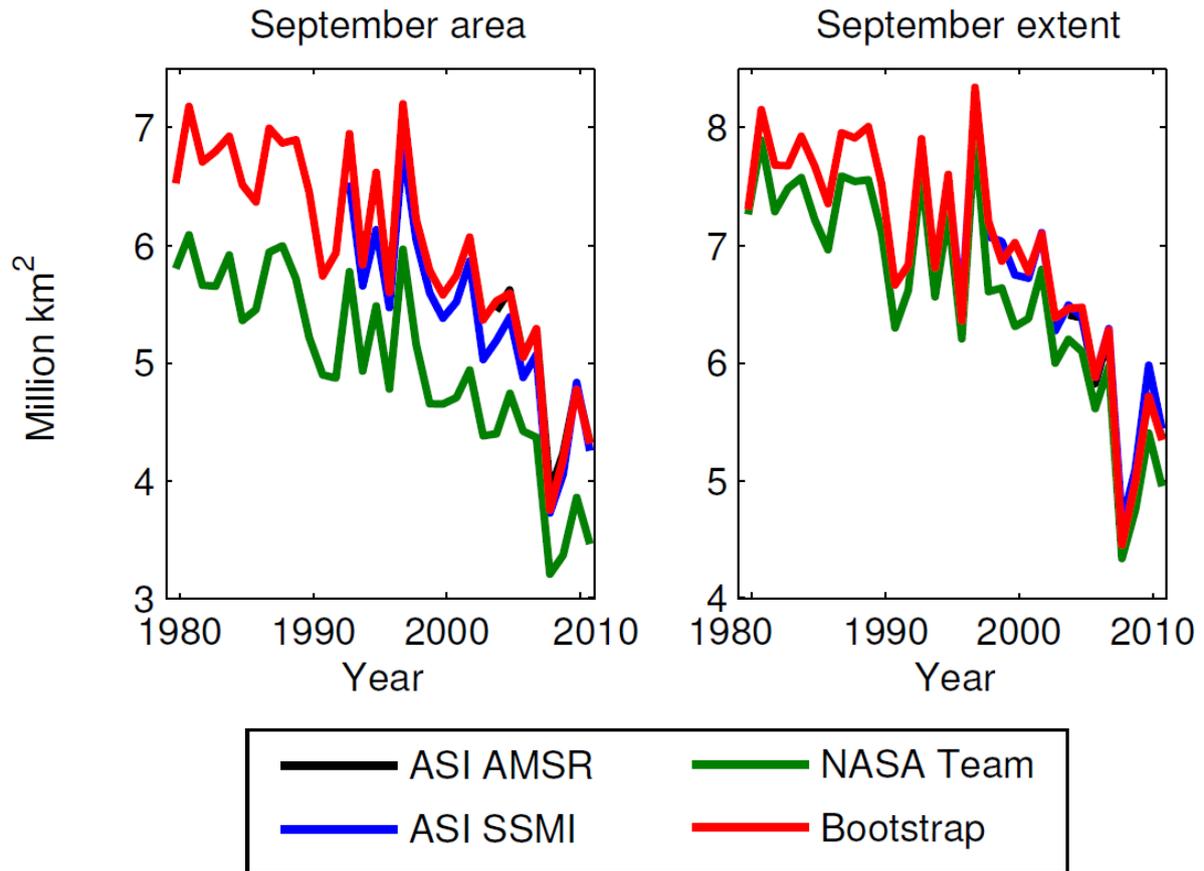


Figure 3.3: The trend of the September minimum sea ice area is displayed on the left graph while the right graph shows the trend of the minimum extent. Note that, because the ASI AMSR data accurately match the Bootstrap data, the black curve is almost invisible. Figure from [Notz \(2014\)](#).

In addition to these error sources linked to the forecast system, the satellite measurements are also affected by errors which have been already briefly discussed in the previous section. To summarize, the main contributions are the atmospheric influences (mainly storms), the melting ponds recognized as open water and the coastal regions detected as sea ice. These contributions influence the observed position of the sea ice edge and introduce errors sources into the verification metrics.

Verification of the S2S sea ice forecasts

This section describes the verification analysis of the S2S sea ice forecasts, focusing in particular on the sea ice edge position as relevant variable for final users. Our main findings, regarding both the behavior of the verification metrics and the performances of the models, are extensively described combining the results from the verification metrics with the direct visualization of the forecast and observation edges.

The verification of the forecasts is based on two different strategies. Firstly, Sec. 4.1 describes the verification of the forecasts against the observations based on satellite retrievals. Secondly, in Sec. 4.3 the forecasts are verified against virtual observations, which are built based on the control forecasts evaluated at the initial time of the forecast. We will refer to these virtual observations as models own analysis. The sea ice concentration derived from satellite data, which is widely employed in the first verification approach, is likely a better representation of the true state of the Arctic sea ice. However, this leads to a relevant discrepancy between the forecasted and observed state, already the beginning of the forecast. This is mainly a consequence of the data assimilation process, which initializes the observations into the forecast system. The use of models own analysis in the verification of the forecasts becomes therefore useful because, by definition, the own analysis initial state coincides with the initial forecast state. This second verification strategy will show how forecast error grows starting from zero, instead of exhibiting large errors already at the beginning of each forecast. However, we are aware of the fact that the close-to-zero errors close to initial time that this verification strategy gives are unrealistic. A similar approach is adopted by [Jung and Matsueda \(2016\)](#) for the verification of TIGGE atmospheric

forecasts. Furthermore, avoiding the use of satellite data allows an easy and extensive calculation of the MHD, which is much less demanding to compute at the model resolution or the even coarser S2S grid.

The verification metrics computation is entirely conducted with the forecast interpolated to the longitude and latitude coordinates of the observations, except for the MHD. Subsequently, grid points where either the observations or the forecast have missing data (due to the respective land-sea mask) have been masked. An alternative approach consists in defining a common mask for all the models, which would allow an easier comparison between the results from different models. However, the land masks of the six models considered in this study are quite different. In particular, the CMA model presents a quite restrictive land mask, meaning that it features more land than the other models. Therefore, the application of this mask to the other models would cause the loss of relevant informations about the predictive skills of the other systems. Carrying the verification analysis with six different land masks, one for each model, is considered the best option in this particular case.

The analysis do not distinguish between forecasts and reforecasts, even though, as already explained in Sec. 3.3, these could differ for forecasts lengths, ensemble sizes and initialization frequency. In total, a period of 15 years (06/2002-06/2017) is covered by most of the models.

4.1 | Verification against satellite observations

4.1.1 | Example verification of a single forecast

The result of a typical S2S sea ice forecast verification analysis is displayed in the top graph of Fig 4.1. The model is Météo France initialized on 01/01/2016, with 60 days forecast length and 51 ensemble members. The plot shows the evolution of verification metrics, the SPS and the IIEE, including the decomposition of the IIEE in ME and AEE. IIEE, ME and AEE are computed for the ensemble-median ice edge. The gray area in the background indicates the range of the benchmark values. As already explained in Sec. 2.4, when the verification metrics are below the benchmark values, the forecast estimates the sea ice position better than the climatology or the persistence. In this specific case, the Météo France model shows predictive skills in the second half of the forecast. The observed initial state provides a better description of the ice edge position for the 15-20 days after the initialization.

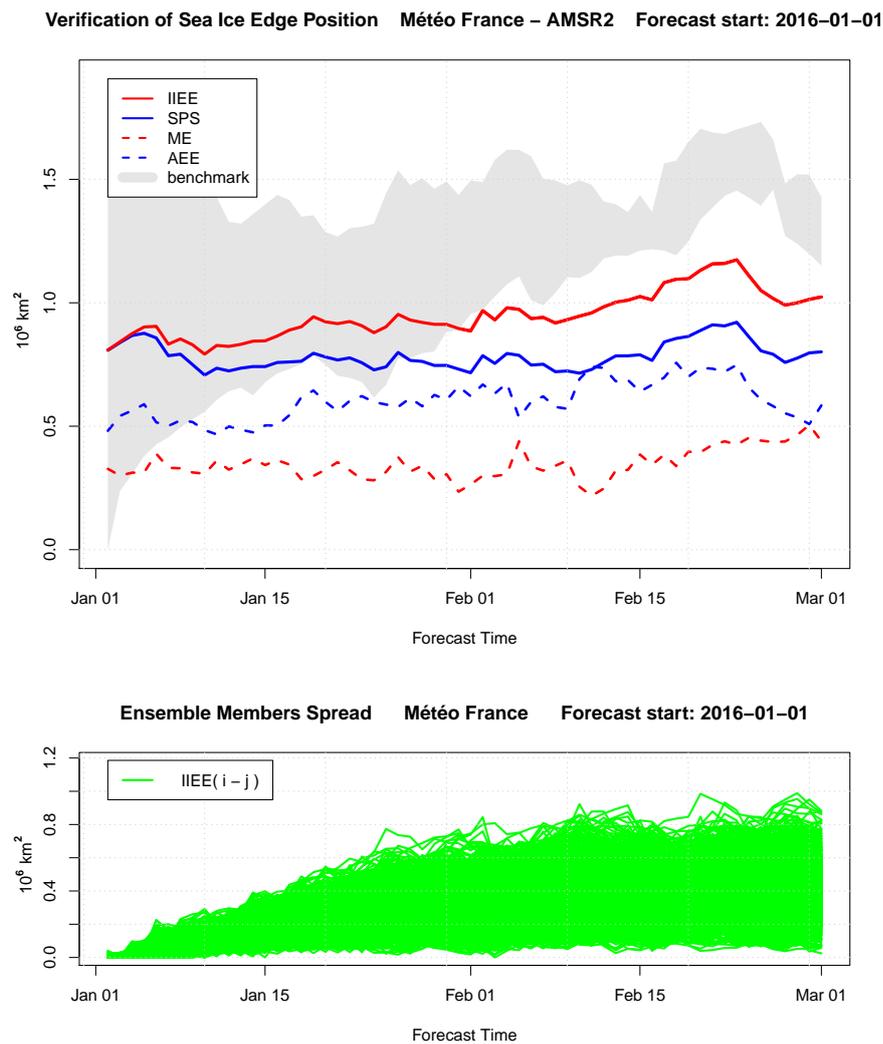


Figure 4.1: Result of one S2S sea ice forecast verification analysis. The model is Météo France and the verification metrics are IIEE, SPS, ME and AEE.

The SPS coincides with the IIEE at the beginning of the forecast, when all the ensemble members are equal. This is consistent with the formulation of the SPS given in Sec. 2.3. For the rest of the forecast, the SPS is increasingly less than the IIEE. Interestingly, even if the sea ice edge as we defined it in Chap. 2 is well determined object, a probabilistic description of this minimizes the error of the forecast. Thus, considering directly the probability derived from several perturbed ensemble members, instead computing a the median ice edge, reduces, to a certain extent, the error of the forecasts.

The bottom plot in Fig 4.1 shows the IIEE computed for all the possible pairs of ensemble members i, j . The clear feature that emerges from the graph is the overall increase of the ensemble spread as time passes. This clearly agrees with the non-linearity of the system and with its

strong initial condition dependence, which cause a rapid divergence of the model solutions.

A useful strategy to better understand the forecast evolution and to find eventual errors in the calculations, is the visualization of both forecast and observation on the same map. Consequently, we mapped together the sea ice probability and the observed sea ice edge, for each day of the forecast. This allows the creation of animations, which make the temporal evolution of our variables of interest highly accessible. In Fig. 4.2 we present four snapshots corresponding to the forecast previously analyzed. The maps confirm the evidence emerged from the verification analysis. The sea ice edge position is slightly overestimated in a systematic way. The shape of the edge is however well forecasted. A very simple potential bias correction strategy that could be applied to reduce the systematic overestimation of the edge, is the adjustment of the threshold value for the sea ice concentration. In particular, choosing a higher threshold value, as example 40% instead of the usual 15%, would cause a general retreat of the ice edge, ensuring a better agreement of the forecast with the observations. Further details about this and other potential bias correction strategies will be briefly discussed in the next sections of thesis.

4.1.2 | Extensive verification of the S2S database

Analyzing in detail a single forecast, as done in the previous paragraphs, is helpful to illustrate the concept and to understand how certain errors influence the verification metrics. However, the large amount of real time forecasts and reforecasts (from 400 to 5400 for each model, depending of the initialization frequency) requires a more adequate visualization strategy. Moreover, understanding recurrent errors and seasonal trends in the forecast systems is certainly not straight forward if all the attention is focused on a single case. In Fig. 4.3, we illustrate how multiple forecasts are merged together in a common plot. The idea is to extract the values from the verification metrics at some fixed reference days corresponding to a certain forecast lead time to construct a global time series. The reference days (day 1, day 8, day 18, day 32, day 44 and day 60) have been chosen to achieve a coverage as uniform as possible for all the forecast systems.

The next pages include the verification analysis of the forecasts formulated by the six predictive models in the S2S database. The verification is performed against the satellite observations. The verification metrics employed in this analysis are the SPS, the IIEE, the ME, and the AEE, with the last three computed for the ensemble-median ice edge. A slightly different definition of AEE is used in this context. In particular, $EE = O - U$ (Extent Error) instead of the usual

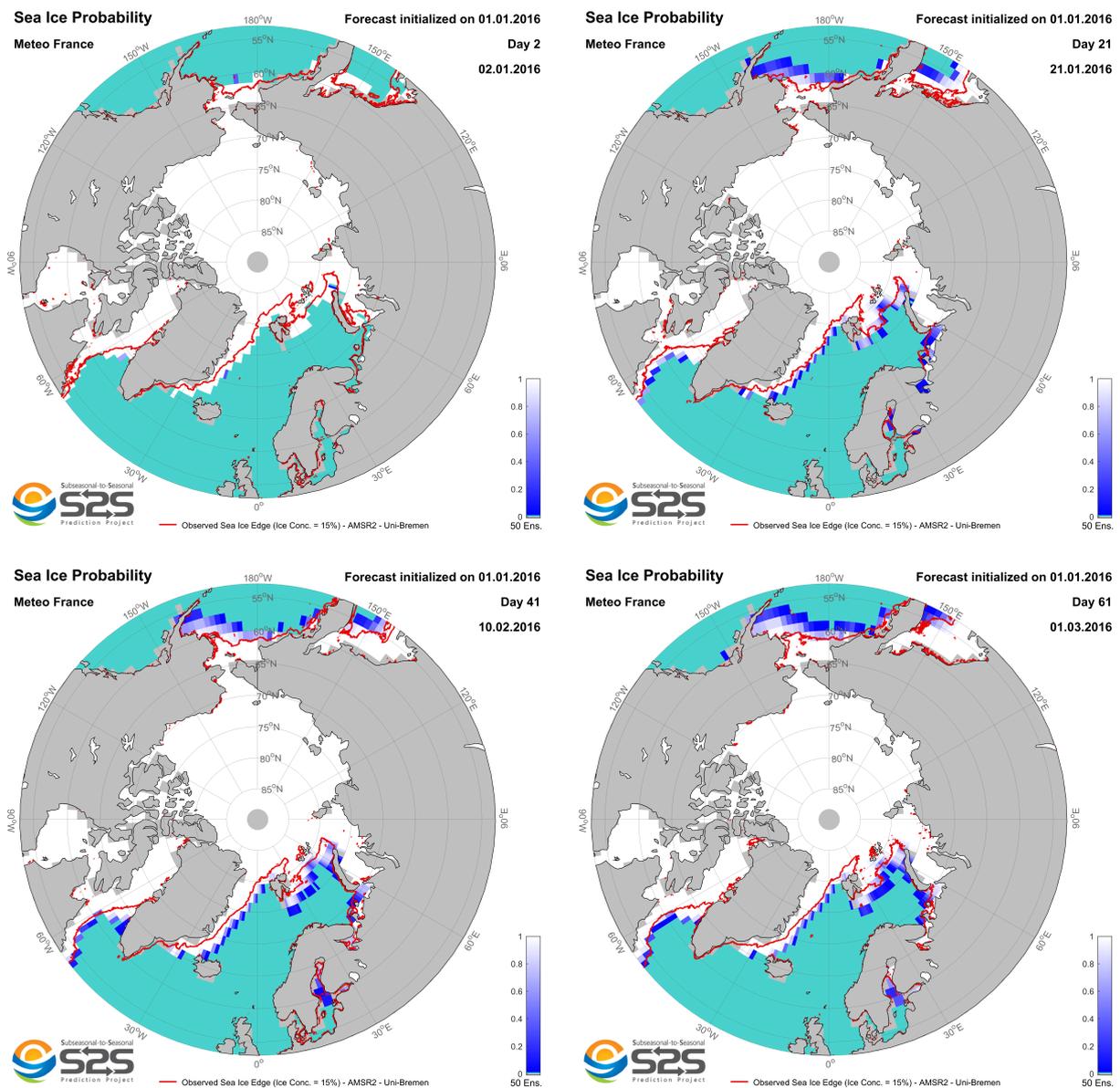


Figure 4.2: Snapshots of the Météo France forecast

$AEE = |O - U|$ (Absolute Extent Error). Omitting the absolute value allows to evaluate if the position of the ice edge is overestimated, in case of positive EE, or underestimated, in case of negative EE. The computation of MHD was too expensive to allow the application of this metric to such an extensive analysis.

