

Uncertainty estimation for operational ocean forecast products—a multi-model ensemble for the North Sea and the Baltic Sea

Inga Golbeck¹ · Xin Li¹ · Frank Janssen¹ · Thorger Brüning¹ · Jacob W. Nielsen² · Vibeke Huess² · Johan Söderkvist³ · Bjarne Büchmann³ · Simo-Matti Siiriä⁴ · Olga Vähä-Piikkiö⁴ · Bruce Hackett⁵ · Nils M. Kristensen⁵ · Harald Engedahl⁵ · Ed Blockley⁶ · Alistair Sellar⁶ · Priidik Lagemaa⁷ · Jose Ozer⁸ · Sebastien Legrand⁸ · Patrik Ljungemyr⁹ · Lars Axell⁹

Received: 18 February 2015 / Accepted: 8 September 2015 / Published online: 21 October 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Multi-model ensembles for sea surface temperature (SST), sea surface salinity (SSS), sea surface currents (SSC), and water transports have been developed for the North Sea and the Baltic Sea using outputs from several operational ocean forecasting models provided by different institutes. The individual models differ in model code, resolution, boundary conditions, atmospheric forcing, and data assimilation. The ensembles are produced on a daily basis. Daily statistics are calculated for each parameter giving information about the spread of the forecasts with standard deviation, ensemble mean and median, and coefficient of variation. High forecast uncertainty, i.e., for SSS and SSC, was found in the

Skagerrak, Kattegat (Transition Area between North Sea and Baltic Sea), and the Norwegian Channel. Based on the data collected, longer-term statistical analyses have been done, such as a comparison with satellite data for SST and evaluation of the deviation between forecasts in temporal and spatial scale. Regions of high forecast uncertainty for SSS and SSC have been detected in the Transition Area and the Norwegian Channel where a large spread between the models might evolve due to differences in simulating the frontal structures and their movements. A distinct seasonal pattern could be distinguished for SST with high uncertainty between the forecasts during summer. Forecasts with relatively high deviation from the multi-model ensemble (MME) products or the other individual forecasts were detected for each region and each parameter. The comparison with satellite data showed that the error of the MME products is lowest compared to those of the ensemble members.

Responsible Editor: Emil Vassilev Stanev

✉ Inga Golbeck
inga.golbeck@bsh.de

- ¹ Bundesamt für Seeschifffahrt und Hydrographie, Hamburg, Germany
- ² Danish Meteorological Institute, Copenhagen, Denmark
- ³ Defence Centre for Operational Oceanography, Ballerup, Denmark
- ⁴ Finnish Meteorological Institute, Helsinki, Finland
- ⁵ Norwegian Meteorological Institute, Oslo, Norway
- ⁶ Met Office, Exeter, UK
- ⁷ Marine Systems Institute at Tallinn University of Technology, Tallinn, Estonia
- ⁸ Operational Directorate Natural Environment, Royal Belgian Institute of Natural Sciences, Brussels, Belgium
- ⁹ Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Keywords Multi-model ensemble · Uncertainty estimates · Ocean forecast · Baltic Sea · North Sea · Satellite data

1 Introduction

Several operational ocean forecasting models are currently available for the North Sea and the Baltic Sea providing a wide range of realizations for the uncertain future situation. There is a strong demand to make the best out of the available forecasts, e.g., for sea level warnings or oil drift forecasts. One solution now commonly applied in weather forecasting is the estimation of forecast uncertainties with the aid of ensemble prediction systems (EPSs). There exist several different types

of EPS. A single-model EPS uses one model with perturbed initial, boundary, and/or forcing conditions and provides a more skillful indication of how likely an event occurs compared to single forecasts (Toth and Kalnay 1993; Molteni et al. 1996). But, this approach assumes that the model itself is well verified and that uncertainty arises only from errors in the applied conditions. Aside from the high computational effort, another disadvantage of this method is the difficulty in attaining a sufficient spread of the ensemble and thereby missing the full range of uncertainty (Houtekamer et al. 1996). Also, systematic biases or errors in model parameterizations can impact the skill of EPS (Molteni et al. 1996). Another method is to combine single-model ensembles from different models, creating a multi-model multi-analysis ensemble (MMAE) that has more skill than any one single-model EPS also due to an increased ensemble spread (Evans et al. 2000; Richardson 2001; Mylne et al. 2002). A third approach is the construction of a so-called poor-man's ensemble system (PEPS) using independent forecast models from different operational centers. An advantage of PEPS is that the model uncertainty can be sampled through the variety of model resolutions, model numerics and physical formulations, initialization methods, boundary data, and forcing data (Ebert 2001). Since the PEPS members typically are operational model runs, this approach has little added computational cost. Compared to a single-model EPS with perturbed initial conditions, PEPS is not prone to systematic biases and often has the advantage of higher spatial resolution in the individual member models (Ebert 2001). In several studies, PEPS has been compared to single-model EPS for short-range forecast, with the result that the skillful PEPS is shown to be highly competitive with the EPS (Atger 1999; Ziehmman 2000; Ebert 2001; Buizza et al. 2003; Arribas et al. 2005). However, the main disadvantages of the PEPS are the low ensemble size and the fact that one model with low forecast skill might have a strong negative impact on the whole ensemble system (Ebert 2001). Ziehmman (2000) also found a low number of contributing independent models to be a limiting factor for operational use. Nevertheless, the equally weighted four-member PEPS outperformed larger ensembles in some key aspects. A modified and improved approach was conducted by Krishnamurti et al. (1999) who developed a PEPS by applying a multiple regression technique on each forecast in order to determine the optimal weight of the models. The so-called super-ensemble outperformed other models to which it was compared.

First approaches to PEPS-type ensemble systems have also been developed for ocean forecasting. In 2000, partners of the Northwest European Shelf Operational Oceanographic System (NOOS, www.noos.cc, accessed 24 October 2014) established an exchange of surge forecasts as well as water level measurements in order to support the national water level forecasting services in the NOOS area. Later, in 2007,

a weighting method, Bayesian model averaging (BMA), was applied on the Multifunctional Access Tool for Operational Ocean data Services (MATROOS) system to gain more information about model uncertainty (Becker 2007; Ebel and Becker 2010). The Ensemble Surge Forecast (ENSURF) system was further developed by Pérez et al. (2012), by applying BMA to independent operational sea level forecasts in the region of the Ireland-Biscay-Iberia Regional Operational Oceanographic System (IBI-ROOS) and the western Mediterranean coast. Recently, the Group for High Resolution Sea Surface Temperature (GHRSSST) developed a Multi-Product Ensemble (GMPE) for the global ocean by using various individual level 4 SST analyses and calculating the ensemble median and standard deviations. A comparison to independent Argo data demonstrated that the GMPE median yields a more accurate estimate of SST than the individual analyses (Martin et al. 2012). Weisheimer et al. (2009) used five equally weighted coupled atmosphere–ocean circulation models to study Pacific SST by comparison with a previous-generation ensemble, DEMETER (Palmer et al. 2004; Doblas-Reyes et al. 2005), yielding a higher skill for the new multi-model ensemble. More weighted 3D multi-model ensembles have been developed for SST forecasts by applying BMA or a Kalman Filter over a learning period for determining the optimal weights between the models (Logutov and Robinson 2005; Raftery et al. 2005; Lenartz et al. 2010; Mourre et al. 2012). The super-ensembles have been validated against in situ data of CTD, gliders, drifter, and scan fish.

The main goal of this paper is to present a new multi-model approach for the North Sea and the Baltic Sea, which is used to illustrate uncertainties between operational ocean forecasting products. The new PEPS, hereafter referred to as multi-model ensemble (MME), uses outputs from existing operational ocean forecasting models as provided by the modeling groups, and all models have individual model codes, resolution, boundary conditions, atmospheric forcing, and methods for data assimilation. The uncertainties are described on a temporal and spatial scale by ensemble statistics and spatio-temporal statistics. The aim is to identify the amount, spatial, and temporal distribution of uncertainties for several physical parameters and by this to provide some added value to the users of the single-model forecasts. It has to be noted that computation of a best estimate for all parameters, which would need more sophisticated averaging methods, or the in-depth explanation of the causes of uncertainties, which would need full access to the four dimensional model outputs, is beyond the scope of this paper.

The development of the MME was done in the framework of the MyOcean project, funded by the EU research framework programme (FP7) (<http://www.myocean.eu/>, accessed 24 October 2014), and is now continued in the Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu>, accessed 30 June 2015). The MME was developed in the framework of two MyOcean work

packages forming the regional monitoring and forecasting centers for the Northwest Shelf (NWS) and the Baltic Sea (BAL) which have been transformed to parts of CMEMS in May 2015. The nominal forecast products for the two regions, namely, FOAM_AMM for NWS and DMI HBM for BAL (see Table 1), are implemented in the MME. One goal in both MyOcean work packages was to provide additional uncertainty information on the nominal products. In addition, the MME is now established as an independent service, taking advantage of the various existing operational ocean forecasting models, and benefitting the participating agencies and institutes. This service is basically a supplement to validation and provides a comparison of the contributing forecasts in order to reveal the degree of agreement and deviation for different parameters. The comparison is done on a daily basis to keep track of the actual variations and to detect potential problems in individual model systems. Based on the daily and long-term results, the model systems can be improved and further developed.

To enhance the sustainability and user uptake of the MME, the development is done in close cooperation with the communities of the Baltic Operational Oceanographic System (BOOS, www.boos.org, accessed 24 October 2014) and the Northwest European Shelf Operational Oceanographic System (NOOS, www.noos.cc, accessed 24 October 2014). These two communities are regional services integrated in the European Global Ocean Observing System (EuroGOOS). Both systems focus on the provision and improvement of high-quality operational marine data.

This study is carried out for the North Sea and the Baltic Sea which are connected by a Transition Area, the Skagerrak and Kattegat. The Baltic Sea is characterized by brackish waters with a surface salinity around 20 in the south decreasing towards the north and east (about 2 in the Bothnian Sea and the eastern Gulf of Finland). During winter months, the northern parts of the Baltic Sea are regularly covered by sea ice. The surface salinity is influenced by freshwater inflow from rivers and melting sea ice in spring (Feistel et al. 2008; Leppäranta and Myrberg 2009). The exchange of water masses between the North Sea and Baltic Sea is characterized by high-saline water entering the Baltic Sea via the Great Belt, the Little Belt, and the Oresund by near-bottom currents. Low-saline surface water is flowing out of the Baltic Sea. Sea surface currents are mainly induced by wind and density gradients, as well as by differences of water level. The dominant feature of currents in the North Sea is the tidal motion (Otto et al. 1990). The residual circulation is characterized by a major inflow from the North Atlantic and the English Channel and a major outflow from the Baltic Sea as the Norwegian Coastal Current. The surface salinity averages between 34 and 35 in the central North Sea. There are freshwater inflows from rivers, such as Rhine and Elbe,

and from the Baltic Sea affecting the surface salinity. The surface temperature has a strong annual cycle in both regions.

The MME systems of the North Sea and Baltic Sea as well as the contributing models are presented in Sect. 2. Ensemble statistics of the MME are explained and some examples are displayed in Sect. 3. The uncertainty estimates between the products are based on spatio-temporal statistics of the data collected. As a result, regions with high and low uncertainties as well as seasonal patterns can be identified. A comparison to satellite data is presented. Results of the spatio-temporal statistics are shown in Sect. 4 and a summary is given in Sect. 5.

2 MME system

2.1 Overview of contributing models

Thirteen different operational ocean forecasting models covering either the North Sea or the Baltic Sea or both regions contribute to the MME. Details on model area, boundary conditions, and forcing are listed in Table 1. A brief overview of each system is provided below. It should be noted, however, that the forecasts are not fully independent of each other since most of the models covering the Baltic Sea are based on the same kernel of model code (CMOD) and are therefore related to a certain degree (Berg and Poulsen 2012). Furthermore, some models are using the same forcing and boundary conditions. Accordingly, it could be expected that the statistical evaluation might be influenced by this dependency.

CMOD and HBM at BSH The Federal Maritime and Hydrographic Agency (BSH) runs two forecast models with dynamical vertical coordinates covering the North and Baltic Sea: the operational Circulation Model CMOD (Dick et al. 2001; Dick and Kleine 2008) and the pre-operational HIROMB-BOOS Model, HBM (Berg and Poulsen 2012). Both model setups consist of a coarse grid with a horizontal resolution of 3 nautical miles (NM) and a maximum of 35 vertical layers, and a two-way nested fine grid with a horizontal resolution of 0.5 NM and a maximum of 25 vertical layers covering the inner German Bight and Western Baltic. While CMOD uses a simple algebraic turbulence model (Kleine 1994), HBM runs with a k-omega turbulence model (Berg 2012).

DKSS2013 and HBM at DMI The Danish Meteorological Institute (DMI) runs the storm surge model DKSS2013 and HBM as the nominal MyOcean product, which both cover the North Sea and Baltic Sea (Berg and Poulsen 2012). The model runs on a two-way nested rectangular grid. The horizontal resolution of DKSS2013 is 3 NM in the main domain, 1 NM in the Wadden Sea, and 0.5 NM in the Transition Area. The model setup consists of vertical z-coordinates with a

Table 1 Overview of general model settings

Institute, model version	Region; approx. coordinates	Horizontal res. lon × lat (°)	Vertical res./no. of layers (thickness)	Atmospheric forcing	Open boundary conditions	Other forcing (RR=river runoff, S=salinity, T=temp.)	Data assimilation	Key references
BSH, CMOD v4	NWS, BAL 48.6° N/4.0° W 65.9° N/30.4° E	1/12×1/20 [3 nm]	dynamical z-coordinates/25 (fine), 35 (coarse) layers	COSMO-EU 7 km; DWD	2D surge model BSHsmod.na, 10 km	RR by HBV (SMHD), true RR from BFG	None	(Kleine 1994; Dick et al. 2001; Dick and Kleine 2008)
BSH, HBM v4 (based on CMOD)	NWS, BAL 48.6° N/4.0° W 65.9° N/30.4° E	1/12×1/20 [3 nm]	dynamical z-coordinates/25 (fine) 35 (coarse) layers	COSMO-EU 7 km; DWD	2D surge model BSH-HBM_v4.na, 10 km	RR by HBV (SMHD), true RR from BFG	None	(Berg 2012; Berg and Poulsen 2012; Brüning et al. 2014)
DML, DKSS2013 (based on HBM)	NWS, BAL 48.5° N/4.1° W 65.9° N/30.2° E	1/12.05×1/20 [3 nm], 0.5 nm Transition Area, 1 nm Wadden Sea	z-coordinates/52 layers (8 m surface layer, 2 m in Transition Area)	HIRLAM-SKA 3 km; ECMWF	2D surge model North Atlantic	RR by HBV (SMHD), true RR from BFG	None	(Berg and Poulsen 2012)
DML, HBM v4	NWS, BAL 48.5° N/4.1° W 65.5° N/30.3° E	1/36×1/60 [1 nm] BAL, 0.5 nm Transition Area, 3 nm NWS	z-coordinates/25 layers	HIRLAM	2D surge model North Atlantic NOAMOD	RR by HBV (SMHD), true RR from BFG	None	(Berg and Poulsen 2012)
FCOO, GETM v6 (using GOTM)	NWS, BAL 50.1° N/4.1° W 65.9° N/30.3° E	1/36×1/60 [1 nm], 600 m Kattegat-Arkona	general vert. coordinates/60 layers (<0.45 m surface layer)	HIRLAM SKA 3 km; DMI	2D storm surge model NA3, GETM-UTILS	T & S by monthly climatological values, RR by HBV (SMHD), true RR by BSH	None	(Burchard et al. 1999, 2009, 2010; Büchmann et al. 2011)
FMI, HBM v2	NWS, BAL 48.5° N/4.1° W 65.9° N/30.3° E	1/12×1/20 [3 nm], 0.5 nm Danish Straits, Wadden Sea	z-coordinates/50 layers (8 m surface layer)	HIRLAM	Inverse barometer calculation for water levels	Climatological values for T & S, RR by climatology	None	(Berg and Poulsen 2012; Poulsen and Berg 2012)
FMI, HBM v2	NWS, BAL 48.5° N/4.1° W 65.9° N/30.3° E	1/12×1/20 [3 nm], 0.5 nm Danish Straits, Wadden Sea	z-coordinates/50 layers (8 m surface layer)	ECMWF	Inverse barometer calculation for water levels	Climatological values for T & S, RR by climatology	None	(Berg and Poulsen 2012; Poulsen and Berg 2012)
Met Office, FOAM AMM v7 (based on NEMO)	NWS 40° N/19.9° W 65.0° N/13.0° E	1/9×1/15 [~7 km]	s-coordinates/32 layers	Met Office Global Unified Model	Tidal model from NE Atlantic, 15 constituents	RR by climatological values; T, S, sea level, currents from FOAM 1/12° NA model; Baltic outflow=river	SST: in situ & satellite sensors SEVIRI, NOAA-AVHRR, METOP-AVHRR	(Maded 2008; O'Dea et al. 2012)
MET Norway, ROMS v3.5	NWS 40° N/19.9° W 65.0° N/13.0° E	1/9×1/15 [~7 km]	s-coordinates/35 layers	HIRLAM, 12 km & ECMWF	Barotropic, baroclinic, MET Arctic-20 km	Tides from TPXO v7.2, RR by climatological values, Baltic outflow=river (18, monthly climatology)	SST: OSTIA SST analysis	(Shepepkin and McWilliams 2005)
MSI, HIROMB v4.5 (EST setup)	Gulf of Riga & Finland 56.9° N/21.6° E 60.7° N/30.3° E	1/60×1/120 [0.5 nm]	z-coordinates (3 m layer thickness in 0-90 m, 5 m in 90-120 m)	HIRLAM, 11 km, EWS	Along western bound 21.55°E, 1 nm HIROMB-BS01	Climatic mean RR	None	(Funkquist and Kleine 2007; Lagemaa 2012)
RBINS, OPTOS NOS v1 (based on COHERENS v1)	NWS 48.5° N/3.9° W 57.0° N/9.1° E	1/12×1/24 [~6 km]	s-coordinates/20 layers	Met Office NWP	2D storm surge model OPTOS_CSM	Neumann condition for T, Ditchlet condition for S (ICES), RR by monthly climatological values	None	(Luyten et al. 1999; Luyten 2014)
SMHI, HIROMB v4.6 NS03 setup	NWS, BAL 48.4° N/4.2° W 65.9° N/30.3° E	1/12×1/20 [3 nm]	z-coordinates/4 m surface layer	HIRLAM, 11 km	2D storm surge model 24 nm NOAMOD	RR by HBV, 73 rivers specified; T, S from met.no ocean model or climatology	SST: observation from Sea Ice Service, OSI-SAF; T, S: observation from ship, buoys, ferry boxes	(Daley 1991; Funkquist and Kleine 2007; Axell 2013)
SMHI, HIROMB v4.6 BS01 setup	BAL 53.7° N/9.4° E 65.9° N/30.3° E	1/36×1/60 [1 nm]	z-coordinates/4 m surface layer	HIRLAM, 11 km	Exchange with NS03 set-up at western bound	RR by HBV; 56 rivers specified	SST: observation from Sea Ice Service, OSI-SAF; T, S: observation from ship, buoys, ferry boxes	(Daley 1991; Funkquist and Kleine 2007; Axell 2013)

maximum of 52 layers and finest resolution of 2 m (Transition Area, 1 m) coarsening toward the sea bed. The thickness of the top layer is 8 m in general but reduced to 2 m in the Transition Area. The horizontal resolution of HBM is 1 NM and the number of vertical layers is 122, where only 25 layers are provided in the MyOcean product.