Given the high resolution of the satellite data the IIEE and SPS plots include a gray line. This line represents the CMID (Climatological Median Ice Edge) in IIEE plots and CSPS (Climatological Spatial Probability Score) in SPS plots. Those benchmark values put of the predictive skills of each model in context.

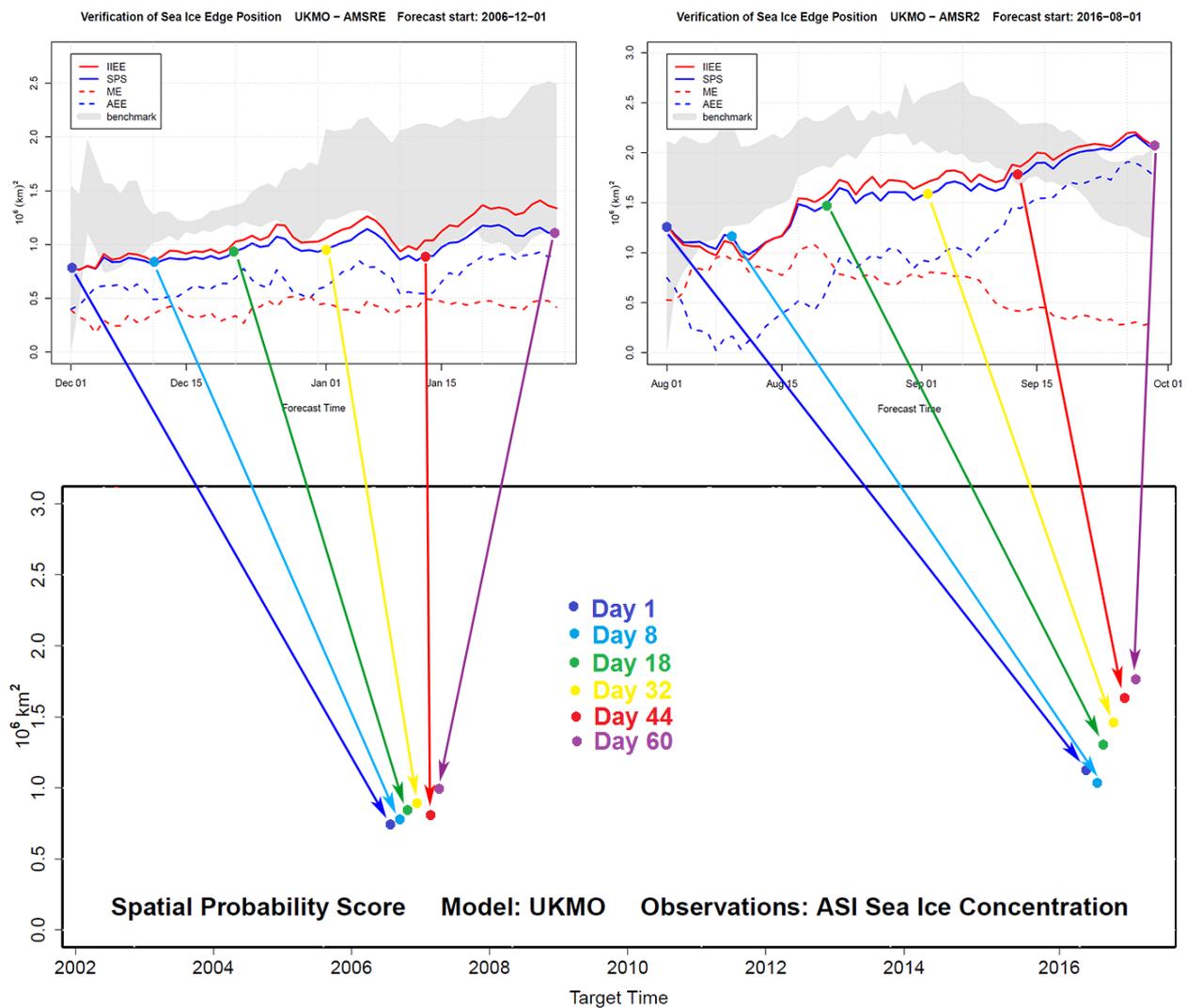


Figure 4.3: Scheme illustrating the simultaneous visualization strategy of multiple forecasts analysis. The model (UKMO) and the variable (SPS) are here chosen to illustrate the mechanism on which the plot is based. The concept remains the same for all the other models and verification metrics.

ECMWF and KMA models exhibit annually recurrent gaps in the time series, between the months of June and November. This is because the two forecast systems started to produce forecasts only since November 2016, thus, they have not been operative for an entire year yet. The on-the-fly reforecasts strategy used in ECMWF and KMA propagates the gaps also to the past years, fact that is reflected in the plots. This implies that the filling of the gaps will not be complete until November 2017. For the moment, we can provide only a partial evaluation of these two forecast systems.

Integrated Ice Edge Error - IIEE

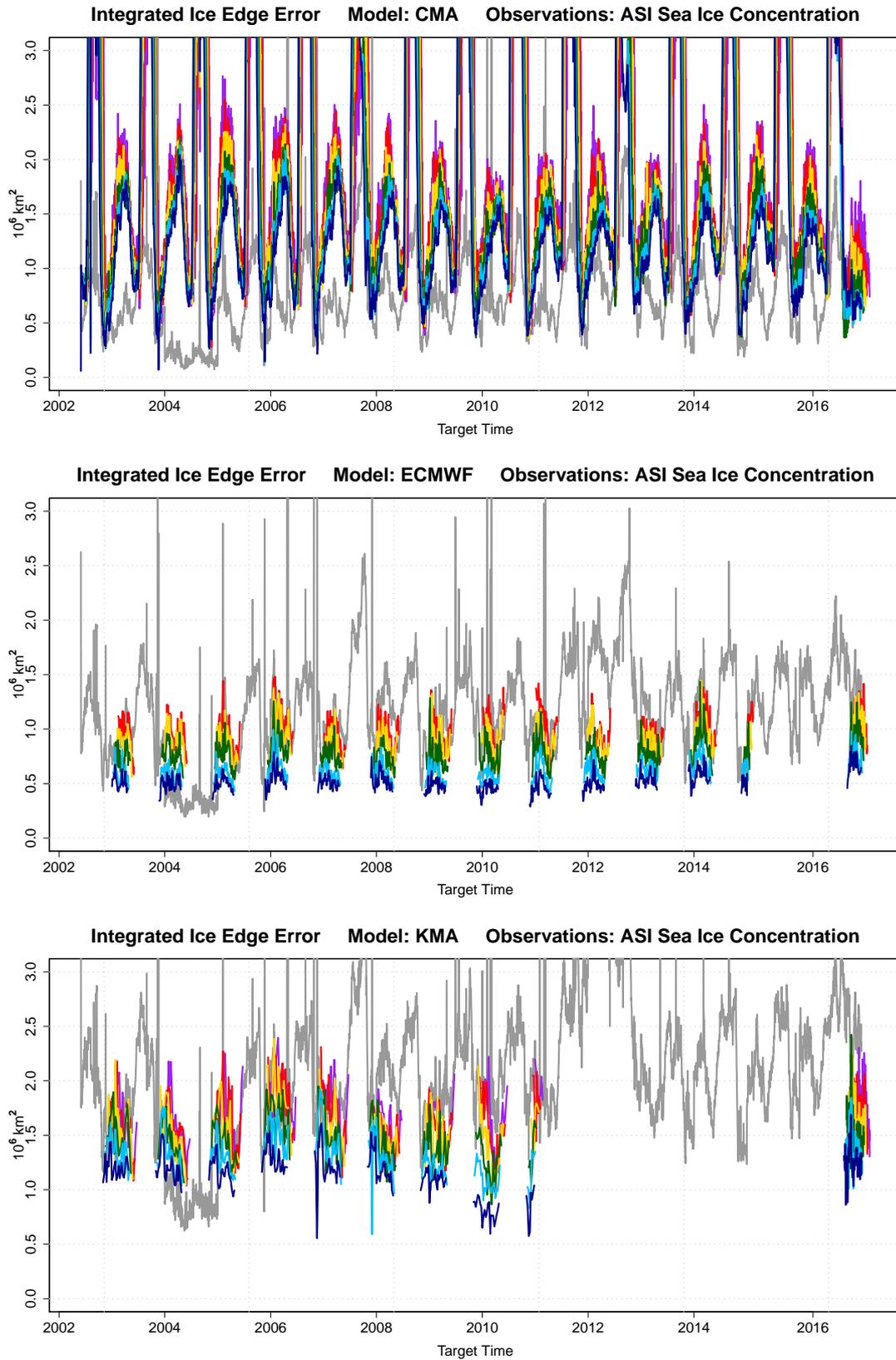


Figure 4.4

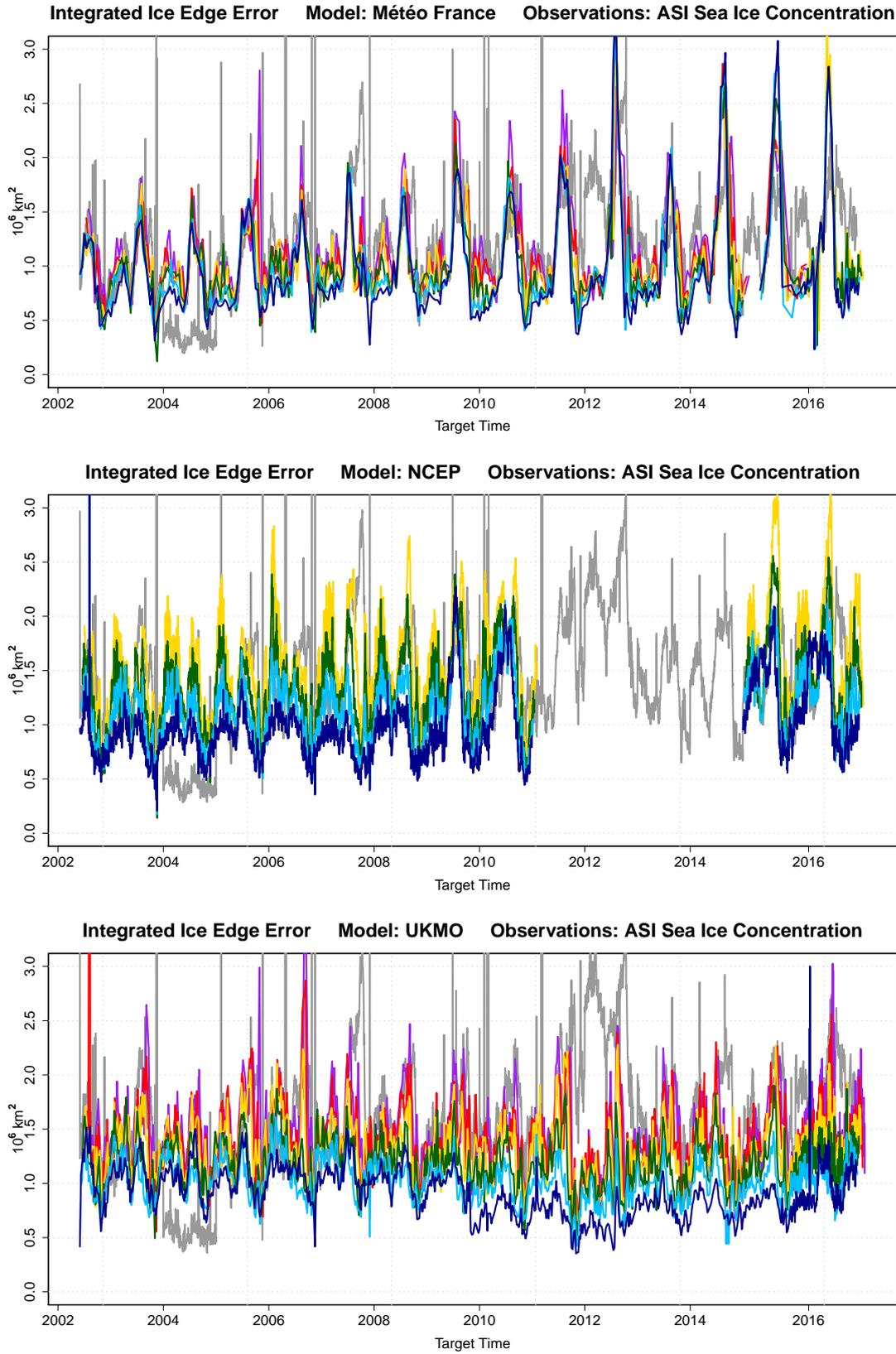


Figure 4.5

Spatial Probability Score - SPS

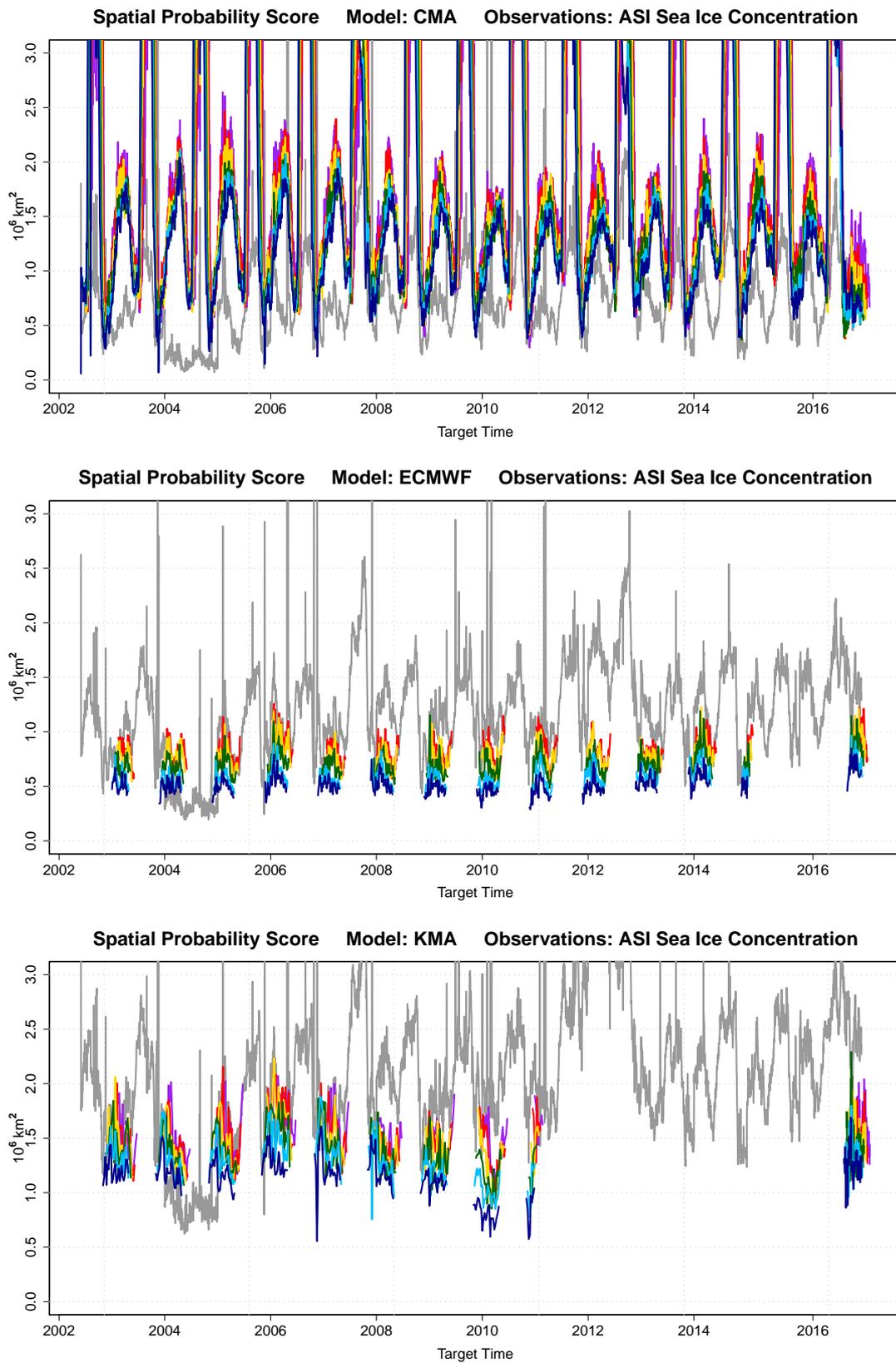


Figure 4.6

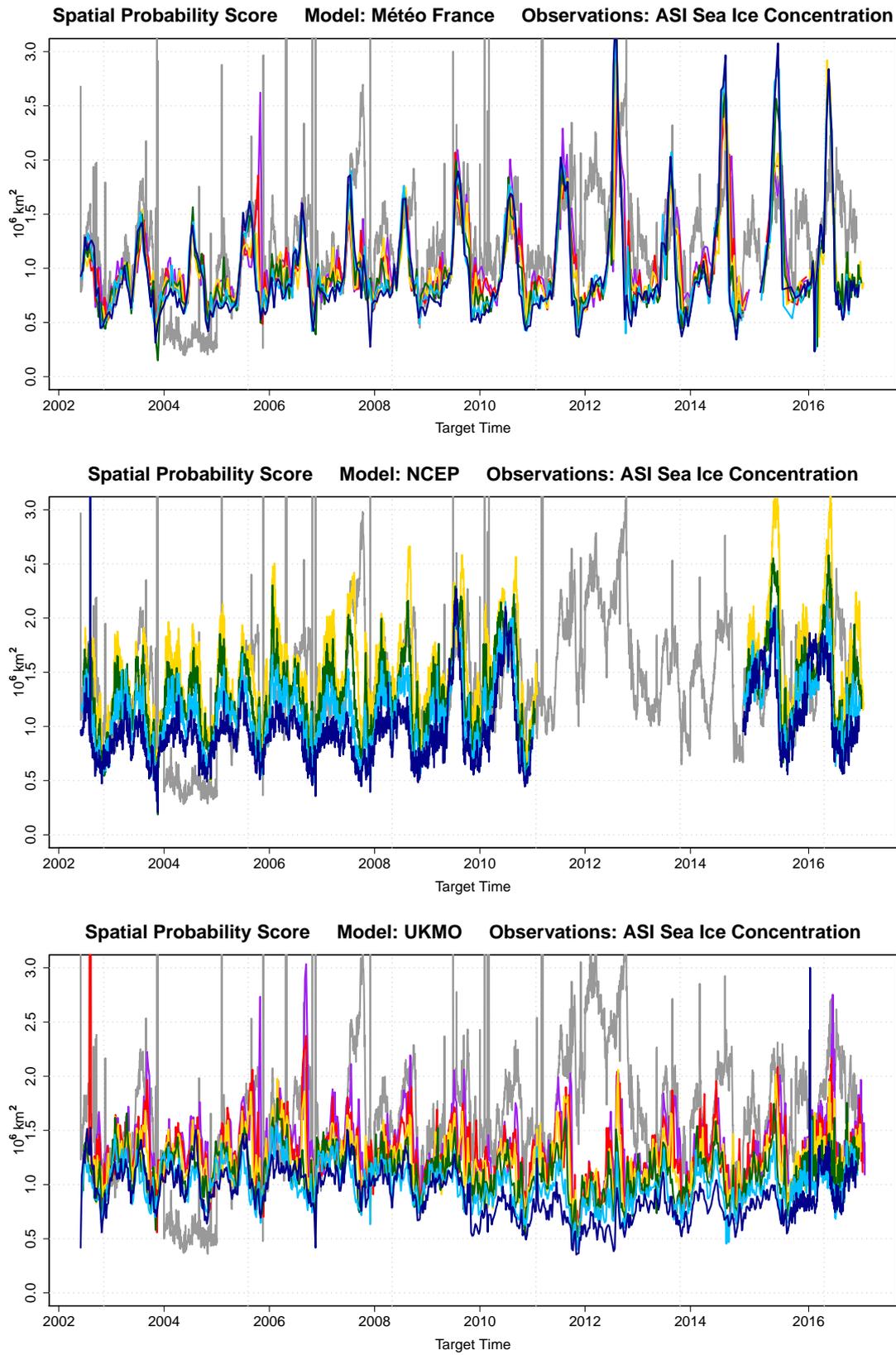


Figure 4.7

Extent Error - EE

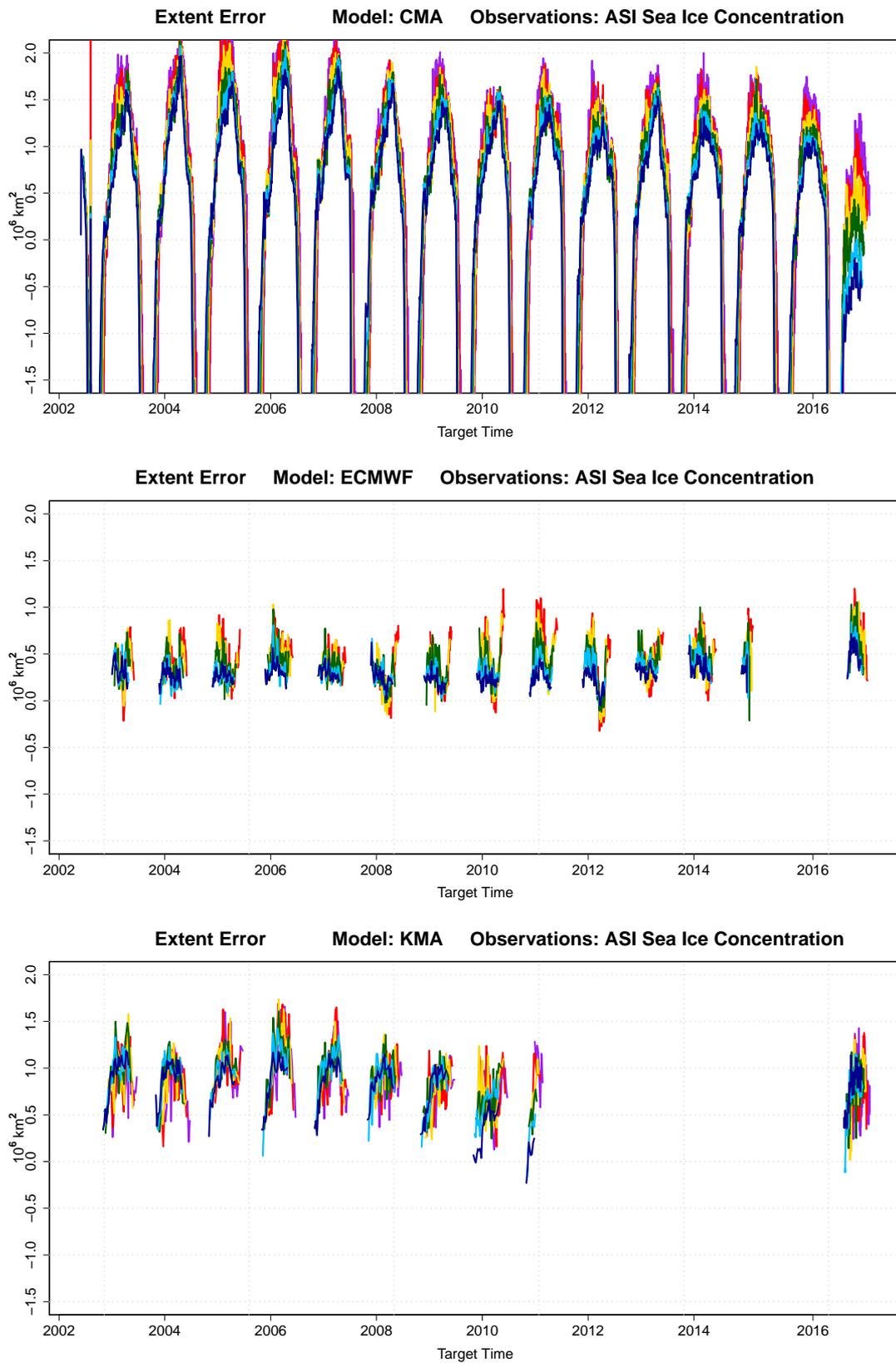


Figure 4.8

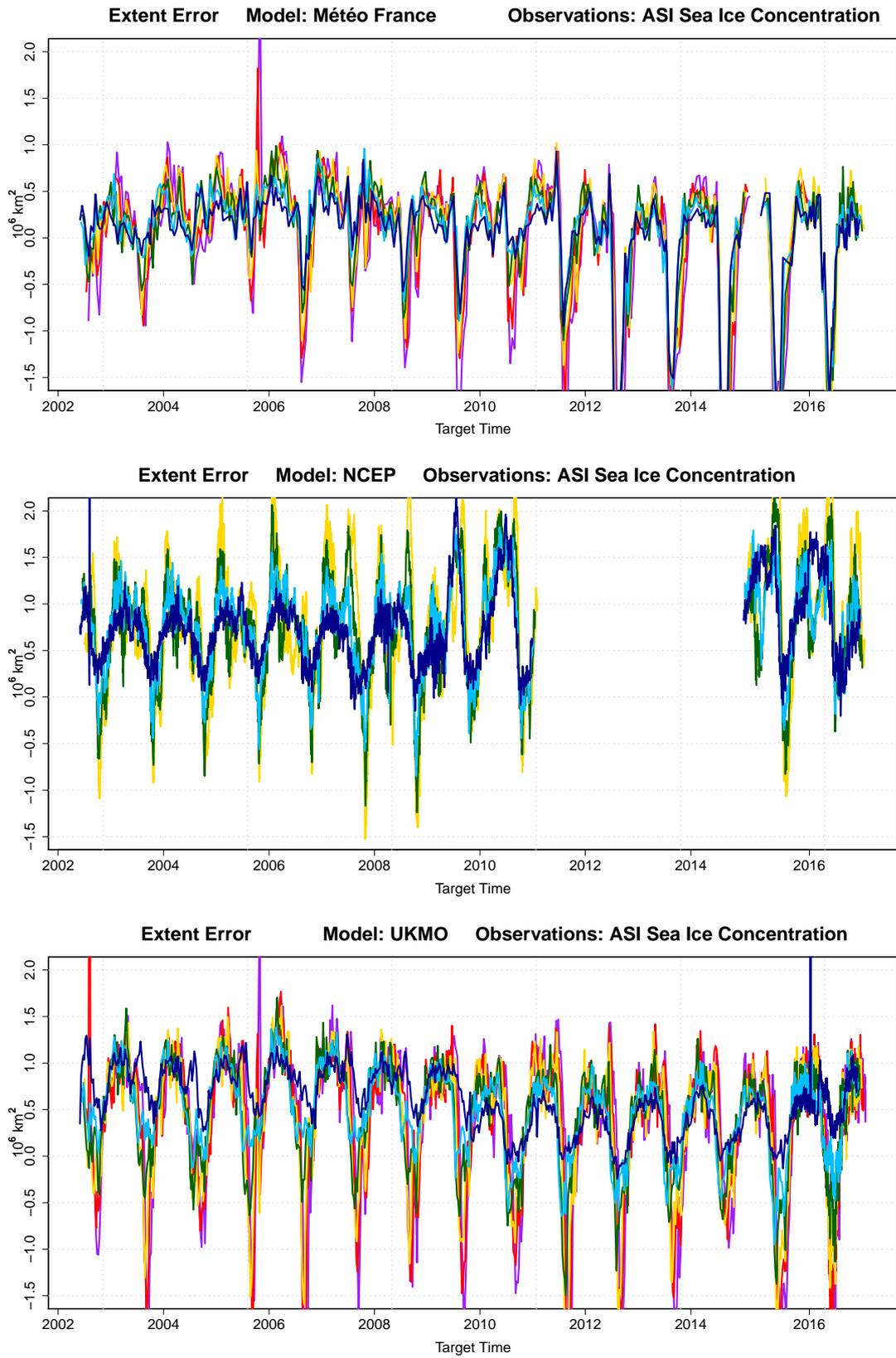


Figure 4.9

Misplacement Error - ME

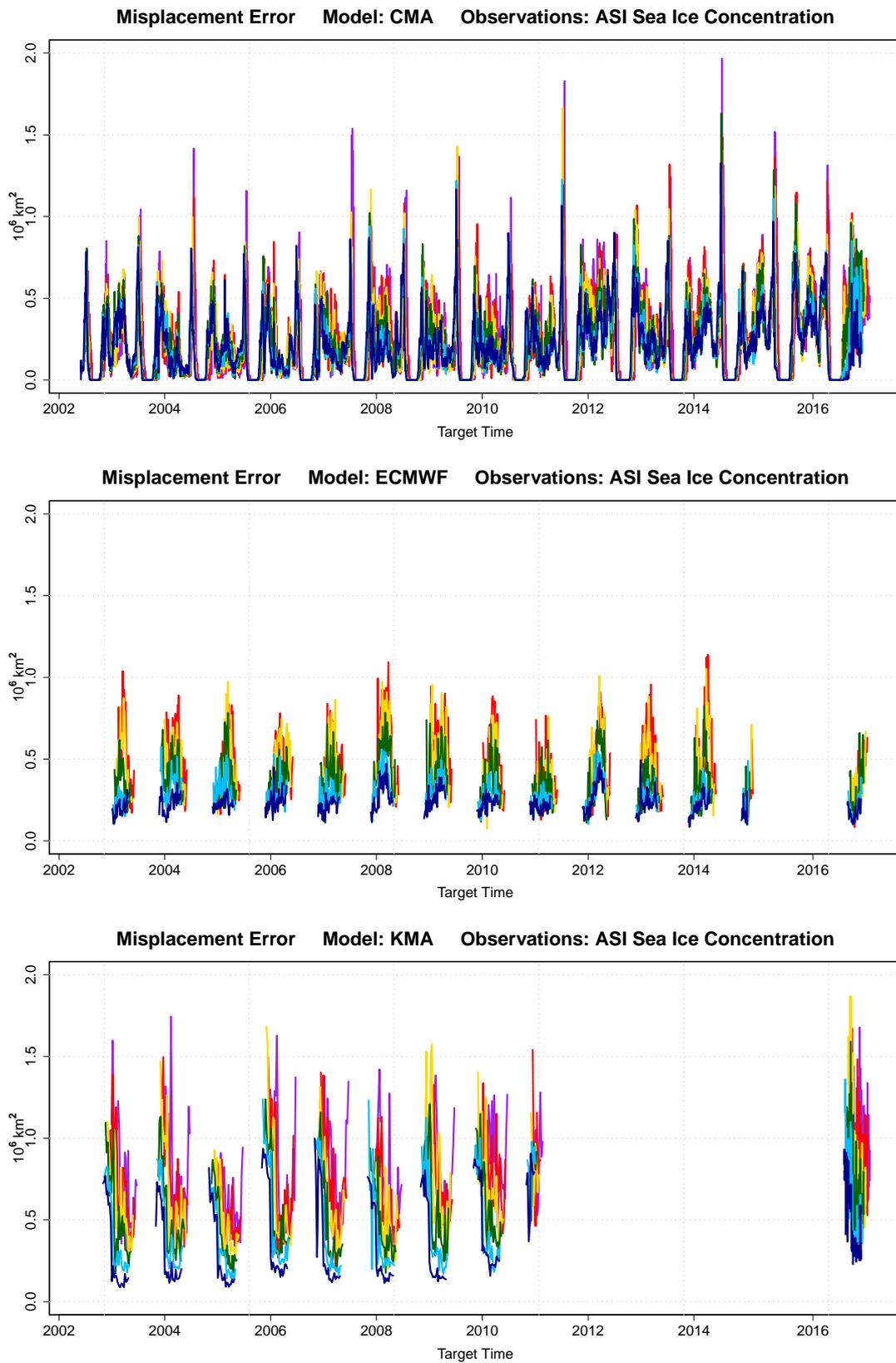


Figure 4.10

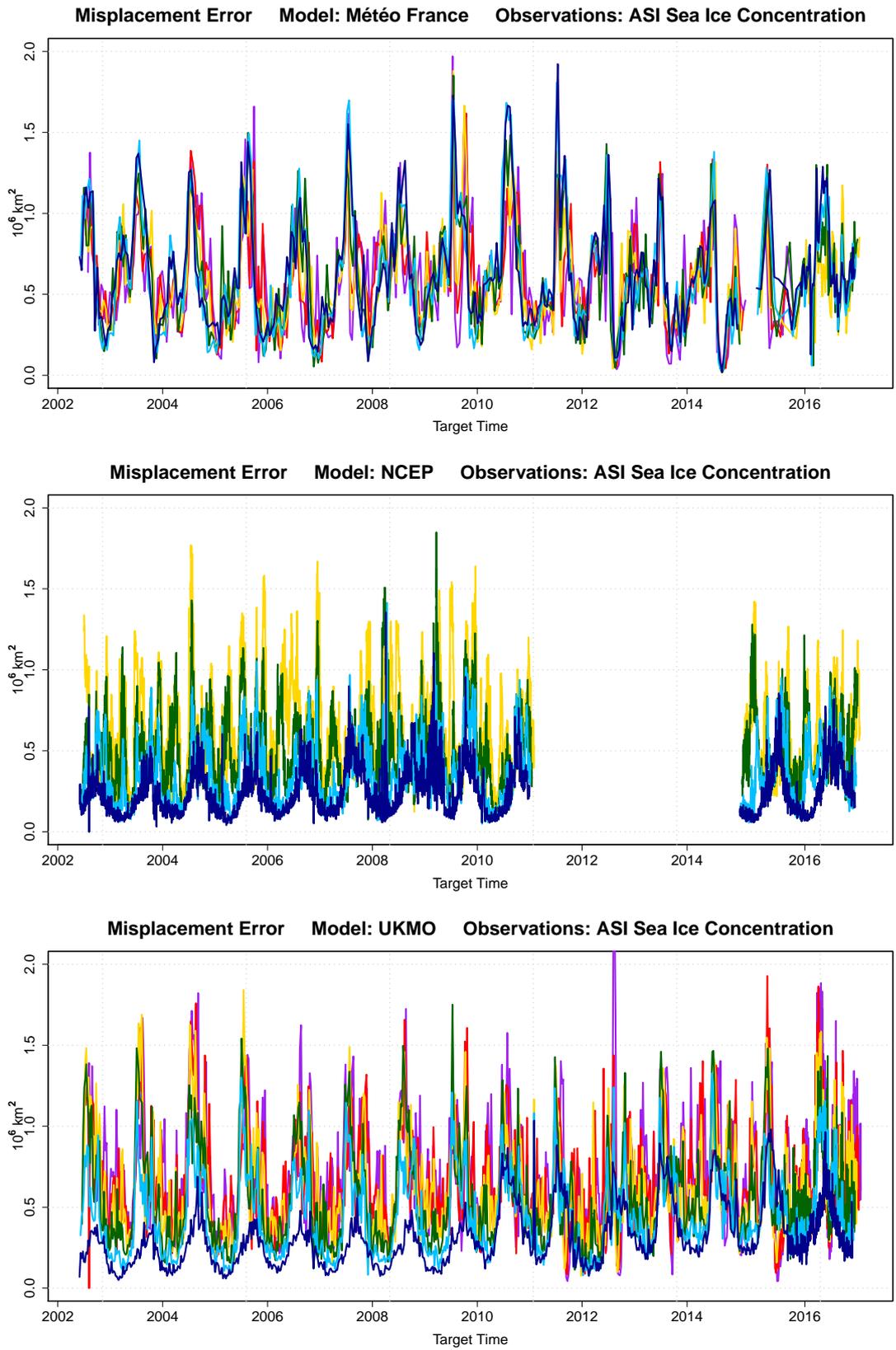


Figure 4.11

The previous plots provide a solid base for the evaluation the S2S forecast models. After emphasizing some features common to all the systems, we move to a specific analysis of the predictive skills of the individual models.