GETM at FCOO The Danish Defence Centre for Operational Oceanography (FCOO) runs three nested setups (Büchmann et al. 2011) of the General Estuarine Transport Model, GETM, in operational production (Burchard et al. 2009, 2010). The three GETM setups are configured as one-way nesting, with differing horizontal resolutions of 3 NM, 1 NM, and 600 m. The 1-NM North Sea–Baltic Sea setup and the 600-m setup, covering the Kattegat–Arkona region, are both baroclinic setups, which use 60 layers of general vertical coordinates (Hofmeister et al. 2011) with zooming toward surface and sea bed. The maximum thicknesses of the upper layers in the Skagerrak are 0.45, 0.6, 0.8, 1.0, 1.25, and 1.42 m. Elsewhere, the thicknesses of the vertical layers are thinner.

HBM at FMI The Finnish Meteorological Institute (FMI) uses HBM in operational mode covering the North Sea and Baltic Sea (Berg and Poulsen 2012; Poulsen and Berg 2012). The grid of the baroclinic model consists of regular horizontal coordinates with a 3-NM resolution and up to 50 depth layers, and a two-way nesting with a 0.5-NM grid, covering the Danish Straits and the Wadden Sea. The thickness of the surface layer is 8 m. Two separate model runs are made with different atmospheric forcing using European Centre for Medium-Range Weather Forecasts (ECMWF) and High Resolution Limited Area Model (HIRLAM).

FOAM_AMM at the Met Office The Met Office runs a coupled hydrodynamic-biogeochemical Forecasting Ocean Assimilation Model 7 km Atlantic Margin Model (FOAM-AMM) (O’Dea et al. 2012) covering the Northwest Shelf, including Skagerrak and Kattegat, and parts of the North-East Atlantic. In the current model version the Little Belt, Great Belt, and the Sound are defined as big rivers for transition to the Baltic Sea. The model is run on a regular horizontal grid with about 7-km resolution. The vertical resolution of 32 levels is determined by a hybrid s-sigma terrain following coordinate system (following Song and Haidvogel (1994)). Sea surface temperature (SST) is assimilated utilizing infra-red satellite observations from the SEVIRI, NOAA-AVHRR, and METOP-AVHRR instruments along with in situ measurements.

ROMS at MET Norway The Norwegian Meteorological Institute (MET Norway) runs the Regional Ocean Modeling System (ROMS) covering the Northwest Shelf, including Skagerrak and Kattegat, and the Nordic Seas (Shchepetkin

and McWilliams 2005). The model setup consists of a horizontal grid with orthogonal polar-stereographic coordinates of 4-km resolution and a vertical S-coordinate system with 35 levels. The transition to the Baltic Sea is handled the same way as for FOAM_AMM in the current model version of ROMS. Data assimilation is applied for SST using OSTIA SST analysis. For the MME, the data are interpolated to the Met Office FOAM_AMM horizontal grid (~7 km).

HIROMB at MSI The Marine Systems Institute (MSI) uses the baroclinic and eddy-resolving High-Resolution Operational Model for the Baltic Sea (HIROMB) (Funkquist and Kleine 2007) in operational mode for forecasts in Estonian marine areas, the Gulf of Finland, and the Gulf of Riga (HIROMB-EST). The model setup itself is configured without nested grids and uses the boundary conditions of HIROMB-BS01, described below. The horizontal resolution is 0.5 NM. It uses fixed z-coordinates with a vertical resolution of 3 m from the surface down to 90-m depth and 5 m between 90- and 135-m depth.

OPTOS_NOS at RBINS The Royal Belgian Institute of Natural Sciences (RBINS) runs OPTOS_NOS, covering the English Channel and the southern North Sea, nested with the high-resolution OPTOS_BCZ, covering the Belgian waters and its approaches from Dunkirk to Rotterdam. The model setup consists of a regular latitude–longitude grid with about 6-km resolution, and a vertical sigma coordinate system consisting of 20 layers (Luyten et al. 1999).

HIROMB at SMHI The Swedish Meteorological and Hydrological Institute (SMHI) runs the baroclinic model HIROMB (Funkquist and Kleine 2007; Axell 2013) with two different configurations: The NS03 grid has 3-NM horizontal resolution, covering the North Sea and Baltic Sea, and the BS01 grid has a horizontal resolution of 1 NM, covering only the Baltic Sea, Kattegat, and Skagerrak. The horizontal grid is set up on regular coordinates and the vertical grid consists of z-coordinates. Data assimilation is applied for SST, using observations analyzed by the Swedish Ice Service as well as from OSI-SAF, near-real-time in situ temperature (T) and salinity (S) profiles from the Finnish research ship Aranda, real-time in situ S/T profiles from Swedish and German buoys, and in situ surface measurements of S and T from Ferry Boxes on several merchant ships and ice breakers.

2.2 MME processes and ensemble statistics

Since April 2013, most of the above mentioned forecasts have been provided by the partners to the MME system. It is important to note that the contribution to the MME is made on voluntary basis. As a consequence, not all partners started to deliver all parameters at the same time; hence, some forecasts

Table 2 List of physical parameters, spatial coverage, and start of delivery (5 m mean) of data, provided by the partners for the MME

Institute, Model Version	MME Output				Region	Start of delivery (5 m)	
	SST	SSS	SSC	TRA			
BSH, CMOD v4	X	X	X	X	North Sea	Baltic Sea	01.2013
BSH, HBM v4	X	X	X	X	North Sea	Baltic Sea	02.2013
DMI, DKSS2013	X	X	X		North Sea	Baltic Sea	12.2013
DMI, HBM v4	X	X	X	X		Baltic Sea	02.2013
FCOO, GETM v6	X	X	X	X	North Sea	Baltic Sea	01.2013
FMI, HBM v2 - HIRLAM	X	X	X			Baltic Sea	04.2014
FMI, HBM v2 - ECMWF	X	X	X			Baltic Sea	04.2014
Met Office, FOAM_AMM v7	X	X	X	X	North Sea		01.2013
MET Norway, ROMS v3.5	X	X	X		North Sea		02.2014
MSI, HIROMB v4.5	X	X	X			Baltic Sea	04.2013
RBINS, OPTOS_NOS v1	X	X		X	North Sea		04.2013
SMHI, HIROMB v4.6 NS03	X	X	X		North Sea	Baltic Sea	09.2014
SMHI, HIROMB v4.6 BS01	X	X	X			Baltic Sea	09.2014

The spatio-temporal statistics are calculated for different study periods defined in Sect. 2.3

were included later (see Table 2). MMEs are produced separately for four physical parameters: sea surface temperature (SST), sea surface salinity (SSS), sea surface currents (SSC), and water transport (TRA). For SST, SSS, and SSC, hourly model results for 48-h forecasts have been used, starting from forecast time step at 01 h (Fig. 1). Since the thickness of the surface layer differs in space and time within the models and among the models, due to different vertical coordinate systems, it was suggested to provide a 5-m mean of the upper model layers for SST, SSS, and SSC. In this way, the data sets are better comparable and a compromise is made between optimal and available vertical resolution. The exchange of daily water transport through a series of transects in the North Sea and the Baltic Sea started earlier in the NOOS and BOOS communities and is used for the MME. The calculation of daily TRA is described in Sect. 2.2.3.

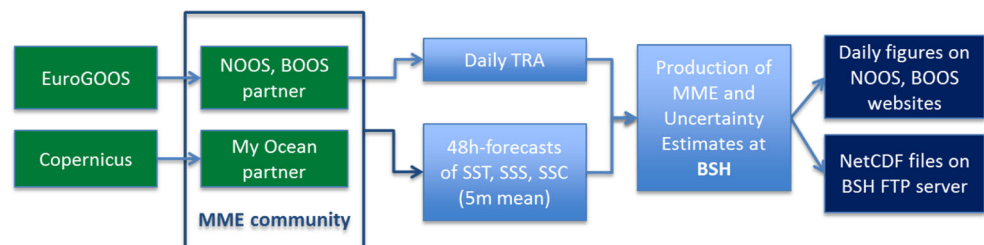
Model forecasts of the parameters are supplied daily on the local ftp servers of each partner. The data sets provided by the individual partners for the MME are listed in Table 2. For each region, all forecasts are interpolated on common reference grids, which are defined by the nominal MyOcean products.