The first feature clearly visible in all the plots is a strong seasonal dependence of the verification metrics. This appears to be more or less pronounced in each model, remaining a cross behavior of all the systems. On one hand, this could be an evidence of biases in the models, which systematically occur with regular annual cycles despite possible differences in the initial sea ice state. On the other hand, a further possible explanation for the observed fluctuations in the metrics, relevant especially for the error peaks occurring during summer, is the higher variability that the system has in this season compared to the winter when the whole Arctic is ice covered. In the next paragraphs, we will try to determine which of the two contributions is the preeminent one to explain specific systematic fluctuations of the metrics.

A substantial non-zero initial error is the second general feature of all the models. As anticipated in the introduction of this chapter, this is mainly caused by the data assimilation process which normally includes some bias correction strategies to improve the overall forecast. Clearly, the forecast models are just a simplified representations of the real world and thus not perfect, especially when dealing with the seasonal timescale. A direct consequence of this, is that initializing the model with values as close as possible to the observations does not necessarily produce a better forecast. Indeed, the opposite is often true. The observed state must be modified and adapted to the model behavior to improve the performances of the forecast system, generating however a big error already at day 1. Moreover, errors can be caused by the employment of a different sea ice concentration product to initialize the model, compared to the one employed in the verification procedure.

A further evidence emerging from the plots is the fact that forecast system performance is seemingly not correlated with the ensemble size. Furthermore, as evidenced in Tab. 3.2 and Tab. 3.3, the ensemble size of most of the models changes between real-time-forecasts and re-forecasts. However, this seems not to have a direct influence on the performance of the forecasts systems, fact that could have two possible explanations. Firstly, the ensemble members could be strongly underdispersive. However, this does not emerge neither from the IIEE calculated between single ensemble members nor from the direct visualization of the probability fields, where a clear spreading of the ensemble members is visible. Secondly, biases in the models and high

spatial resolution differences between forecasts and observations could mask eventual benefits which a larger ensemble could generate.

The last common feature that is worth to mention in this context, is the high similarity of the median-edge IIEE and the SPS. Even if the IIEE is always slightly larger than the SPS, especially in the second half of the forecasts where substantial ensemble-spread has developed (yellow, red and purple curves), these two metrics provide similar information. The IIEE is thus omitted from the analysis in the next section.

4.1.3 | Evaluation of the models' predictive skills

The next paragraphs include a more detailed analysis that focuses on the predictive skills of each single model. The system maintains its seasonality when averaging over time of the verification results, thus we averaged the results of different years. This allows a more robust evaluation of the systems. The gray curve in the SPS panels is the benchmark value calculated for the probabilistic climatological field. In the EE panels a gray curve has been introduced to enhance the 0 km^2 value. This separate the overestimation regime ($EE > 0 \text{ km}^2$) from the underestimation regime ($EE < 0 \text{ km}^2$).

CMA

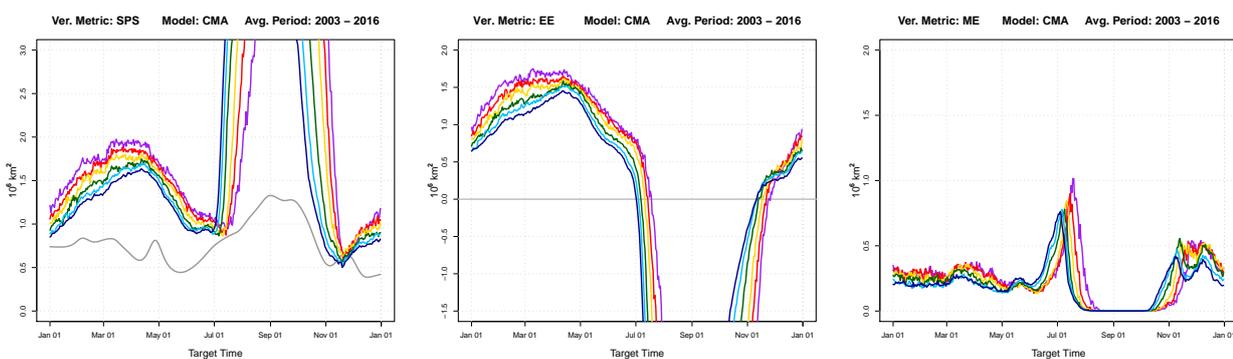


Figure 4.12: CMA model. SPS, EE and ME. Average period from 2003 to 2016.

Fig.4.12 clearly shows that CMA model does not have predictive skill compared to the climatological baseline. The SPS values (colored lines) are always above the CSPS (gray line). Thus, the climatological ice edge position is a better estimation for the sea ice conditions in the Arctic than the CMA forecasts. This is particularly enhanced in spring and late summer, when the model shows strong biases respectively overestimating and underestimating the ice edge posi-

tion. Furthermore, a second critical aspect that emerges from the plots is a large error already at the beginning of the forecast (dark blue curves). This suggests that the data assimilation procedure of sea ice variables is not properly conducted.

We recommend a revision of the sea ice component of CMA model. Errors of this magnitude can potentially strongly affect the atmospheric circulation. Thus, the improvement of the sea ice model could be beneficial also to the predictive skills of the atmospheric forecasts in polar and sub-polar regions.

ECMWF - 2nd Version

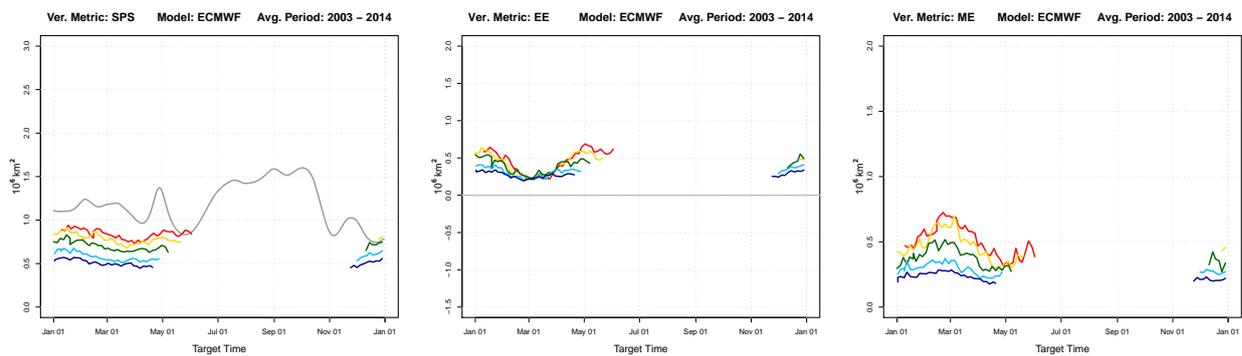


Figure 4.13: ECMWF model - 2nd Version. SPS, EE and ME. Average period from 2003 to 2014.

The verification of ECMWF forecasts, even though limited to the winter and early spring, shows promising results. The model has relevant predictive skills, providing always a better estimation of the ice edge position than the climatology. Interestingly, the error at the beginning of the forecast is lower than the initial error of the other forecasts systems, indicating that the data assimilation strategy adopted at ECMWF is effective. The SPS gradually increases as the forecast proceeds, consistently with chaotic behavior of the system that gradually loses predictive skills. The contributions of EE and ME to the total IIEE are comparable. Each component in fact accounts, on average, for half of the IIEE and similarly for the SPS.

KMA

The KMA model shows predictive skills for the winter and early spring. Its SPS at the end of the forecast (day 60) is always slightly lower than the CSPS, meaning that the system has a moderate potential to provide valuable informations about the edge position even after two months. The contribution to the total error seems to be dominated by the EE at the beginning of the forecast, with the ME contribution becoming more relevant as time passes and accounting in large part for the increase in IIEE and SPS.

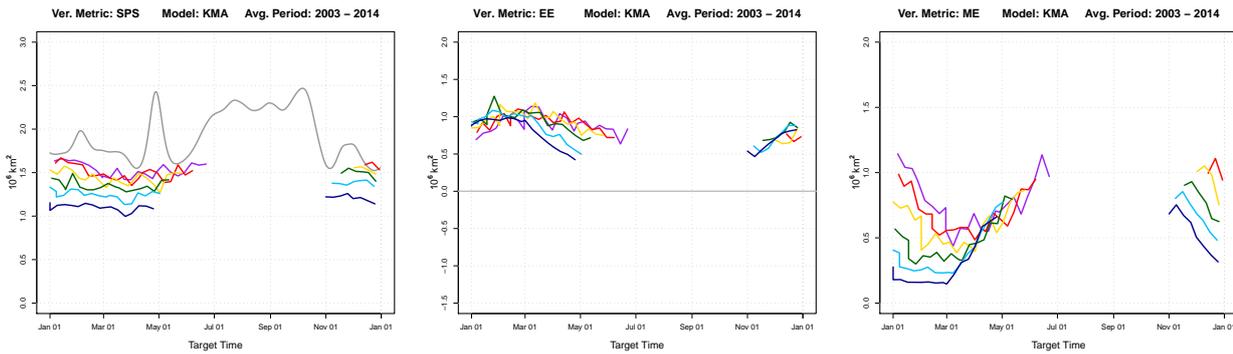


Figure 4.14: KMA model. SPS, EE and ME. Average period from 2003 to 2010.

Météo France

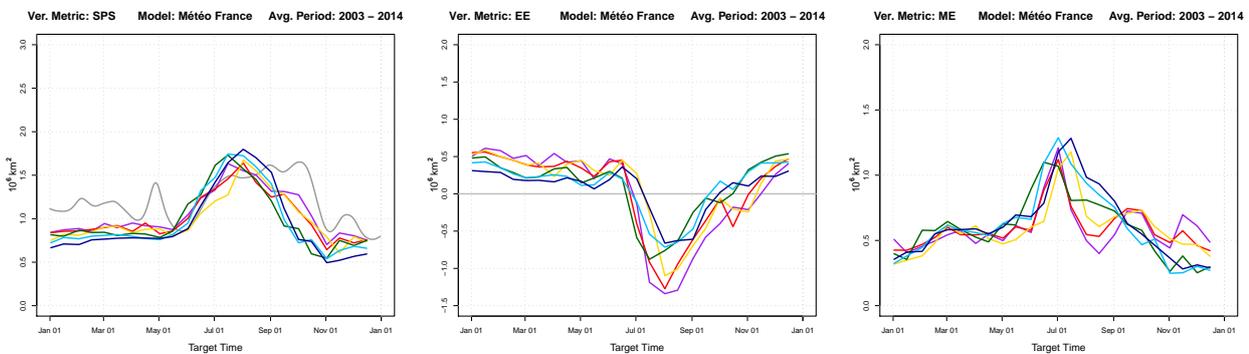


Figure 4.15: Météo France model. SPS, EE and ME. Average period from 2003 to 2014.

The Météo France model exhibits a behavior similar to ECMWF and KMA during late autumn, winter and spring. The SPS at the end of the forecast is slightly lower than the climatological benchmark values, which is an indication of moderate predictive skills. In summer we observe a reduction of the predictive skills, associated with an excessive ice melting in the model. This is already valid at initial time, indicating that this low-ice model bias already affects the initial state during the assimilation procedure. This is clearly visible from the central plot of Fig. 4.15. Such a negative peak in the EE is the direct consequence of the underestimation of the ice edge position in the forecast. The ME is also enhanced during summer, fact that can could associated with a higher variability of the Arctic sea ice edge in summer compared to the other seasons.

NCEP

The NCEP model shows moderate predictive skills only in autumn. During the rest of the year NCEP forecasts are not satisfactory. In particular, a relevant overestimation of the ice edge position is present since the beginning of the forecast. The initial overestimation then is main-

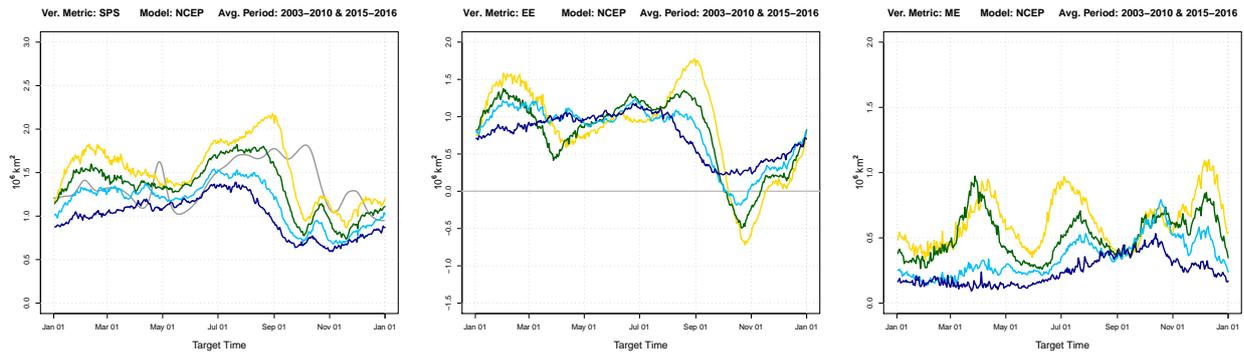


Figure 4.16: NCEP model. SPS, EE and ME. Average period from 2003 to 2011 and from 2015 to 2016.

tained and occasionally even enhanced in the continuation of the forecast. This means that an improvement of the data assimilation strategy of this model would likely lead to better forecasts for the ice edge position.

UKMO

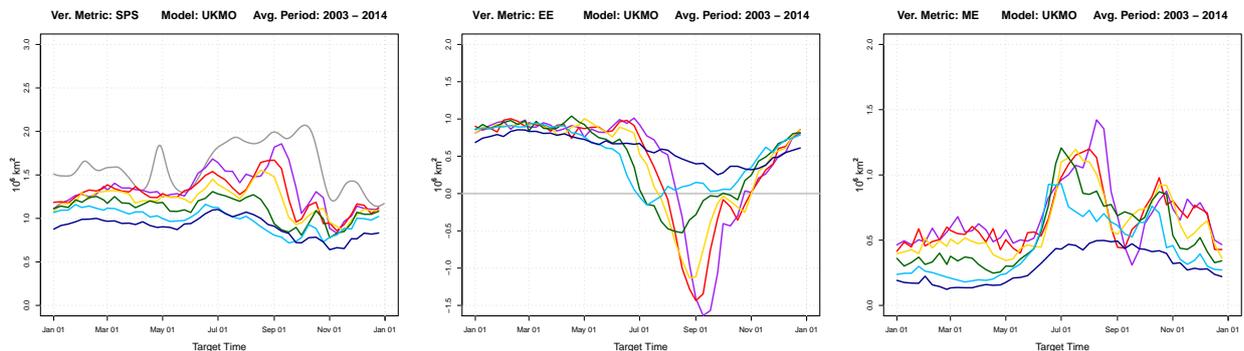


Figure 4.17: UKMO model. SPS, EE and ME. Average period from 2003 to 2014).