The domain of the MME for the North Sea therefore ranges from 4.11° W/48.60° N to 13.00° E/60.33° N. For the Baltic Sea, the MME products cover the area from 9.04° E/53.03° N to 30.29° E/65.88° N. At present, the resolution is 1/9° latitude and 1/15° longitude for the North Sea and 1/20° latitude × 1/12° longitude for the Baltic Sea. Slightly different further processing procedures are applied for each parameter, described in detail in the following subchapters. The hourly MME outputs are produced each day for the 48-h forecast period using all data available on that day. Erroneous data, i.e., with shorter forecast length or obvious errors in model results such as incorrectly interpolated SSC, are excluded from the MME. The daily output of each MME includes figures, published on the NOOS and BOOS websites, and NetCDF files of SST, SSS, and SSC MME, which are freely available on the BSH ftp server.

2.2.1 Sea surface temperature and salinity

Currently, there are up to eight different forecasts of SST and SSS available for the North Sea and ten for the Baltic Sea. The most covered region in the North Sea is the central part of the

Fig. 1 Schematic illustration showing the general structure of the participating community and the production and outputs of the MME



domain. For the Baltic Sea, the maximal number of different forecasts can be assembled in the Gulf of Finland. In most areas of the Baltic Sea, there are up to nine forecasts used for the current version of the MME for SST and SSS. For the MME, only areas covered by more than three forecasts are taken into account, resulting in a smaller region of the MME compared to the area of the MyOcean product. The number of contributing models, MME maximum, MME minimum, MME median, and MME mean and standard deviation between the models of SST and SSS are calculated at each grid point for each time step of the 48-h forecast. These outputs are provided in the NetCDF-files. On the NOOS and BOOS website, the figures of the first forecast time are shown (i.e., 01:00 UTC).

2.2.2 Sea surface current

At present, up to seven forecasts of SSC are available for the North Sea and up to ten forecasts for the Baltic Sea. A first overview of the ensemble spread is given by progressive vector diagrams (PVD) (Emery and Thomson 2001) of the hourly surface currents calculated for each 48-h forecast at selected points distributed over the whole study area (see example of PVD in Fig. 4). Since the NOOS and BOOS transects are situated in hydrodynamically important areas, i.e., English Channel, Kattegat, or the Danish Straits, PVDs are calculated at the centers of all transect (see Fig. 7 for transect locations and numbering). The PVD is a type of water particle trajectory calculated by summing up the travelled distance of the particle using the hourly u and v velocities of the surface currents. In addition, the ensemble mean and the standard deviation of the velocity components are calculated on an hourly basis, and the resulting mean PVD (MME PVD) is determined.

More recently, a MME and corresponding statistics of the 2D SSC fields are produced on an hourly basis for the 48-h forecast period. On the NOOS and BOOS websites, only figures for the first 24 h are displayed. The MME and statistical values are calculated as follows (where $i=1,2,\dots,n$ for number of forecast):

1. The mean current field of each velocity component and the resulting magnitude, the vector mean current (\overline{VM}), is determined with

$$\overline{VM} = \sqrt{\overline{u}^2 + \overline{v}^2}, \quad \text{with} \quad \overline{u} = \frac{1}{n} \sum_{i=1}^n u_i \quad \overline{v} = \frac{1}{n} \sum_{i=1}^n v_i$$

It should be noted that this definition may average out current components of opposite directions, which means that even though the models predict strong current of varying directions, the average \overline{VM} may be small.

2. The standard deviation ($S_{\overline{VM}}$), which represents the dispersion between the models, is given by

$$S_{\overline{VM}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (VM_i - \overline{VM})^2} \quad \text{with} \quad VM_i = \sqrt{u_i^2 + v_i^2}$$

3. The stability (P) between the forecasts, expressed by the ratio of the vector mean current \overline{VM} to the mean magnitude (\overline{MM}), is calculated with

$$P = \frac{\overline{VM}}{\overline{MM}} * 100, \quad \text{with} \quad \overline{MM} = \frac{1}{n} \sum_{i=1}^n M_i \quad M_i = \sqrt{u_i^2 + v_i^2}$$

Areas characterized by, i.e., low stability indicate that either magnitude or directions of the forecasts are not consistent.

4. The angular difference, which is the difference between the current fields of the MME mean and the MyOcean (MyO) product, is displayed as angular degree (α) and given by

$$\cos\alpha = \frac{(\overline{u} * u_{MyO}) + (\overline{v} * v_{MyO})}{\overline{VM} * VM_{MyO}}, \quad \text{with} \quad VM_{MyO} = \sqrt{u_{MyO}^2 + v_{MyO}^2}$$

5. The difference-to-standard-deviation ratio (DSR), calculated by dividing the difference between the MME mean (\overline{VM}) and the MyOcean product (VM_{MyO}) by the standard deviation of the MME, is expressed as

$$DSR = \frac{|\overline{VM} - VM_{MyO}|}{S_{\overline{VM}}}$$

The ratio shows where the difference is smaller than the standard deviation, i.e., if below 1.

2.2.3 Water transport

The MME of water transport is based on an ongoing project in the NOOS and BOOS communities, which has been running since 2004, focusing on the exchange of computed transport to get a better understanding of the hydrodynamic situation in the North Sea and Baltic Sea. In the project, heat transport, salt transport, and water transport across several transects in the North and Baltic Sea are calculated on a daily basis using the outputs from different circulation models. The main tidal contribution is removed by averaging the transport at each grid cell along the transect over a time interval of 24 h and 50 min centered around noon of the first day of each forecast. The

resulting positive and negative transport values along a transect are summarized yielding the total inflow and outflow. The net transport is given by summing up inflow and outflow. The transport data of all contributing models are displayed in charts and vertical profiles on the NOOS and BOOS websites (www.noos.cc/index.php?id=151, www.boos.org/index.php?id=24, accessed 24 October 2014).

Based on data from this ongoing project, a MME of vertically integrated and surface water transport is developed to provide information about model uncertainty. Daily data across the defined transects are provided by up to six models for NOOS transects and by up to four models for BOOS transects (see Fig. 7 for transect locations and numbering). The ensemble mean and standard deviation of the model data are calculated and displayed on daily maps. An additional statistical parameter, the coefficient of variation (CV), helps to compare the dispersion between the data (i.e., Brown (1998)). The CV is the ratio of the standard deviation (T_{std}) to the absolute ensemble mean of transports (T_{mean}):

$$CV = \frac{T_{std}}{|T_{mean}|}, \text{ with } T_{mean} = \frac{1}{n} \sum_{i=1}^n T_i, \text{ and}$$

$$T_{std} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (T_i - T_{mean})^2}$$

A low CV index means low variability between the models. If the standard deviation is larger than the mean transport, the CV index is higher than 1. For this study, the CV index is subdivided into three categories: category 1 ($CV \leq 1$), category 2 ($1 < CV \leq 3$), and category 3 ($CV > 3$), where a CV above 3 is often associated with high variability or even outliers (Brown 1998).

2.3 Spatio-temporal statistics

For the statistical evaluations, only complete data sets were included, thus only those days and regions where all model data are available. The amount of complete data sets varies with region and parameter and is also due to the late inclusion of some forecasts. Accordingly, the study period varies between the parameters: For SST and SSS, the period is 01.01.2014–31.12.2014, SSC are evaluated for the time period 01.05.2014–31.05.2015, and TRA is studied for the period 01.04.2013–31.05.2015.

2.3.1 Comparison of sea surface temperature forecasts to satellite observations

Sea surface temperature of the MME mean, the MME median (MME products), and the individual forecasts are

compared to remote sensing (satellite) data. It should be mentioned that satellite SST measures skin temperature, while the SST used for the MME is the 5-m mean of the upper model layers. Due to the diverse performance of satellite observations, several products are selected for the comparison: For the North Sea, the daily level 3 MyOcean SST nighttime satellite data is used, which is from the mono sensor AVHRR. For the Baltic Sea, the comparison is carried out by using the daily level 3 MyOcean SST nighttime satellite product, which is provided by various sensors: AATSR, AVHRR, AVHRR_GAC, SEVIRI, GOES_Imager, MODIS, and TMI. It has to be noted that satellite data is affected by cloud cover. In comparison to the Baltic Sea, less satellite data are used for the North Sea, where the satellite products are from mono sensor.

Due to the limitation of the spatial coverage of SST satellite data in the North Sea and the Baltic Sea, the comparison is carried out on a monthly basis. The SST 01-h forecast is selected for comparison, since it is closest to the nighttime satellite data. Satellite data at 0 h UTC are interpolated to the reference grids of the MME products. The bias between the individual SST 01-h forecast and the satellite data (forecast–satellite data) is averaged over each month at each grid point. In addition the root-mean-square deviation (RMSD) of each forecast is calculated for each month at each grid point. Moreover, the number of days with available satellite data is divided by the length of the month giving the available satellite data (%) for each grid cell. It has to be noted that only grid points are taken into account, where the satellite data are available for more than 7 days per month. The monthly mean values for bias, RMSD, and available satellite data are further spatially averaged. Annual means of bias and RMSD are compared respectively. The comparison is done for the time period January–December 2014 using the MATLAB package CalVal-toolbox (Lagemaa et al. 2013; Jandt et al. 2014). Results are presented in Sect. 4.1.

2.3.2 Seasonal changes of sea surface salinity

For SSS, the differences among the individual forecasts are evaluated for the time period January–December 2014. Therefore, the temporal mean of the MME mean and the standard deviation between the forecasts is calculated at each grid point. In addition, the daily spatial mean for each region is calculated for each forecast and the MME products. The ensemble spread, expressed as the ensemble standard deviation, is taken into account for the comparison. Results are presented in Sect. 4.2.

2.3.3 Regional pattern in forecast deviation for sea surface current

The daily PVDs of the North Sea and Baltic Sea are evaluated by determining the final displacements between the MME PVD and the PVD of each forecast separately. The result is a matrix for each forecast showing distances in kilometers for each day at the points covered by the model grids. Moreover, the temporal mean of final displacement is calculated for every forecast at the corresponding transects. Another way to display the deviation between the forecasts is to determine the temporal mean of standard deviations of SSC magnitude (c). The mean standard deviation between the forecasts over the 48-h time period (msd_f) is normalized by the mean of forecast standard deviations (msd_{Si}) to get comparable relative values independent of the transect location. The temporal mean of the resulting daily deviations (SD) was calculated at each transect T :

$$SD(T) = \frac{1}{j} \sum_{l=1}^j \frac{msd_f(l)}{msd_{Si}(l)}, \text{ with}$$

$$msd_f(l) = \frac{1}{k} \sum_{t=1}^k S_n(t), \text{ and } S_n(t) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (c_{i,t} - \bar{c}_t)^2}$$

$$msd_{Si}(l) = \frac{1}{n} \sum_{i=1}^n S_k(i), \text{ and } S_k(i) = \sqrt{\frac{1}{k-1} \sum_{t=1}^k (c_{i,t} - \bar{c}_{i,t})^2}$$

where $l=1,2,\dots,j$ for number of days, $i=1,2,\dots,n$ for number of forecast, and $t=1,2,\dots,k$ for each hourly output of the 48-h forecast. Results are presented in Sect. 4.3

2.3.4 Regional pattern in forecast deviation for water transport

Statistical analyses are only performed for surface water transport for a better comparison to the surface fields of the other parameters. To estimate regional differences in model consistency, the occurrence of every CV category in percent (see Sect. 2.2.3) at each transect is determined. Since not all models included in the MME provide transport data for all transects, the number of products and thus the resulting percentages of complete data sets differ accordingly between transects. To detect differences in daily transport patterns, the correlations between each time series were determined and the mean of all correlations was calculated. This was done for each transect separately. The results were compared to the mean of the correlations between the MME time series and each product time series, also computed for each transect. To determine which product deviates most from the others, the RMSD between the time series of each product and the MME median is normalized by the standard deviation of the MME median at each transect. Normalization is done to have relative, comparable results similar to the SSC analysis. This measure allows comparison of regions with different transport values. Results are presented in Sect. 4.4.

3 Daily results of the MME and ensemble statistics

3.1 Sea surface temperature and sea surface salinity

Examples of graphical daily output of the MME for SST in the North Sea and for SSS in the Baltic Sea are shown in Figs. 2 and 3, respectively, reflecting obvious differences among the forecasts. The number of ensemble members displays the actual number of forecasts used by the MME system on the current day. The ensemble minimum and maximum of the forecasts indicate the plausible range of simulated SST and SSS. For instance, in Fig. 2, the differences of SST among the forecasts are approximately up to 3 °C in the English Channel. The standard deviation displays the variability among the forecasts. In the Skagerrak and Kattegat, high standard deviation is the dominant characteristic in the SSS field indicating large differences among the forecasts in these areas (Fig. 3). Moreover, the ensemble median is calculated as additional information in order to provide a more robust estimate of the ensemble mean less prone to outliers.

For example, the ensemble mean of SST in the northern North Sea close to the British coast is slightly higher than the ensemble median (Fig. 2). In this case, SST of one forecast might be much higher compared to the other forecasts on the chosen day. This is also reflected by the wide range between ensemble minimum and ensemble maximum where the differences between the individual forecasts are shown. Along the boundaries, where the number of ensemble members changes, discontinuous transitions can often be found in all fields. This characteristic is obvious approximately along 59° N in the North Sea, where the analysis number drops from 6 to 5 and further to 4 northward. This form of discontinuity cannot be found in the Baltic Sea, since most of models in this region cover the same area.

3.2 Sea surface current

The PVD (see Sect. 2.2.2) and the 48-h time series for the u and v components as well as a feather plot are displayed on daily figures for each transect separately. An example of transect 7 (Tr7) in the North Sea is shown in Fig. 4. As the tides are present in the surface currents, the time series at the North Sea transects are dominated by a strong tidal signal which is also visible in the resulting PVD. Surface currents in the Baltic Sea also have a tidal signal which is much weaker, and the strength of currents is in general lower than in the North Sea. However, comparatively strong currents occur also in the Danish Straits. In this example, BSH_HBM seems to be out of phase and overestimates the magnitude of u velocity while it underestimates the magnitude of v velocity, the latter similar to DMI_DKSS. This is reflected in the PVD, where those forecasts exhibit the largest distances from the starting point. Although it is not obvious in the time series, the PVD of

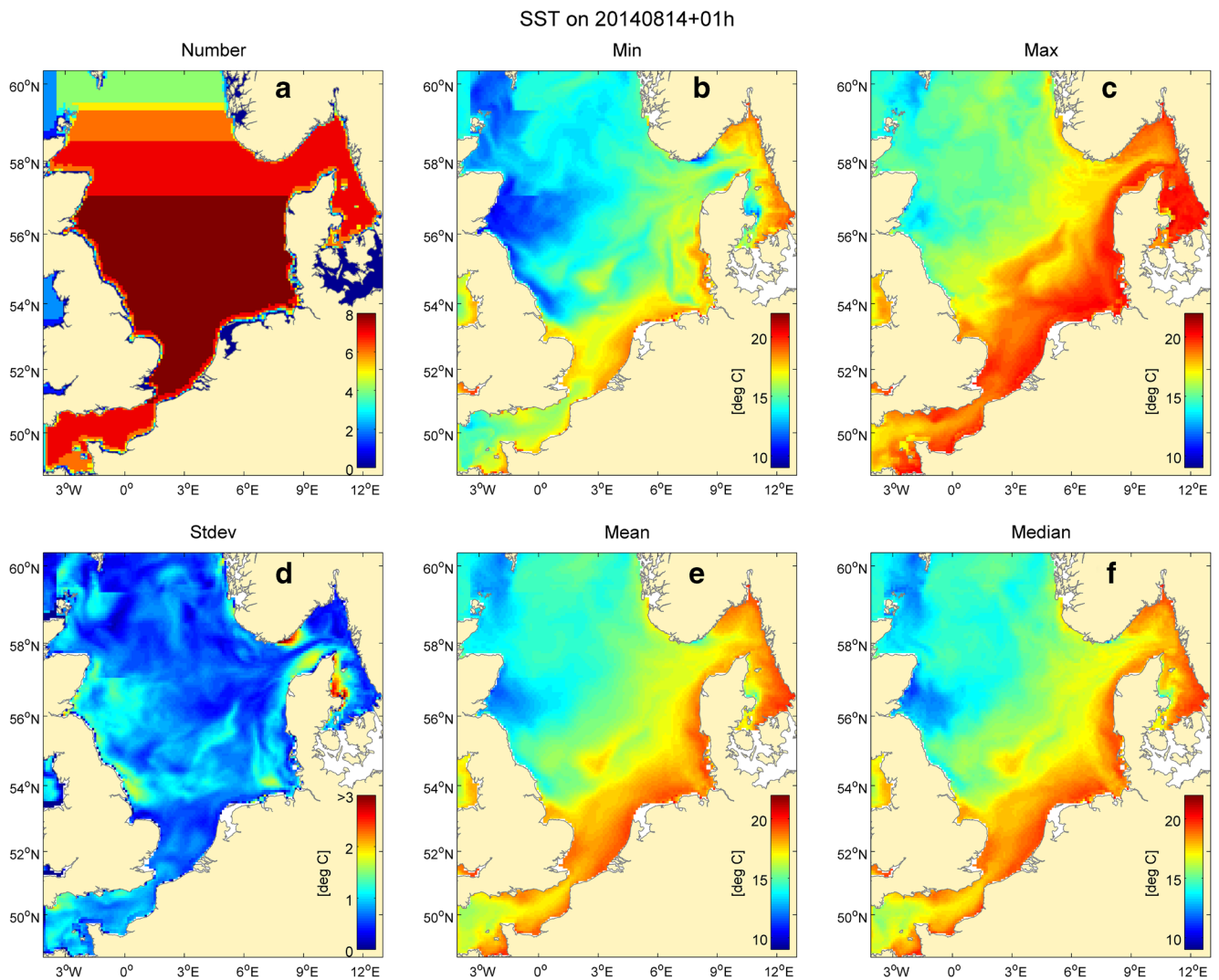


Fig. 2 Example showing the number of forecasts per grid cell (a), ensemble minimum (b) and ensemble maximum (c), standard deviation (d), MME mean (e), and MME median (f) of SST 01-h forecast in the North Sea

FCOO_GETM has similar large distances. However, the high uncertainty in the u velocity between 01 and 06 h of the forecast is not obvious in the PVDs. Differences in SSC possibly occur due to the different boundary conditions of the models with varying tidal constituents and resolutions. The large difference between BSH_HBM and BSH_CM0D might also be due to different turbulence schemes in the models which possibly have an effect on the surface currents. In Fig. 5, some examples of PVDs with various structures on different days are shown.

Depending on the variations in time series, the resulting PVDs exhibit smaller or larger differences also depending on the region (Fig. 5). In the upper Baltic Sea, i.e., Bothnian Sea (Tr53) or Gulf of Finland (Tr44), the tidal signal is quite low and SSC are mainly dominated by the wind or inertial currents. Tr13 displays a pattern in a tidally dominated region. Variations in PVD patterns occur due to differences in phase, current direction, and strength. As mentioned above,

differences in boundary conditions and especially tidal constituents of the models might cause the differences in current patterns. Although mostly current forecasts in the upper 5-m mean are used for the MME, the different layer thicknesses of the original models might still have an impact on the strength and direction of the currents.