The SPS values of the UKMO model are always below the climatological benchmark values. This means that the model exhibit predictive skills during the whole year, which appears to be a good achievement. The summer season appears once more challenging also for UKMO forecasts. In particular, those forecasts initialized between June and July appears to melt too much sea ice, leading to an underestimation of the ice edge position at the beginning of September. However, in contrast to the Météo France forecasts, where an underestimation prevails already at initial time, in the UKMO forecasts an initial overestimation turns into an underestimation with growing lead time.

4.2 | Comparison between predictive and prescriptive ECMWF model

ECMWF updated the sea ice component of its seasonal forecast system at the end of November 2016. In particular, the forecast center abandoned the prescription of the sea ice variables, introducing a dynamical sea ice model coupled with its atmospheric and ocean models. Real time forecasts and reforecasts are available for both the model versions, allowing a comparison which could reveal possible strengths and weaknesses of the two approaches.

Obviously, our considerations are limited to the winter and spring seasons, when the forecasts are available for both the model versions.

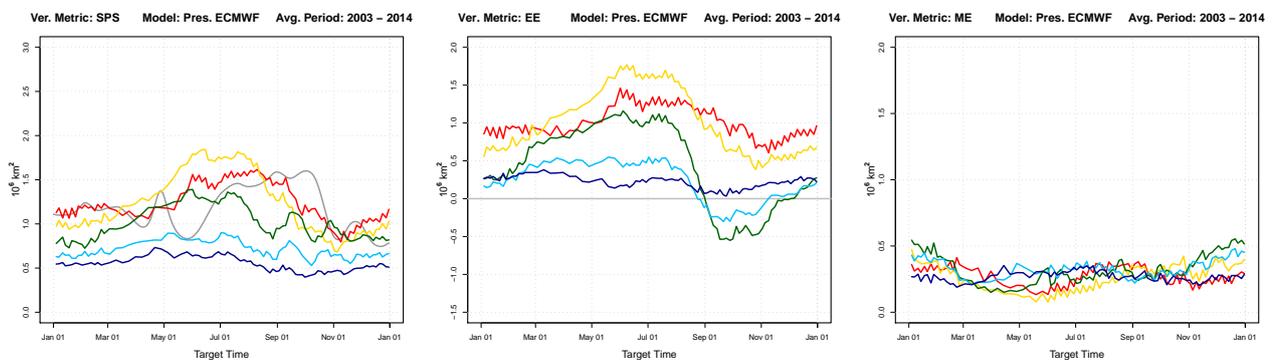


Figure 4.18: ECMWF model - Prescriptive version. SPS, EE and ME. Average period from 2003 to 2014.

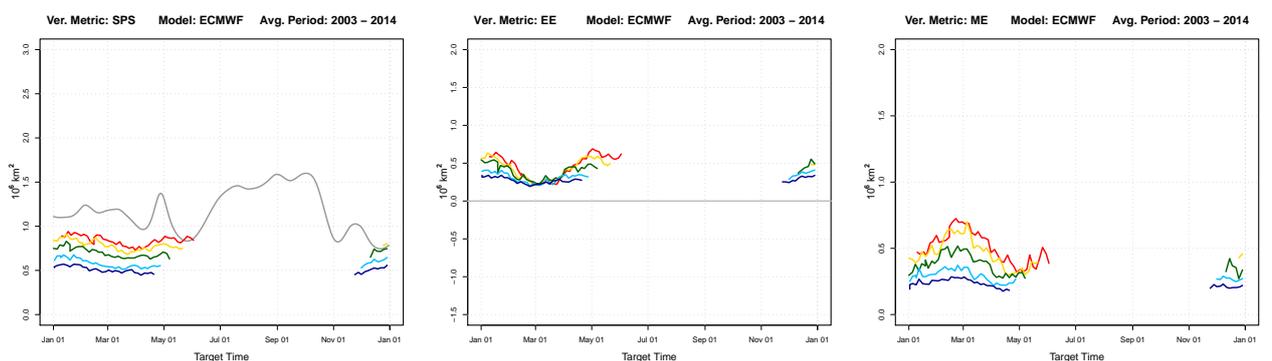


Figure 4.19: ECMWF model - Predictive version. SPS, EE and ME. Average period from 2003 to 2014.

When comparing the SPS of the prescriptive version, in Fig. 4.18, with the SPS of the predictive version, in Fig. 4.19, we observe that the second model version has better predictive skills

compared to the first one. The two versions exhibit similar results at the beginning of the forecast (blue lines), with an initial SPS of $\sim 0.6 \times 10^6 \text{ km}^2$. As the forecasts proceed, the prescribed system overestimates the ice edge position strongly, while the predictive systems seem to maintain a better agreement with the satellite data. This can be clearly noted by comparing the two EE plots, which shows a better performance of the more recent version of ECMWF model. The ME, on the contrary, shows comparable results.

The better performance achieved by the predictive version of ECMWF meets with our expectations. However, the positive results of the predictive system should not only be considered as the direct consequence of the employment of a more sophisticated method. Two other possible causes are likely behind this improvement. Firstly, the climatology to which the prescribed sea ice is relaxed, is built averaging the sea ice conditions over the past decade. The current climate change scenario strongly affects the Arctic sea ice and causes extraordinary low sea ice extent conditions, particularly in late summer but in a minor extent also during the rest of year. A relaxation towards the climatology of our forecasts is therefore not able to fully capture the negative trend of sea ice that we observed in the last years, which could be partially responsible for the overestimation of the ice edge position in the prescribed sea ice. To reduce as much as possible this error contribution, only the last few years are used at ECMWF to build the relaxation climatology. Secondly, the probabilistic formulation of the forecasts with diverging ensemble members, which is a feature proper of the predictive system but not of the prescriptive one, could have a beneficial effect on the performances of the newest ECMWF model version.

4.3 | Verification against models own analysis

This section presents the verification results of the S2S models when verified against the models own analyses. Before presenting and commenting the results, it is worth to explain in detail how the model own analysis have been constructed.

As already mentioned in the introduction of the current chapter, the idea behind the models own analysis is to define virtual observations based on the control forecasts evaluated at the initial time of each single forecast. This should provide an estimation of the true sea ice state, which more or less realistic depending on the biases introduced by the data assimilation procedure. It is relevant to remind the reader that most of the S2S forecast systems do not have a daily initialization frequency. Thus, the models own analysis can not be defined for each day of the time domain. This issue is solved by extrapolating the sea ice concentration state of the missing days by applying a linear interpolation between the closest known day in the past and the closest known day in the future. The Météo France model is not considered for this verification analysis. Its too low initialization frequency, which is 15 days for the reforecasts and 30 days for the real time forecasts, would not allow a reliable estimation of its own analysis for the numerous missing days.

The chosen interpolation method is not very sophisticated but it represent the only reasonable option considering the low resolution of the S2S database. More elaborated strategies, such as a contour shifting algorithm, could be employed to provide a better estimation of the ice edge position in the missing days, for data with higher resolution. The R package IceCast¹ provides some interesting contour shifting tools that would be useful for this purpose.

The SPS, EE and ME are reported for each model for the period 06.2002 - 06.2017.

¹<https://cran.r-project.org/web/packages/IceCast/vignettes/biasCorrectDemo.html>