The hourly figures showing the MME of SSC, accompanied by some statistics, are only created for the first 24 h of the whole forecast period. The example shown in Fig. 6 displays a forecast close to a storm event over the North Sea. Mean and standard deviation are highest close to the eastern coast of Great Britain and in the English Channel. This pattern is also reflected in the difference-to-standard-deviation ratio. The stability of the MME is very low in the regions where the current strength is also low (blue areas in MME mean), and high angular differences between the MME mean and the MyOcean product occur. For this storm event, the values in high angular difference are related to weak currents

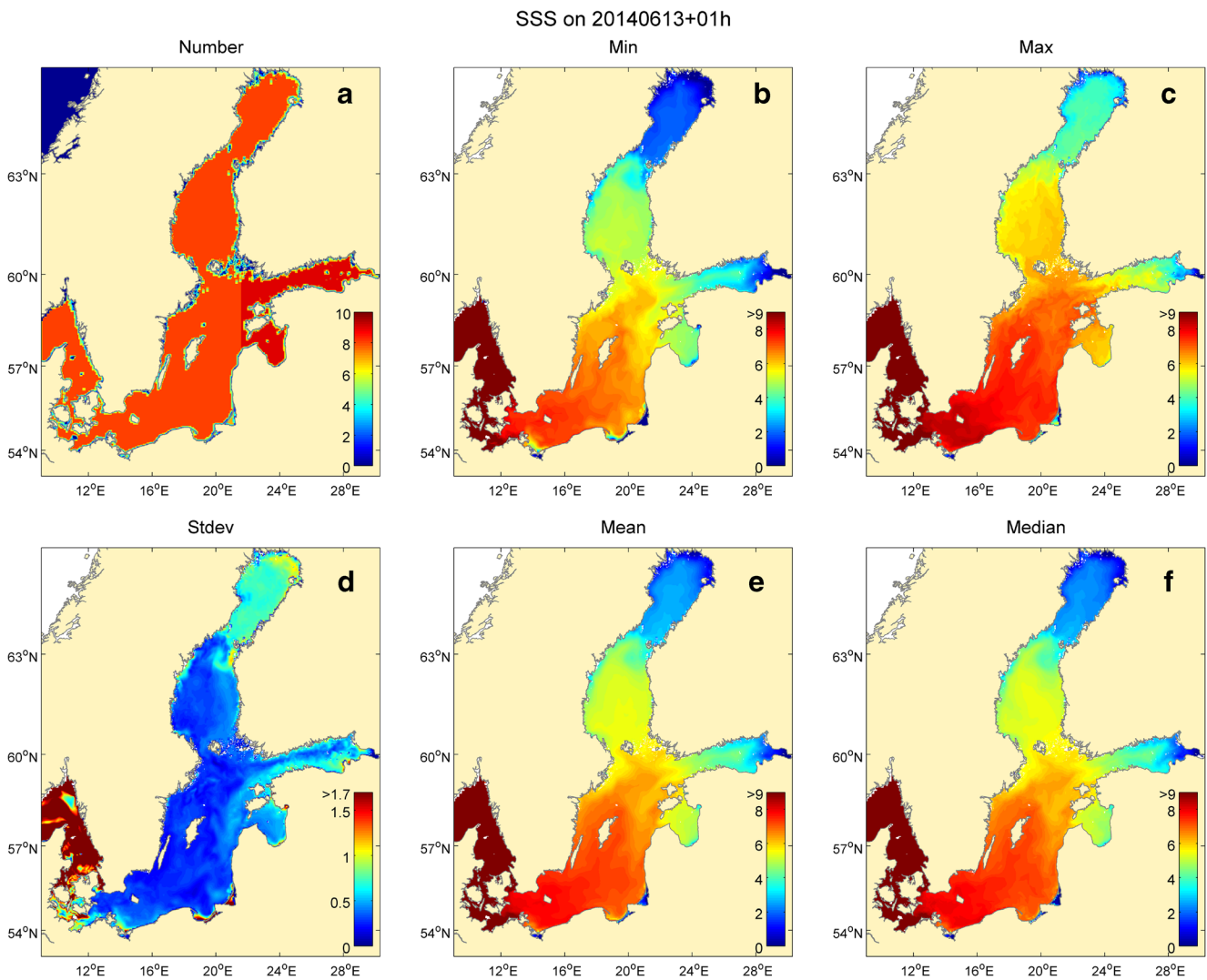


Fig. 3 Example showing the number of forecasts per grid cell (a), ensemble minimum (b) and ensemble maximum (c), standard deviation (d), MME mean (e), and MME median (f) of SSS 01-h forecast in the Baltic Sea

and differences in tidal phase in the models. Thus, high values of angles might give the wrong impression of very strong varying forecasts.

This relationship between the patterns of the statistics during the storm event is displayed on most days in the North Sea and the Baltic Sea, also when the currents are generally lower. High SSC occur in the Skagerrak and Kattegat and are often connected to high standard deviation between the forecasts in this region. Due to the higher resolution of the forecasts in the Baltic Sea, patterns are more detailed than in the North Sea. However, to make a qualitative comparison between the forecasts for the narrow Transition Area, even higher resolution of model grids is needed for this region.

3.3 Water transport

Owing to differing grid extensions and orientations, not every model delivers transport data at each transect. However, the

major central part of the North Sea is covered by six models. Daily maps are produced separately for the North Sea and the Baltic Sea (Fig. 7). Since water transports in the Baltic Sea are typically smaller than in the North Sea, it is important to note that the scale factor of the arrows is different in both maps and thus cannot be directly compared. If the ensemble mean is close to 0, the CV index (see Sect. 2.2.3) is in a critical area and should be handled carefully. A CV index greater than 3 appears mostly when the ensemble mean is close to 0 either due to very low transports or due to opposed transport direction. Temporarily, there occur major product disagreements in some regions which are further evaluated in Sect. 4.4. But, mostly, there seems to be a stable agreement between the products; seasonal differences like higher transports in winter and lower transports in summer are reflected by most of the products. The mean circulation patterns in North Sea and Baltic Sea are generally well represented by the water transports, i.e., major transport through the English Channel,

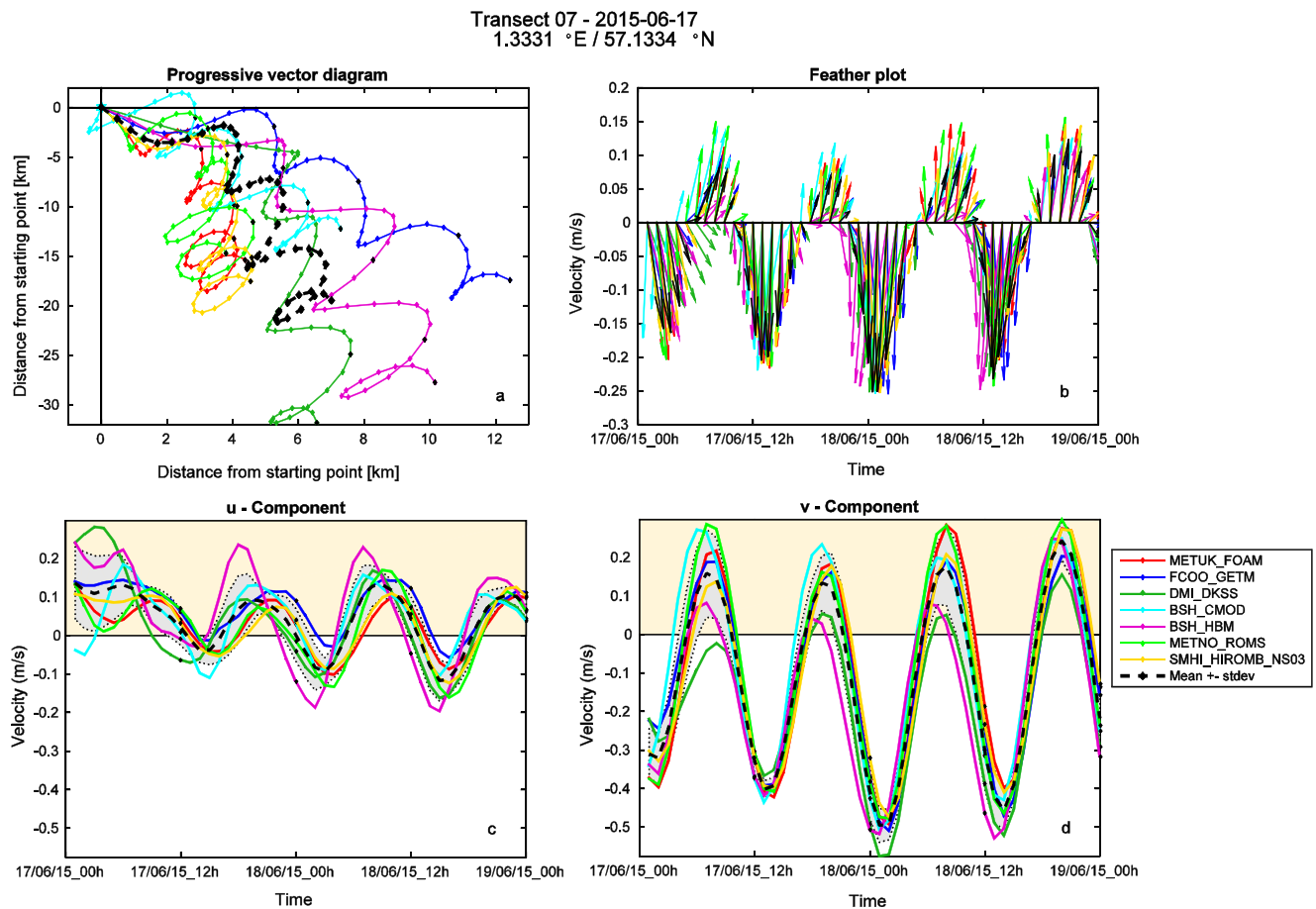


Fig. 4 Example of daily output showing time series of u and v components of near-surface currents (**c**, **d**), resulting PVD (**a**), and feather plot (**b**) at Tr7 (North Sea: see Fig. 7 for transect locations). Mean time series and corresponding MME PVD are marked by *dashed black lines*

inflow from the Atlantic Ocean along western boundary of the North Sea (Tr1, Tr4, Tr7, Tr10), and outflow including the Baltic Sea along the Norwegian Trench (Tr2, Tr5, Tr6, Tr8) or following the main circulation in the central Baltic Sea (Feistel et al. 2008).