Hannah M. Director, Adrian E. Raftery, & Cecilia M. Bitz

CMA

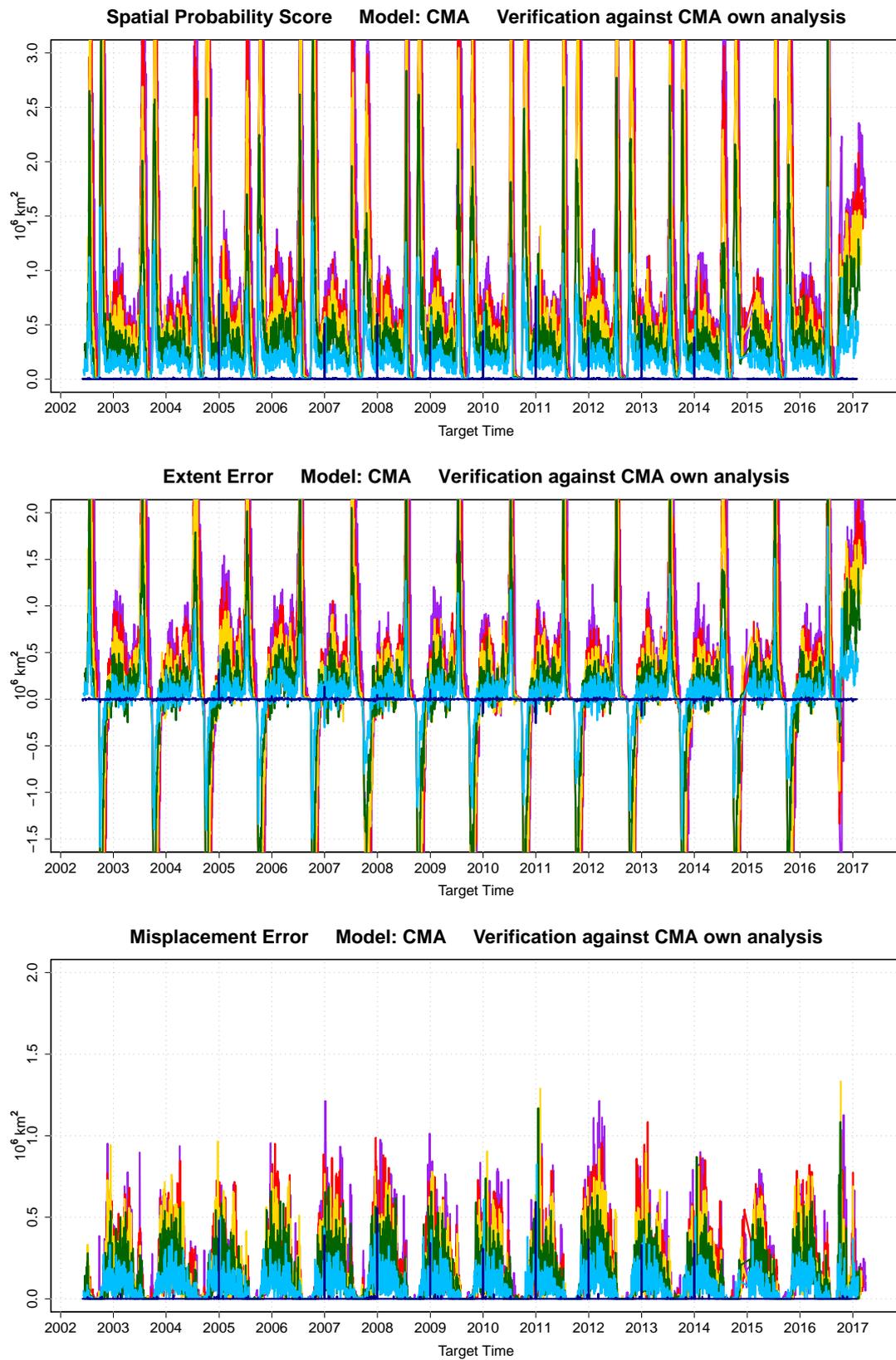


Figure 4.20

ECMWF

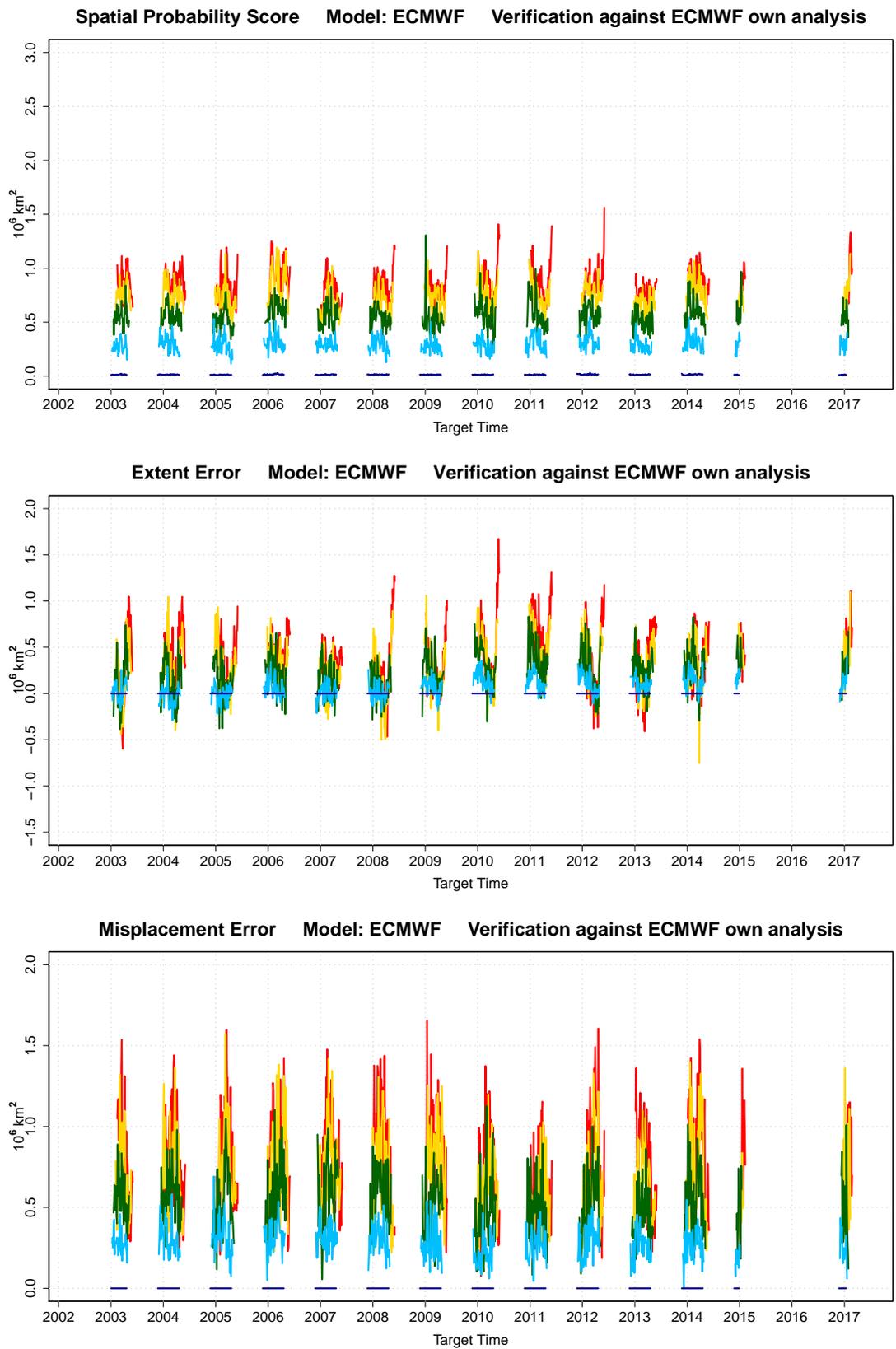


Figure 4.21

KMA

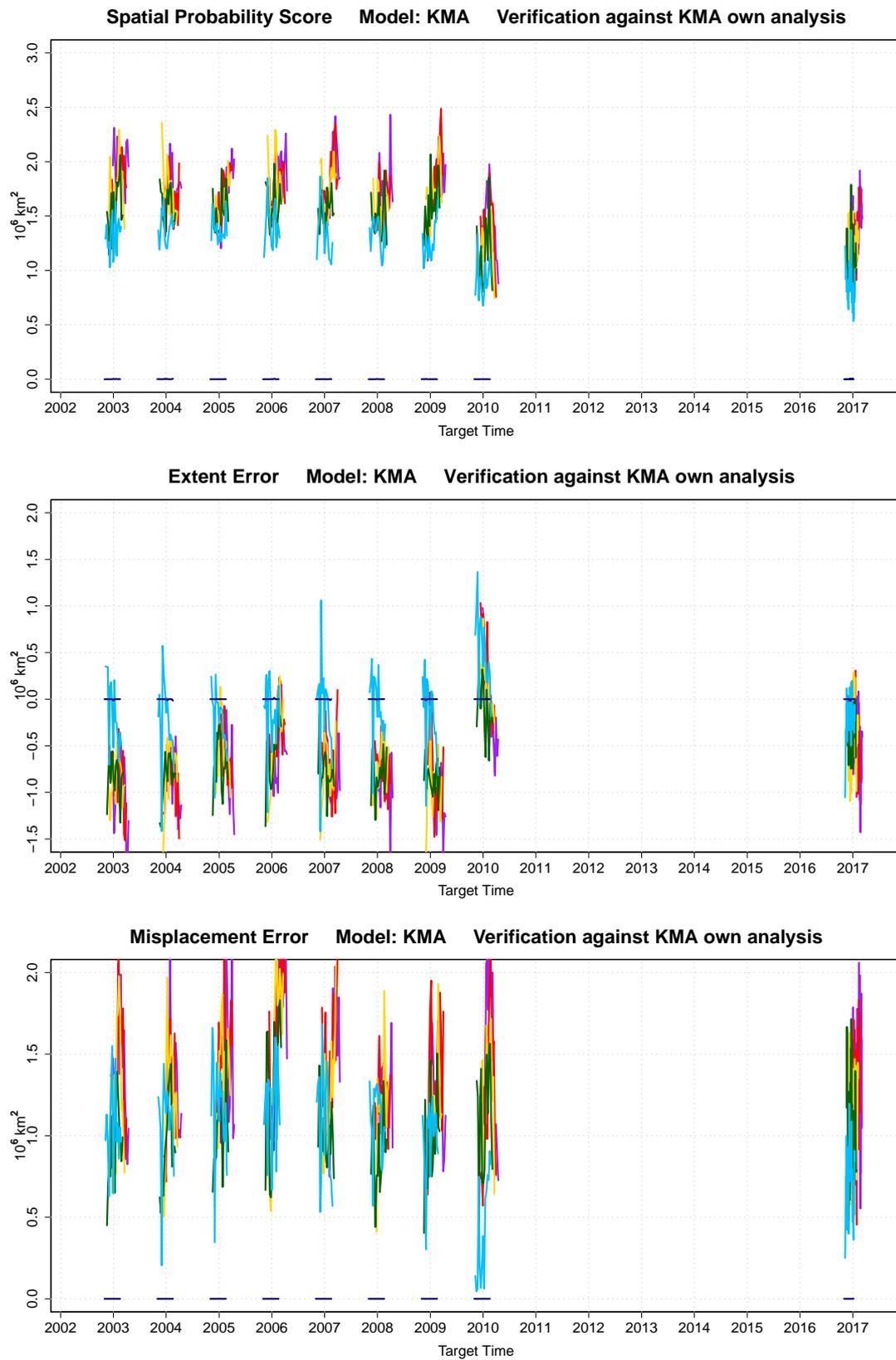


Figure 4.22

NCEP

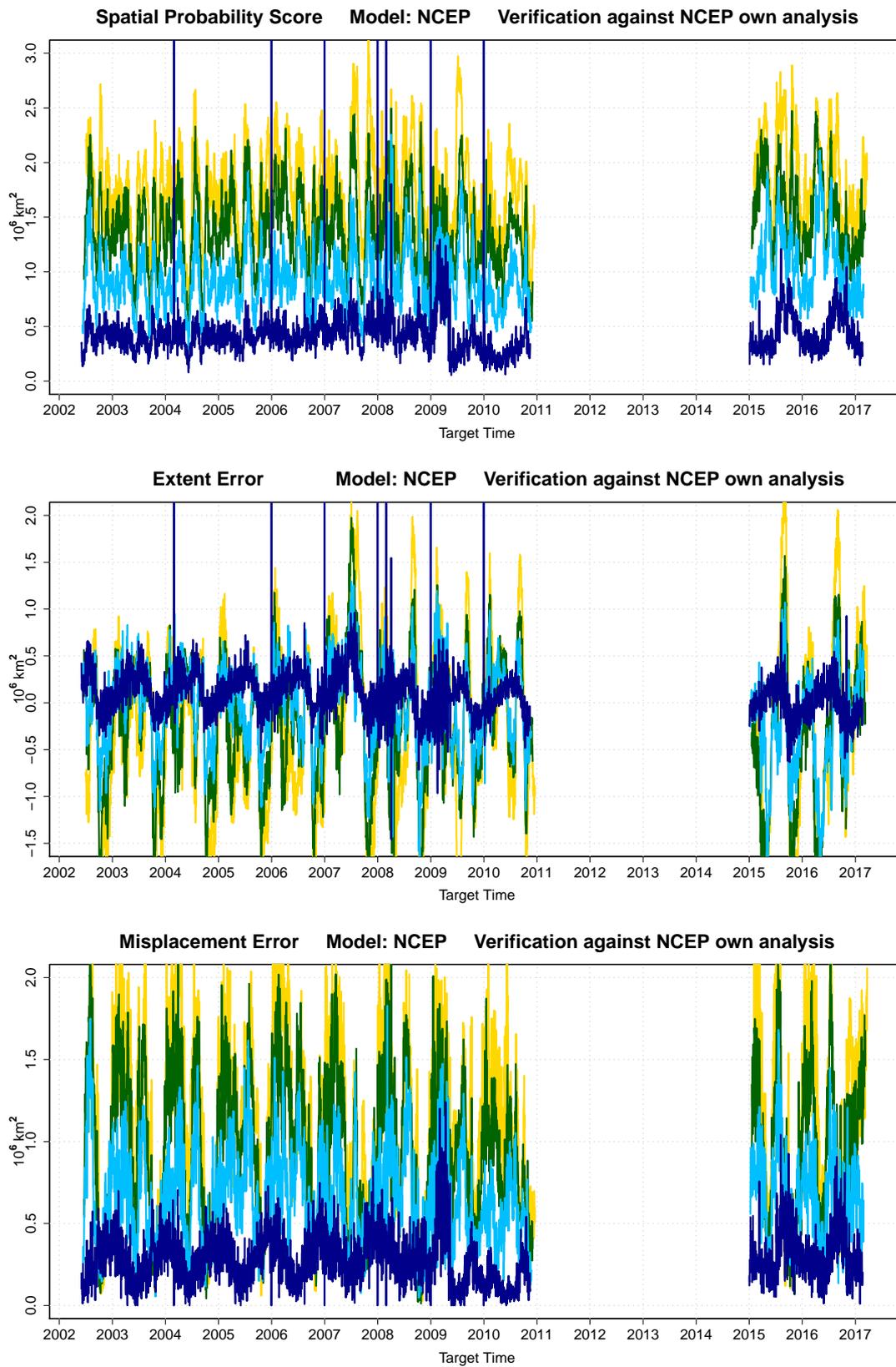


Figure 4.23

UKMO

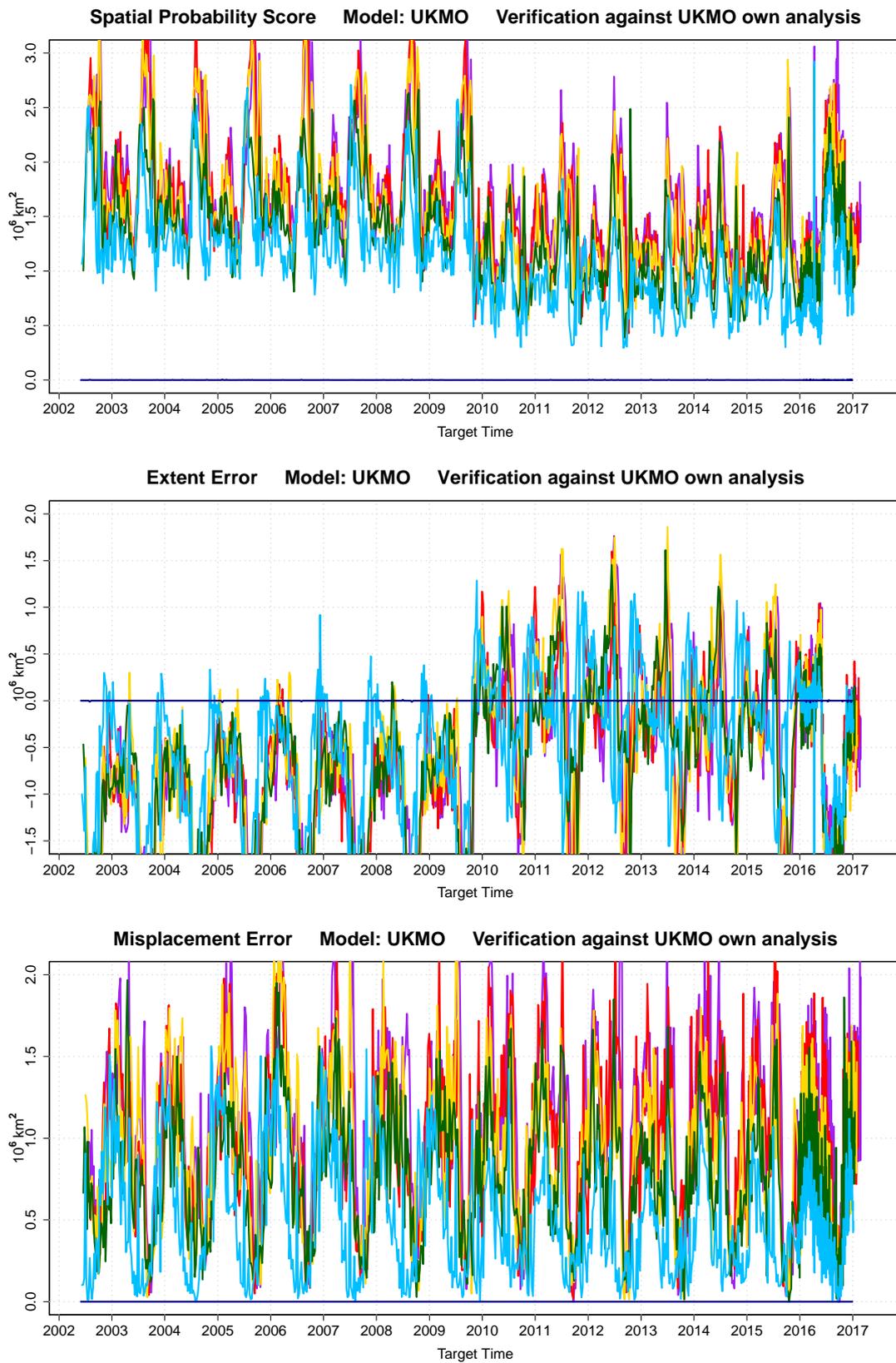


Figure 4.24

As expected, at the beginning of the forecasts errors are negligible for all the models except NCEP. The peculiar behavior of NCEP model is displayed in Fig. 4.23. This exception is explained by the fact that NCEP system saves the forecast output starting from day 2 instead of day 1. At this time the ensemble presents already a certain degree of spreading. Since the NCEP own analysis is based only on the control run and not on the entire perturbed ensemble, the own analysis and the initial state differ, fact that is correctly captured by the verification metrics. The divergence of NCEP ensemble since the beginning of the forecast can be noted in Fig. 4.25, where the IIEE quantifies the spreading between single ensemble members pairs. Furthermore, for most of the models a considerable fraction of the errors seen after the full forecast range is already present after only 8 days. In this context ECMWF is, once more, a promising exception. Indeed, its SPS is gradually growing during the whole lead time of the forecasts.

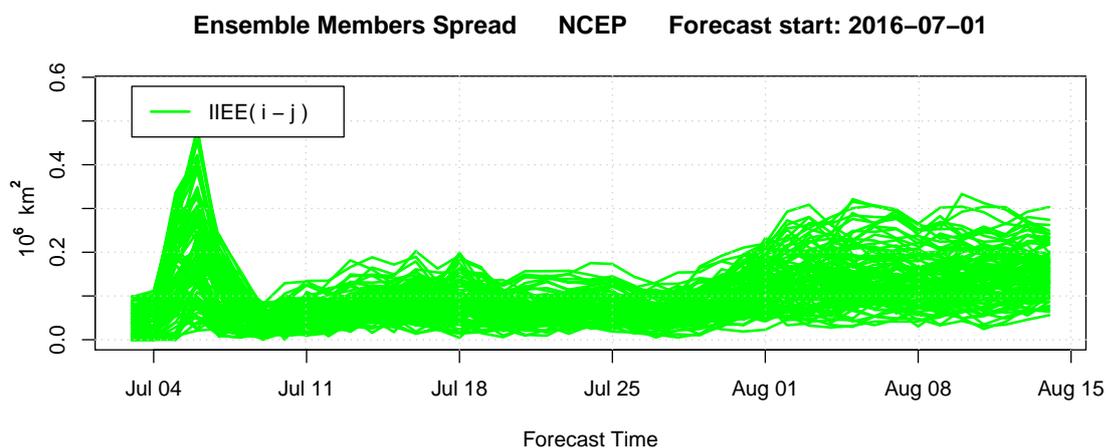


Figure 4.25: IIEE calculated between all the possible ensemble members pairs combinations of a NCEP forecast. A divergence of the ensemble is visible already at the beginning of the forecast.

Generally, the verification against the models own analysis shows worst results compared to the verification against the satellite observations. The only exception seems to be the ECMWF model, which shows comparable results. The overall increase of the SPS, EE and ME in the other models can be explained by the fact that the data assimilation process strongly affects the initial state of the control forecast, causing the models own analysis to be far from the true sea ice condition in the Arctic and introducing a further error source. This would also explain why ECMWF is not significantly affected by this worsening, since it is the model with the better agreement between the initial state of the forecast and observed state. However, we would expect to observe lower errors when verification is performed against models own analyses compared to

observations. Since the same model is used for the assimilation, systematic errors that are shared between the forecasted states and the initial states should disappear. Surprisingly, this is not observed.

The interesting aspect of this verification approach is not the assessment of predictive skills of the models, for which the verification against the satellite observations is certainly more appropriate, but rather understanding the response of our forecast systems to specific changes in the forecast procedure. These changes are likely highlighted by this verification approach.

An interesting example, in this context, is the sudden improvement of the UKMO model from 2010 onwards. This is clearly visible in the SPS plot of Fig. 4.24. This behavior is a quite peculiar and can not be explained with an improvement of the model physics, since the on-the-fly nature of its reforecasts guarantees the same model configuration to be used for all the years. A possible explanation is a change in the data used to initialize the model. The UKMO ocean and sea ice models are initialized using the NEMOVAR system [Mogensen et al. (2009)]. Its technical aspects are described in detail in the dedicated ECMWF technical memorandum [Mogensen et al. (2012)], which indeed reports a change of the initialization database for SST and sea ice variable, exactly at the beginning of 2010. From December 1981 until December 2009, SST and sea ice are taken from the NCEP OI_v2 weekly product [Reynolds et al. (2002)], while from January 2010 onwards the OSTIA SST and sea ice products [Donlon et al. (2011)] are used. As can be seen from the EE plot in Fig. 4.24, the improvement in the SPS is the consequence of a reduction of the forecast underestimation. Likely, the new dataset used to initialize the model provides a better representation of the true state, resulting in a better agreement with the forecasts. A similar behavior is observed for KMA model, with a sudden shift downwards of the SPS for the years 2010 and 2017. This is consistent since UKMO and KMA are sharing the same GloSea5 seasonal forecast system. However, the partial forecast coverage of KMA system do not allow a robust evaluation of this finding for this specific model.

Comparison with verification based on Modified Hausdorff Distance

The MHD can be calculated in the frame of the verification against the models own analysis. The identical low resolution of both forecasts and models own analysis allows an extensive calculation of this metric, avoiding technical issues related to the elevated computational time typical of the satellite observations. In addition, the forecast and the model own analysis are defined on the same grid, excluding the occurrence of biases induced by a different spatial step of the two contours, as discussed in Sec. 2.5. To avoid errors caused by the inclusion of coastlines (Fig. 2.5c), those are not considered in the MHD computation. Thus, the MHD is simply calculated between the forecasted and observed ice edges, without any further attempt of bias correction.

As explained in detail in Chap. 2, the two metrics that provide an estimation of the mean distance between the edges are the NIIEE (or similarly NSPS) and the MHD. The computation procedures behind those metrics are different. On one hand, the focus of NIIEE and NSPS is initially on areas, specifically on the of disagreement area between forecasted and observed edge, which is afterwards normalized to a length that can be interpreted as the mean distance between the edges. The normalization is achieved by dividing the IIEE or SPS by the climatological length of the ice edge, which is calculated from the satellite data. On the other hand, the MHD performs a direct topological computation of the mean distance between the edges. Despite the different nature of the metrics, these should ideally exhibit similar results if we assume that they are both equally effective. These considerations enhance the need of a detailed comparison between NIIEE and MHD, to determine which metric is more reliable and stable.

This section tries to address the previous questions by setting a basic statistical analysis,

which should reveal possible differences and common aspects of the NIIEE and MHD. Note that in this context the NIIEE is considered instead of the NSPS, for consistency with the MHD. Indeed, the formulation of both the MHD and the NIIEE is based on deterministic sea ice edges, in our case the ensemble-median sea ice edge, while the NSPS deals directly with the sea ice probability. At first, we will restrict the analysis to the UKMO model because it covers the whole time period without gaps and appears to be a reliable forecast system. The same analysis conducted on the CMA or NCEP models would not be representative of a realistic configuration, mainly due to the apparent biases in the model physics and in the data assimilation process.

As starting point, the UKMO's MHD and NIIEE verification results are displayed in Fig. 5.1. The elevated noise, especially in the MHD, does not allow a clear visualization of possible annual cycles. Thus, the noise has been filtered out by applying a spline smoothing algorithm. The smoothed curves are reported in Fig. 5.2.

Two aspects emerge from a first visual comparison of the plots. Firstly, the mean distance between the edges is different for MHD and NIIEE. Specifically, MHD estimates a distance which is approximately twice what estimated by the NIIEE. Secondly, the trends of the two metrics are correlated, suggesting that the same features are captured. Furthermore, similarly to the NIIEE, also the MHD capture the 2010 UKMO skills improvement which has been described in Sec. 4.3. However, the annual cycles in the MHD appear to be less precise than those in the NIIEE, probably because of the augmented noise level brought about by the MHD sensitivity to outliers. To quantify the agreement between the two metrics, we plot the MHD and NIIEE in a scatter plot, shown in Fig. 5.3. This suggests a positive correlation between the two metrics. To confirm our visual interpretation, a correlation test is performed. The result is reported in Tab. 5.1.

Table 5.1: Pearson correlation test for MHD and NIIEE.

Forecast Time	Smoothed Data			Original Data		
	r	95% Conf. Interv.	p-value	r	95% Conf. Interv.	p-value
Day 1	0.915	[0.905; 0.924]	$< 2.2 \times 10^{-16}$	0.816	[0.795; 0.836]	$< 2.2 \times 10^{-16}$
Day 8	0.813	[0.792; 0.833]	$< 2.2 \times 10^{-16}$	0.682	[0.649; 0.714]	$< 2.2 \times 10^{-16}$
Day 18	0.872	[0.857; 0.886]	$< 2.2 \times 10^{-16}$	0.756	[0.728; 0.781]	$< 2.2 \times 10^{-16}$
Day 32	0.860	[0.844; 0.875]	$< 2.2 \times 10^{-16}$	0.678	[0.644; 0.710]	$< 2.2 \times 10^{-16}$
Day 44	0.770	[0.744; 0.793]	$< 2.2 \times 10^{-16}$	0.465	[0.416; 0.511]	$< 2.2 \times 10^{-16}$
Day 60	0.672	[0.638; 0.704]	$< 2.2 \times 10^{-16}$	0.347	[0.293; 0.400]	$< 2.2 \times 10^{-16}$

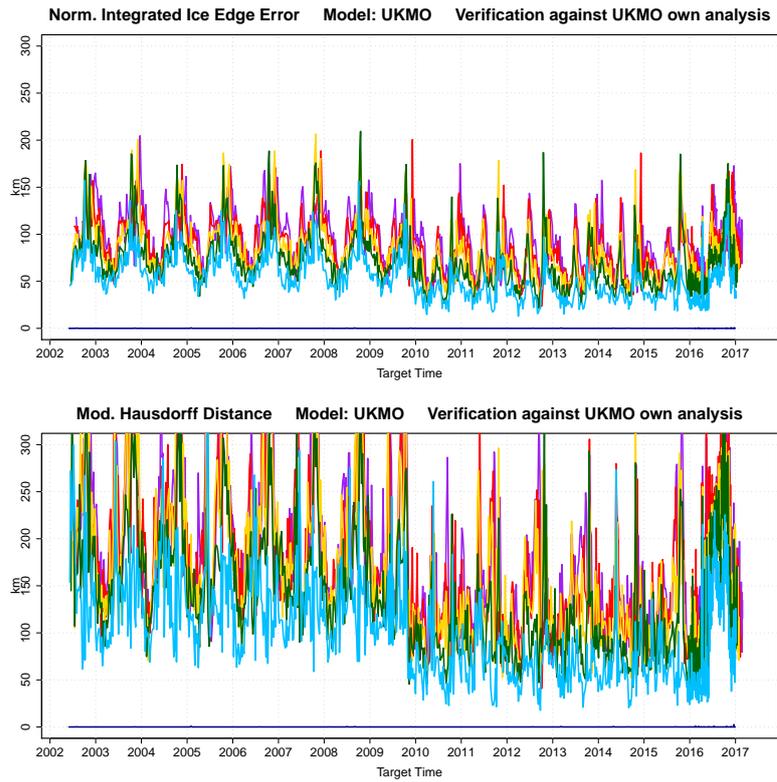


Figure 5.1: NIIEE and MHD calculated for UKMO model. Verification against UKMO own analysis

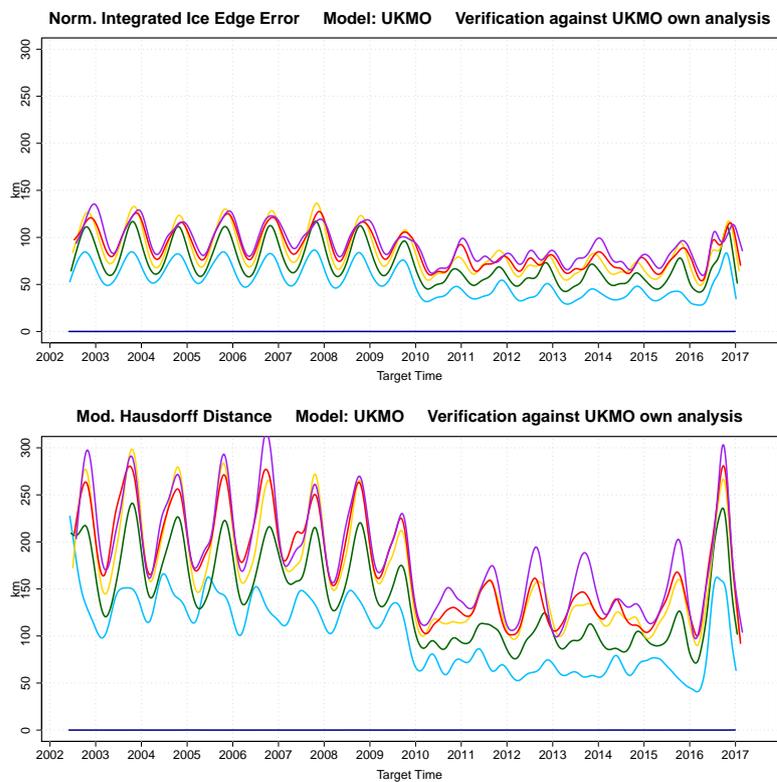


Figure 5.2: NIIEE and MHD calculated for UKMO model, with spline smoothing. Verification against UKMO own analysis

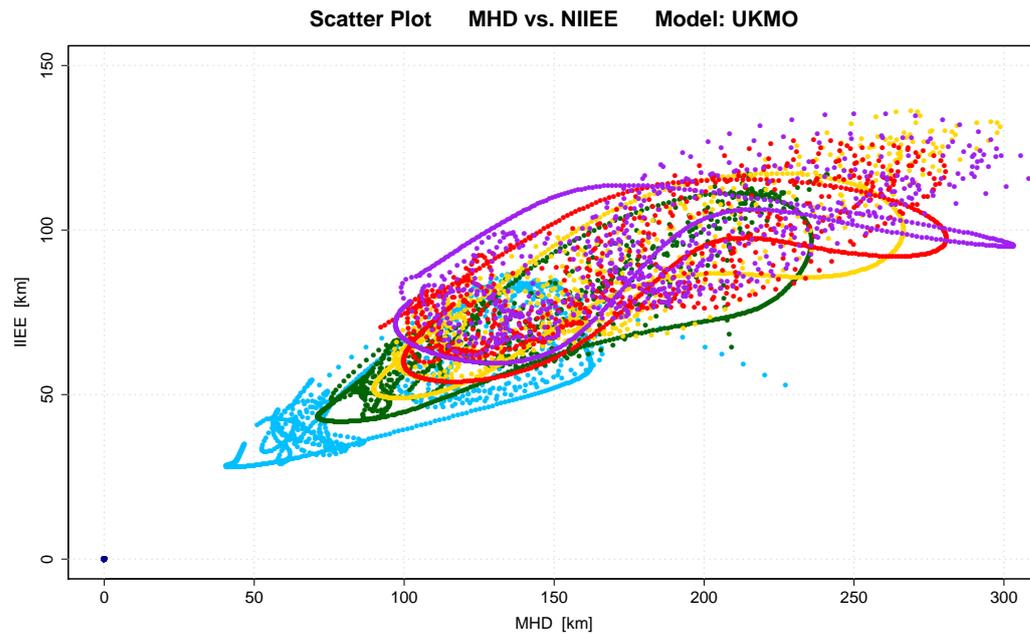


Figure 5.3: NIIEE vs. MHD scatter plot. Smoothed data. UKMO model. The colors of the points are chosen in agreement with the previous time series to indicate the six different lead times.

The test reveals a relevant positive correlation between the MHD and IIEE, both in the original data and in the smoothed data. Generally, the correlation is higher for the smoothed data which are not affected by the noise.

Two aspects of the previous results provide different indications. On one side, the metrics seem not to be consistent since the MHD is twice as big as the NIIEE, which is a discrepancy that needs to be resolved, or at least understood, for a verification analysis. On the other hand, the two metrics are highly positively correlated, suggesting indeed that the same phenomenon is captured. The natural question that arises at this point is what introduces a scaling factor of 2 between the two metrics. In general, the possible explanations are two. Either one of the two metrics is not reliable or we are making some wrong assumptions.

The first candidate that could explain the error is the effect of outliers on the MHD, which we already treated qualitatively in Sec. 2.5. However, outliers-induced errors would be randomly and not coherent, which is not what is observed. Thus, this first explanation is discarded. A better approach to solve this issue, consists in determining which verification metric produces the most meaningful results. The resolution of our data is 1.5° , which corresponds to ~ 165 km in the meridional direction. A qualitative evaluation of the forecast maps reveals a rough edge discrepancy of 1 grid cell already in the first days of the forecast. Thus, in this context, the MHD is the

metric with most reasonable results, while the NIIEE is likely underestimated the mean distance between the edges. The NIIEE underestimation could be introduced in the normalization of the IIEE. The climatological sea ice edge length was indeed calculated from high resolution satellite data, causing small scale features perpendicular to the main propagation direction of the edge to be resolved and included in the edge. These small scale feature are responsible of a substantial overestimation of the edge length, which is then reflected in the NIIEE. Further details about the computation of the climatological ice edge can be found in Appx. A. The length of the ice edge has been recalculated after interpolating the climatology to the model grid. A second option is calculating the ice edge length based on a climatology appositely built with the models own analysis. The corrected NIIEE is displayed in Fig. 5.4. The scaling factor s between MHD and NIIEE has been calculated with and without the NIIEE correction. The calculation has been performed for each lead time. As shown in Eq. 5.1, s is here defined as the ratio between the average MHD and the average NIIEE, again for each lead time separately.

$$s = \frac{\overline{\text{MHD}}}{\overline{\text{NIIEE}}} \quad (5.1)$$

The results are displayed in Tab. 5.2. The agreement between the corrected NIIEE and MHD has clearly improved, and the discrepancy is now tolerable.

Table 5.2: Scaling factor s between MHD and NIIEE, with and without NIIEE correction.

Forecast Lead Time	Scaling Factor	
	Original Data	Corrected NIIEE
Day 1	1.22	0.75
Day 8	1.94	1.18
Day 18	2.01	1.23
Day 32	2.03	1.24
Day 44	2.03	1.24
Day 60	2.00	1.23

To conclude, the NIIEE and the MHD estimation of the mean distance between the edges are comparable. This feature is relevant because could ideally allow a cross validation of the verification results with two independent metrics. However, the limitations of the two metrics should

always be carefully considered. In particular, the NIIEE is sensitive to the normalization procedure while the MHD is subject to noise, which is likely to be caused by outliers. Furthermore, it is worth to repeat that the MHD computation is much more computationally demanding than the NIIEE computation.

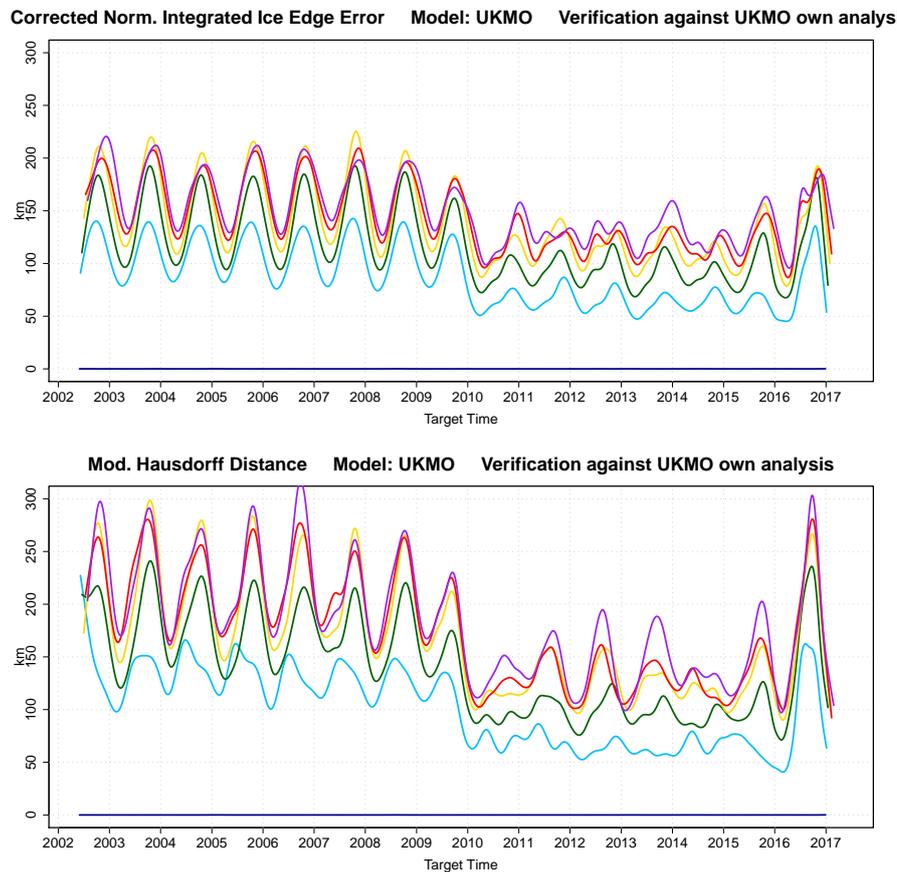


Figure 5.4: Corrected NIIEE and MHD calculated for UKMO model with spline smoothing. Verification against UKMO own analysis.

Conclusions

Sea ice forecasting is a relevant and actual research topic for several reasons. Firstly, sea ice is a fundamental component of the climate system. It regulates the exchange of heat and momentum between atmosphere and ocean and has a strong impact on the oceanic circulation, being one of the actors that drives the deep water formation in polar regions. At the same time, sea ice is strongly sensitive to climate changes and observational records show a clear decreasing trend of Arctic sea ice, both in area and thickness. The system is gradually moving from a multi-year ice regime to a first year ice regime. In this context, the development of improved models for this component could lead to a better understanding of the system itself and its interactions with atmosphere and ocean. This will certainly have positive repercussions on the climate science. A further advantage would be the possibility to better forecast relevant events, such as the extent of the September sea ice minimum, especially in those years when the sea ice conditions are far from the climatological trend.

Secondly, the gradual retreat of sea ice is leading to an increasing human activity in the Arctic regions. This makes sea ice forecasting a demanding service, especially for navigation companies, on a broad timescales range. However, a well established operational network for sea ice forecast is far from being established and improvements are still needed.

The master's thesis aims to contribute to the development of a verification system for sea ice forecasts, both in deterministic and probabilistic scenarios. We present different verification metrics, specifically designed for the evaluation of the spatial distribution of sea ice. We concentrate our attention on the verification of the sea ice edge position, which is a relevant parameter

for potential final forecast users. Our analysis is constrained to the S2S database, which collects global subseasonal-to-seasonal forecasts, including sea ice, from several operational centers and provides a promising base to test the verification metrics. The observations are a fundamental components of every verification process. Therefore the thesis includes a description of the passive microwave remote sensing measurements. These are the best option for verification purposes because of their good spatial coverage and daily availability.

The next sections summarize the main finding of our research and its scientific relevance.

6.1 | Evaluation of the verification metrics

The verification metrics that have been tested in the thesis are the Integrated Ice Edge Error (including its decompositions Misplacement Error and Absolute Extent Error), the Spatial Probability Score and the Modified Hausdorff Distance. The criteria considered to assess the quality of the metrics are listed below, arguably in order of importance.

1. **Ability to verify the ice edge position.** A good metrics should be able to quantify the correctness of the ice edge position and therefore the sea ice distribution in the Arctic.
2. **Comprehensibility.** A good metric should provide a meaningful result. In particular, real physical quantity such as areas or distances are preferred to non-dimensional indexes.
3. **Stability.** The metric should exhibit a good stability, possibly not being affected by noise.
4. **Simplicity.** The computation of the verification metric should be as easy and direct as possible.

The IIEE appears to be a good verification metric. Its formulation allows to compare efficiently the sea ice distribution in the forecasts and the observations. The decomposition into AEE and ME is particularly useful to understand the type of error affecting the forecast. Furthermore, the AEE coincides by definition with the difference of the pan-Arctic sea ice extent of the forecast and of the observations, allowing therefore the inclusion of a further relevant information in the metric. The IIEE does not capture the probabilistic nature of the ensemble forecasts. In this case, the spatial probability has first to be converted into a deterministic field, by calculating the median sea ice edge, in order to compute the IIEE.

The SPS is similar to the IIEE, since both the verification metrics are based on the calculation of disagreement areas. The main advantage of the SPS, compared to the IIEE, is the capability of dealing with a probabilistic description of both observations and forecasts. For this reason, the SPS can be considered the evolution of the IIEE in the probabilistic forecast world. However, contrarily to the IIEE, a decomposition of the SPS into AEE and ME is not possible, at least not in the same way.