4 Results of spatio-temporal statistics

4.1 Comparison of sea surface temperature forecasts to satellite observations

4.1.1 North Sea

The comparison of SST forecasts to satellite observations (see Sect. 2.1) in the North Sea is displayed in Figs. 8, 9, and 10. In February 2014, there is no satellite data available for more than 7 days which means that no results can be obtained for comparison. For the other months, the mean biases show that all forecasts and the MME mean and MME median (MME products) tend to produce negative values (Fig. 8). Only the forecasts from SMHI_HIROMB_NS03 and METNO_ROMS

have positive biases in more than 2 months compared to the other ensemble members. A possible reason for the negative biases could be that different kinds of surface temperatures are compared: Satellites measure skin temperature, while SST provided by the forecasts is a 5-m mean of the upper model layers (see Sect. 2.1). Negative biases are also found in the comparison between satellite and in situ data as demonstrated by Alvera-Azcárate et al. (2011). Moreover, the biases from most forecasts change differently with time. For instance, the bias from BSH_CMOD varies between approximately 0 and -1 °C, while METUK_FOAM has the smallest absolute bias varying only slightly between -0.2 and 0.1 °C. As mentioned in Sect. 2.1, data assimilation is applied in FOAM_AMM. The result reflects the improvement of the forecast due to data assimilation. Although other models such as ROMS and HIROMB from SMHI also apply data assimilation, their monthly mean biases are not close to zero. This might be due to other satellite products and data assimilation techniques applied.

The biases of the MME products are negative during the whole study period and do not change significantly with time, only varying between -0.5 to -0.1 °C (Fig. 8). It indicates that

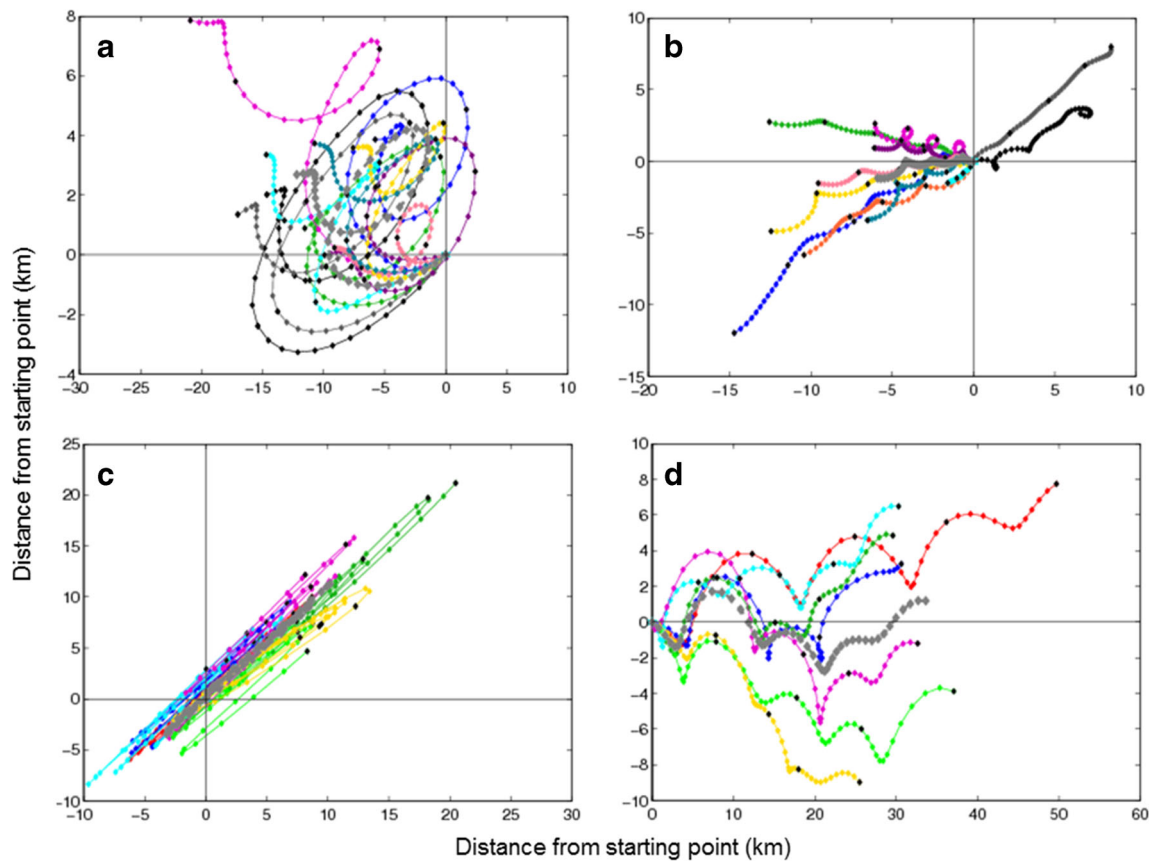


Fig. 5 Example PVDs of the Baltic Sea (a, b) and the North Sea (c, d) with varying structures: Tr53 (a, 01.08.14), Tr44 (b, 18.07.14), Tr13 (c, 07.07.14), Tr8 (d, 15.07.14) (see Fig. 7 for transect locations)

the SST is underestimated on average by the MME products. Even though the biases of the MME products are larger than the bias of METUK_FOAM, they show less variation in comparison to the other models which do not apply data assimilation. Looking at the annual mean bias, METNO_ROMS has the smallest value of about $-0.03\text{ }^{\circ}\text{C}$ which is an average of large positive biases in summer and negative biases in winter (Fig. 8). The annual mean bias from METUK_FOAM is also low accounting for $-0.04\text{ }^{\circ}\text{C}$, while the values for all other models vary between -0.2 and $-0.5\text{ }^{\circ}\text{C}$. The biases from the MME products are slightly higher with -0.27 and $-0.23\text{ }^{\circ}\text{C}$, respectively, but still smaller than the biases from most individual forecasts.

The monthly mean RMSDs between forecasts and observations display the errors of the MME products and the individual forecasts (Fig. 9). Except for October, the MME products are more accurate than any of the individual forecasts, reflected by lower errors. Although the RMSD of METUK_FOAM is smaller than the errors of the other models, it is still slightly higher than those of the MME products. In October, only the RMSD of METUK_FOAM is lower than the RMSDs of the MME products.

Seasonal features are reflected by highest values for RMSD of all products between May and August with maximum

values around $0.62\text{ }^{\circ}\text{C}$ (Fig. 9). The RMSDs of the MME products in these months are lower, accounting for $0.54\text{ }^{\circ}\text{C}$ and decreasing to about $0.42\text{ }^{\circ}\text{C}$ in August. From January to April and in September, November, and December, the maximum values of RMSDs for the individual forecasts vary between 0.28 and $0.45\text{ }^{\circ}\text{C}$. During these months, the RMSDs of the ensemble members are close to each other. In this time, also the RMSDs of the MME products deviate only by approximately $0.05\text{ }^{\circ}\text{C}$ from the maximum values of the individual forecasts. Looking at the annual mean RMSDs, there are only marginal differences among the forecasts where the values vary between 0.42 and $0.46\text{ }^{\circ}\text{C}$. The annual mean RMSD from METUK_FOAM is lower than those from the other forecasts. Nevertheless, the values from the MME products are lowest with about $0.40\text{ }^{\circ}\text{C}$.