In Sec. 4.1, we compare the IIEE and SPS calculated verifying the S2S database against satellite data (Figs. 4.4, 4.5, 4.6, 4.7). The two metrics show the same variations but the SPS is always slightly lower than the IIEE. This similarity confirms the consistency of the two metrics, which together represent a solid verification tool for probabilistic sea ice forecasts. An alternative formulation of the two metrics can be achieved through their normalization, specifically dividing the IIEE or SPS by the climatological length of the ice edge. These are called NIIEE and NSPS and produce an estimation of the mean distance between the forecast and observation edge, which is an interesting information for potential final users. The analysis reported in Chap. 2 reveals that the normalized versions of the metrics are sensitive to the resolution at which the climatological ice edge length is computed. To avoid mistakes, we suggest to interpolate the climatology to a grid with approximately the same resolution as the forecasts before calculating the climatological edge length. In general, biases are more likely to be introduced with the employment of high resolution sea ice concentration fields. In this case, small scale features are resolved along the edge, causing the overestimation of the total ice edge length.

Similarly to the NIIEE and NSPS, the MHD estimates the mean distance between the observation and forecast edges. Even if the MHD is able to evaluate the quality of the forecasted ice edge position, this metric seems to be less flexible than the two previous ones. In particular, the MHD can not be decomposed into an extent error and a misplacement error, which makes difficult to evaluate whether the forecasted edge position is overestimated or underestimated. The MHD is particularly sensible to outliers and thus the metric result is not as stable as the NIIEE and NSPS. A possible approach to reduce outliers-induced biases is the inclusion of the coast line in the MHD computation. However, we demonstrated that this approach fails for particular configurations when the forecast either strongly overestimates or strongly underestimates the sea ice, causing the MHD to become meaningless. Furthermore, the MHD is overestimated when the resolution of observations and forecast is different. Except for this issue, this metric is not affected by other resolution-related bias because, contrary to the NIIEE and NSPS, the MHD computa-

tion do not include any intermediate normalization step. A relevant downside of the MHD is its computational demand, which is significantly higher for this metric than it is for the NIIEE and NSPS.

It is important, at this point, to underline the relevance of the verification against the models own analysis. This procedure is significant for all the verification metrics for the following reasons.

- The error at the beginning of the forecast is set to zero by definition, allowing to study the error growth more effectively.
- Eventual errors in the data assimilation become clear. Strong discrepancies between the verification against satellite data and verification against models own analysis are the direct consequence of an inaccurate initialization of the sea ice.
- The effects of eventual changes in the forecast setup, such as improvements associated with the introduction of new observations, are more clearly captured by this verification strategy. As example, we analyzed the improvement in UKMO forecast system after the modification of the sea ice concentration data used to initialize the forecast in 2010.

6.2 | Predictive skills of the S2S forecast systems

The verification of the S2S real-time-forecasts and reforecasts against satellite data provides a robust base for the evaluation of the predictive skills of these seasonal forecast systems. The performances of the S2S models are distributed among a wide range, depending on the features of each single model and on the data assimilation strategy. The following points summarize the main aspects that influence the predictive skills of the forecast systems.

- **The forecast model.** A sea ice models which is consistent with the observations and well tuned will likely provide a better and more stable forecast than a models which is not reliable. Errors associated with model biases are recurrent in our verification analysis. As example, CMA and Météo France models exhibit too high melting rates during summer and consequently underestimate the ice edge position. On the contrary, NCEP has an high freezing rate during winter and does not melt enough ice during summer, resulting in a systematic overestimation of the sea ice edge.

- **The data assimilation procedure.** Imperfections in the assimilation of the observation into the forecast system can generally causes a discrepancy between the forecasts and the observations already at the beginning of the forecasts. This does not necessarily mean that the long term forecast will be inaccurate. However, most of the S2S models exhibit error that can be associated with the propagation of erroneous data assimilation of the sea ice concentration field into the system. In this context, it is worth to mention the overall deficient data assimilation of the CMA model, which can certainly be improved.
- **The seasonal variability.** An aspect that clearly emerges from all the S2S forecast systems is the dependence of the models predictive skills upon the seasons. The models generally exhibit better predictive skills during winter and spring, when the variability of the ice edge position is lower because the whole Arctic is frozen. In summer and autumn on the contrary, the ice edge can assume several configurations because the ice strength is less, which results in an intrinsic loss of predictability of the ice system.

The comparison of the metrics results with the climatological benchmark value shows that the predictive skills at 45-60 days of ECMWF, KMA, Météo France and UKMO forecast systems are in general moderately better than the climatological baseline. The CMA and NCEP forecast systems shows lower predictive skills than the climatology during the whole year. Overall, the persistence of the observed initial condition appears to be a better estimation of the ice edge position for the first 2-3 weeks of forecast. Furthermore, the expected benefits from an increased ensemble size could not be detected. This specific behavior is not expected, since in theory a broader ensemble would lead to a better probabilistic description of the sea ice state. However, the biases of the models and the coarse resolution of the forecast probably mask the benefit introduced by a larger ensemble.

6.3 | Outlook

This research represents one of the first attempts to conduct an extensive verification analysis of real sea ice forecasts for the seasonal timescale. Our findings suggest that the sea ice system exhibits relevant potential predictability for this timescale. The natural tool to investigate the sea ice component at this timescale is a coupled atmospheric-ocean-ice model, run in a perturbed configuration. Our work reveals that a correct probabilistic and dynamic description of the sea

ice system is more effective than a simple climatological prescription, particularly if the variable of interest is the sea ice edge.

Even though sea ice forecasts are already a product of interest for final users operating in the Arctic regions, the operational centers are in general not yet able to deliver reliable sea ice forecasts. The predictive skills of their model can certainly be improved to achieve better results. In particular, the operational forecast centers which are contributing to the S2S Prediction Project, at different level, could improve their models and data assimilation procedures. Specifically, a broader assimilation of sea ice observations, which goes beyond the simple sea ice concentration, would have a positive relapse on the skills of the systems. The growing availability and regularity of new observations, first and foremost the sea ice thickness from satellite retrievals, should be considered a primary opportunity to formulate better seasonal sea ice forecasts [Day et al. (2014)].

Seasonal sea ice forecast will certainly be a priority topic in the next years and will likely be extensively studied and developed by the scientific community. In parallel to the development of this product, a continuous monitoring of the models skills through verification analysis is certainly needed. This includes the improvement and the test of new verification metrics to achieve the creation of common protocols to evaluate our forecasts and meet the needs of the final users.

A

Climatological length of the sea ice edge

The climatological ice edge length is calculated as the median ice edge of the daily climatological sea ice concentration. The climatology is built by daily averaging the sea ice concentration fields for the period 2003-2016. The red line in Fig. A.1 correspond to the edge length calculated from the satellite data at full resolution, which is 6.250 km. The blue line refers to the ice length calculated from the satellite data interpolated to the low resolution model grid. In this second case the resolution is ~ 165 km.

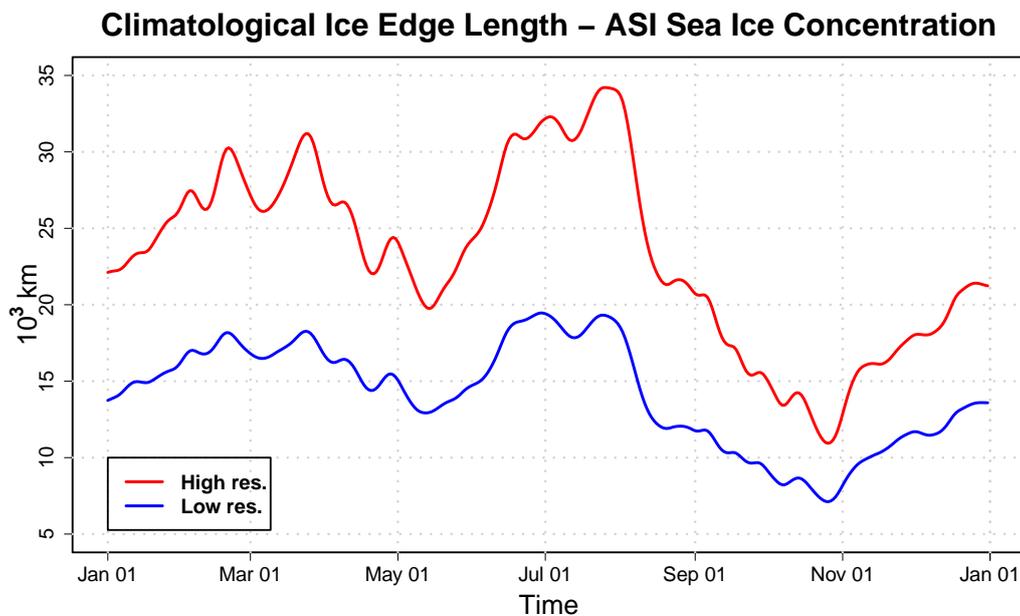


Figure A.1: Climatological ice edge length. Computation based on low resolution edge (blue line) and high resolution edge (red line). The lines are smoothed by applying a spline interpolation algorithm.

The discrepancy between the two curves is caused by small scale features that are resolved in the high resolution configuration but do not appear in the low resolution version. These features are recognized by the edge detection algorithm and they cause an increase of the edge length.

The algorithm that has been used to compute the edge length is based on a contour recognition function. This extracts those grid points of the sea ice concentration field (*sic*) with $sic \geq 0.15$ that have at least one neighbor with $sic < 0.15$. The points are then ordered based on multiple proximity criteria to form a contour. Several segments are detected, depending on the complexity of the edge. Afterwards, the length of each single segment is computed and summed up. Short segment are discarded to avoid the inclusion of either small scale ice islands or holes in the sea ice coverage. Note that the points of each segment have to be in the correct order to obtain meaningful results. The original curves present a modest noise level. This is filtered out by applying a smoothing algorithm based on spline interpolation. In this way, the values obtained are stable and do not generate a further increase in the noise of the verification metrics.



Relevant Acronyms

AEE Absolute Extent Error.

AMSR Advanced Microwave Scanning Radiometer.

ASI ARTIST Sea Ice.

BoM Bureau of Meteorology.

CMA China Meteorological Administration.

CMID Climatological median ice edge.

CSPS Climatological spatial probability score.

ECCC Environment and Climate Change Canada.

ECMWF European Centre for Medium-Range Weather Forecasts.

EE Extent Error.

GIOPS Global Ice Ocean Prediction System.

GODAE Global Data Assimilation Experiment.

HD Hausdorff Distance.

HMCR Russian Academy of Science.

IIEE Integrated Ice-Edge Error.

ISAC-CNR Italian Institute of Atmospheric Sciences and Climate.

JMA Japan Meteorological Agency.

KMA Korean Meteorological Administration.

ME Misplacement Error.

MetFrance Meteo France.

MHD Modified Hausdorff Distance.

NCEP National Centers for Environmental Prediction.

NEMO Nucleus for European Modelling of the Ocean.

NOAA National Oceanic and Atmospheric Administration.

O Overestimation.

PER1 Persistence from the previous year.

PERF Persistence from forecast initial time.

S2S Sub-seasonal to Seasonal Prediction Project.

SIO Sea Ice Outlook.

SPS Spatial Probability Score.

SST Sea-surface Temperature.

U Underestimation.

UKMO UK Met Office.

YOPP Year of Polar Prediction.



Bibliography

- P. Bauer et al. *WWRP Polar Prediction Project Implementation Plan for the Year of Polar Prediction (YOPP), Version 2.0*. World Meteorological Organization (WMO), Geneva, Switzerland, 2016.
- S. Bouillon, M. M. Maqueda, V. Legat, and T. Fichefet. An elastic-viscous-plastic sea ice model formulated on Arakawa B and C grids. *Ocean Modeling*, 2009. doi: 10.1016/j.ocemod.2009.01.004.
- D. J. Cavalieri, P. Gloersen, and W. J. Campbell. Determination of Sea Ice Parameters With the NIMBUS 7 SMMR. *Journal of Geophysical Research*, 1984. doi: 10.1029/JD089iD04p05355.
- J. C. Comiso. Characteristics of Arctic Winter Sea Ice From Satellite Multispectral Microwave Observations. *Journal of Geophysical Research*, 1986. doi: 10.1029/JC091iC01p00975.
- J. C. Comiso, D. J. Cavalieri, P. C. L., and P. G. P. Passive microwave algorithms for sea ice concentration: A comparison of two techniques. *Remote Sensing Environment*, 1997. doi: 10.1016/S0034-4257(96)00220-9.
- J. Day, E. Hawkins, and S. Tietsche. Will Arctic sea ice thickness initialization improve seasonal forecast skill? . *Geophysical Research Letters*, 2014. doi: 10.1002/2014GL061694.
- C. J. Donlon, M. Martin, J. D. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer. The Operational Sea Surface Temperature and Sea Ice analysis (OSTIA). *Remote Sensing of the Environment*, 2011. doi: 10.1016/j.rse.2010.10.017.
- D. S. Dukhovskoy, J. Ubnoske, E. Blanchard-Wrigglesworth, H. R. Hiester, and A. Proshutinsky. Skill metrics for evaluation and comparison of sea ice models. *Geophysical Research Letters*, 2015. doi: 10.1002/2015JC010989.

- C. Emmerson and G. Lahn. *Arctic Opening: Opportunity and Risk in the High North*. Lloyd's, London, United Kingdom, 2012.
- H. F. Goessling, S. Tietsche, J. J. Day, E. Hawkins, and T. Jung. Predictability of the Arctic sea ice edge. *Geophysical Research Letters*, 2016. doi: 10.1002/2015GL067232.
- S. M. Griffies, M. J. Harrison, R. C. Pakanowski, A. Rosati, Z. Lang, M. Schmidt, H. Simmons, and R. Slater. A Technical Guide to MOM4. 2008.
- E. C. Hunke and W. H. Lipscomb. CICE: the Los Alamos Sea Ice Model Documentation and Software User's Manual Version 4.1. 2010. URL http://csdms.colorado.edu/w/images/CICE_documentation_and_software_user%27s_manual.pdf.
- IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- T. Jung and M. Matsueda. Verification of global numerical weather forecasting systems in polar regions using TIGGE data. *Quarterly Journal of the Royal Meteorological Society*, 2016. doi: 10.1002/qj.2437.
- T. Jung et al. *WWRP Polar Prediction Project Implementation Plan*. World Meteorological Organization (WMO), Geneva, Switzerland, 2014.
- G. Madec. *NEMO ocean engine*. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN No 1288-1619, 2008.
- R. Mladek, F. Vitart, and M. Manoussakis. URL <https://software.ecmwf.int/wiki/display/S2S/Models>.
- K. Mogensen, M. Balmaseda, A. Weaver, M. Martin, and A. Vidard. NEMOVAR: A variational data assimilation system for the NEMO model. *ECMWF Newsletter*. 120: 17–22, 2009.
- K. Mogensen, M. A. Balmaseda, and A. Weaver. The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. *ECMWF Technical Memoranda*, 2012.
- D. Notz. Sea-ice extent and its trend provide limited metrics of model performance. *The Cryosphere*, 2014. doi: 10.5194/tc-8-229-2014.
- K. Partington, T. Flynn, D. Lamb, C. Bertioia, and K. Dedrick. Late twentieth century Northern Hemisphere sea-ice record from U.S. National Ice Center ice charts. *Journal of Geophysical Research*, 2003. doi: 10.1029/2002JC001623.

- R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang. An improved in situ and satellite SST analysis for climate. *Journal of Climate*, 2002. doi: 10.1175/1520-0442(2002)015.
- G. C. Smith et al. Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 2016. doi: 10.1002/qj.2555.
- L. C. Smith and S. R. Stephenson. New Trans-Arctic shipping routes navigable by midcentury. *PNAS*, 2013. doi: 10.1073/pnas.1214212110.
- G. Spreen, L. Kaleschke, and G. Heygster. Sea ice remote sensing using AMSR-E 89-GHz channels. *Journal of Geophysical Research*, 2008. doi: 10.1029/2005JC003384.
- J. Stroeve, L. C. Hamilton, C. M. Bitz, and E. Blanchard-Wrigglesworth. Predicting September sea ice: Ensemble skill of the SEARCH Sea Ice Outlook 2008–2013. *Geophysical Research Letters*, 2014. doi: 10.1002/2014GL059388.
- F. Vitart, A. W. Robertson, and D. L. T. Anderson. Subseasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *WMO Bulletin*, 2012.
- F. Vitart et al. The Sub-seasonal to Seasonal Prediction (S2S) Project Database. *Bulletin of the American Meteorological Society*, 2016. doi: 10.1175/BAMS-D-16-0017.1.
- D. Wilks. *Statistical Methods in the Atmospheric Sciences, Second Edition*. Academic Press, Elsevier, Oxford, United Kingdom, 2006.
- M. Winton. A Reformulated Three-Layer Sea Ice Model. *Journal of Atmospheric and Oceanic Technology*, 2000. doi: 10.1175/1520-0426(2000)017<0525:ARTLSI>2.0.CO;2.