There is no direct relation distinguishable between the monthly mean available satellite data and the monthly mean bias or RMSD. Looking at the spatial distribution in May, the available satellite observations vary between 50 and 65% in the central North Sea and the English Channel, accounting for 41% on average for this month. In June 2014, the availability of satellite data reaches values of about 90% only in the Norwegian Channel. Due to the low coverage in most parts of the North Sea (less than 40%), the spatial averaged

SSC on 20141021 14:00

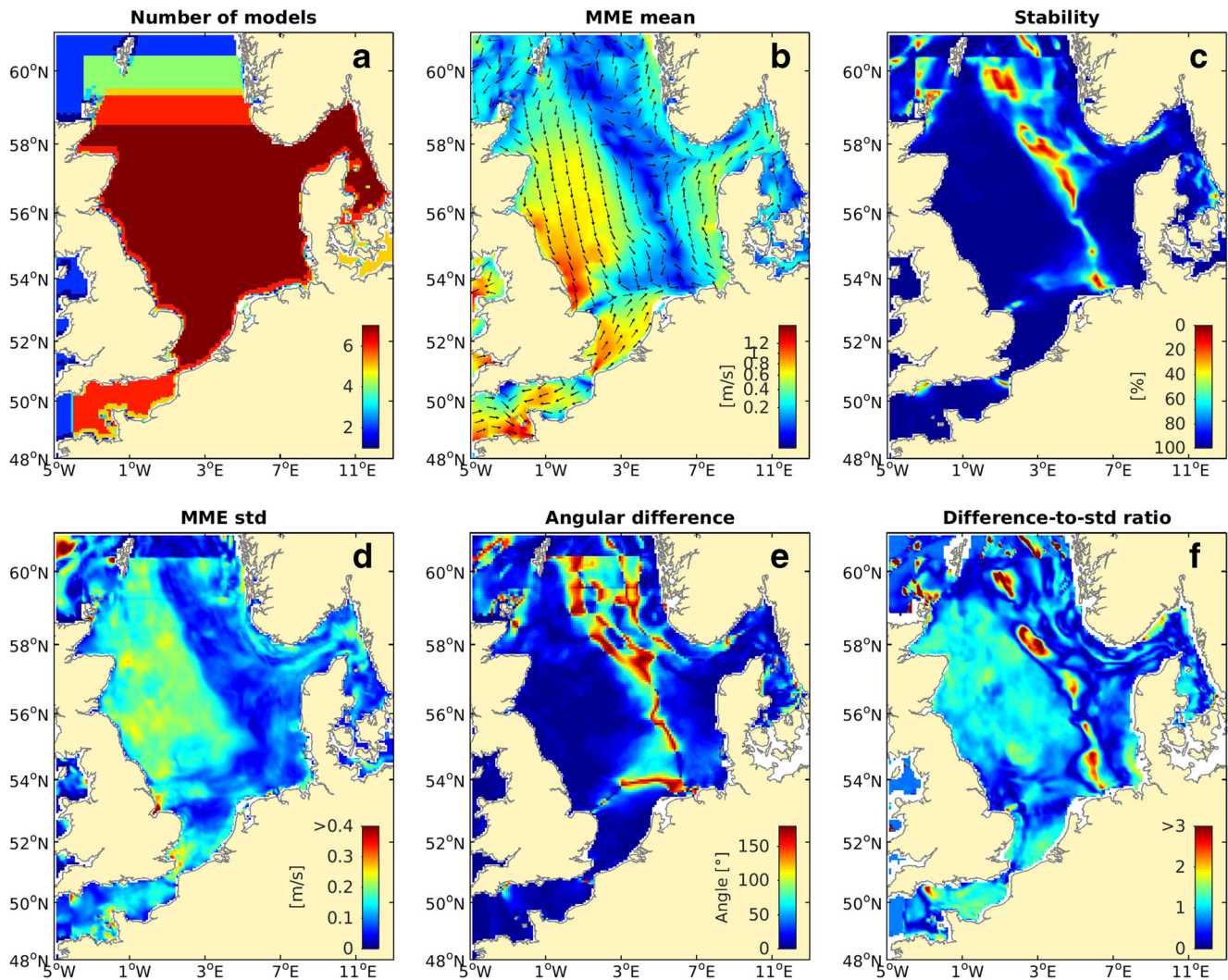


Fig. 6 Example of SSC showing the number of forecasts per grid cell (a), the MME mean (b), stability (c), standard deviation between the forecasts (d), the angular difference between the MME mean and the MyOcean product, (e) and the difference-to-standard-deviation ratio (f) for the North Sea

availability of satellite data is still low, reaching only 28 % in June. During these 2 months, high values and large spread of RMSDs among the forecasts are shown, although the availability of satellite observations is different (Figs. 8 and 9). There might exist local dependency between available satellite data and the errors, but it is not reflected in the comparison of the monthly mean values.

The RMSDs of the ensemble members have the largest differences in August (Fig. 9). Hence, the spatial distribution of RMSDs of the MME mean and the individual forecasts is evaluated for this month (Fig. 10). High RMSDs are found in the southeast of the North Sea for METNO_ROMS, while high values of DMI_DKSS occur in the northwestern part of the North Sea. Moreover, high RMSDs in the area close to the eastern British coast occur only for BSH_CMOD, SMHI_HIROMB_NS03, and DMI_DKSS. The RMSD of the MME mean is lowest compared to the individual forecasts

and distributed evenly with errors less than 0.5 °C at most grid points. No distinct area with high RMSD is reflected for the MME mean (Fig. 10). It shows that the ensemble process helps to reduce the error. In order to fully understand the mechanisms influencing the forecast uncertainties including seasonal features, the atmospheric forcing of each forecast needs to be taken into account, which is not part of this study.

The large spread of spatially averaged monthly mean RMSDs in June, July, and October is due to large differences in the spatial distribution of RMSDs for each product. For instance, in June, almost all forecasts except for DMI_DKSS have large errors in the central North Sea. The RMSDs of SMHI_HIROMB_NS03 and METNO_ROMS are close to 2 °C in this area, while those of the other forecasts vary between 1.25 and 1.75 °C. In July, BSH_CMOD, BSH_HBM, and DMI_DKSS have large errors in the area close to the eastern British coast, which is not reflected by the other

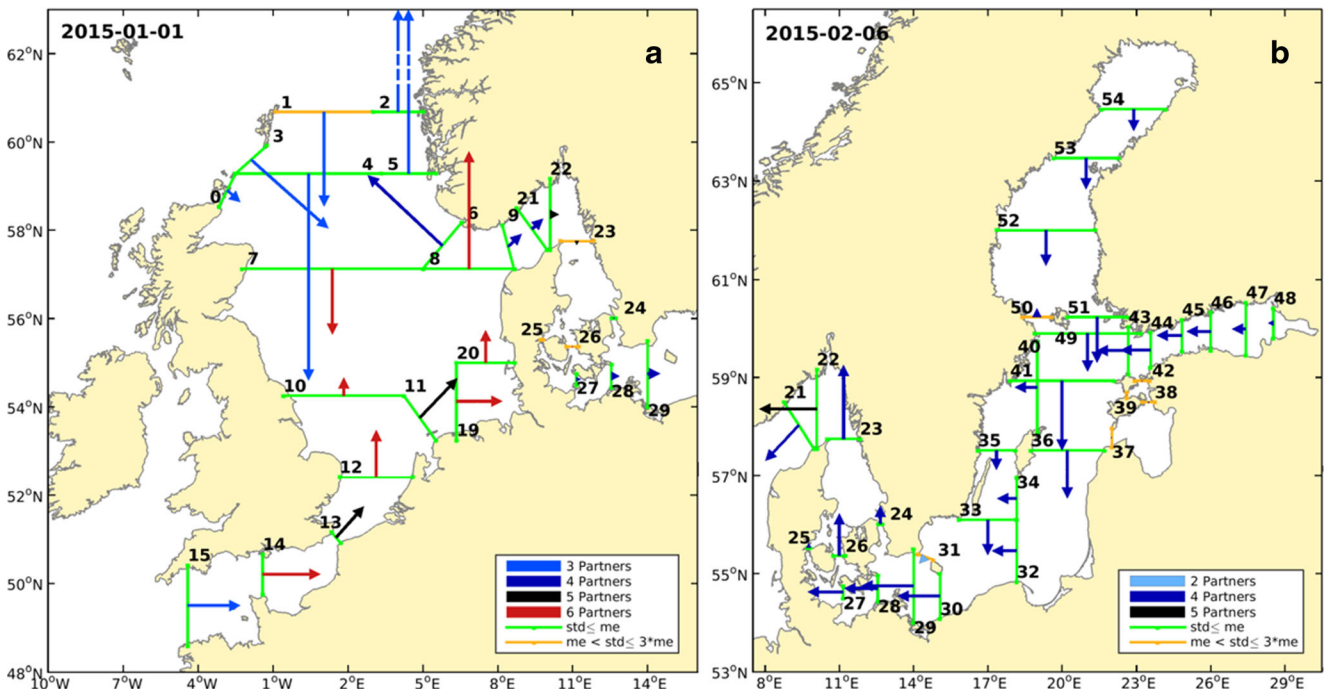


Fig. 7 Example maps of the North Sea (left) and the Baltic Sea (right) showing the ensemble mean of daily vertically integrated transport across the transects defined by NOOS and BOOS. The *arrows* indicate the magnitude of mean transport across each transect whereas the *color* of

the *arrows* marks the number of forecasts contributing to the mean. Each transect is colored according to the corresponding CV with *green* for $CV \leq 1$, *yellow* for $1 < CV \leq 3$, and *purple* for $CV > 3$

ensemble members. In October, extremely large errors occur only for BSH_HBM and SMHI_HIROMB_NS03 along the eastern British coast. For the remaining months (January–April, September, November, and December), the RMSDs are relatively low with an even distribution in the whole area. No region with large RMSDs can be distinguished.

4.1.2 Baltic Sea

The same comparison of SST forecast with satellite observations is carried out for the Baltic Sea for 2014 and displayed in Figs. 11, 12, and 13. The monthly mean biases of the individual forecasts vary strongly with time ranging between -2.5

Fig. 8 Monthly mean (a) and annual mean (b) bias of SST from the MME mean, MME median, and the ensemble members in the North Sea in 2014. The percentage of available satellite data per month is marked as *dotted line*

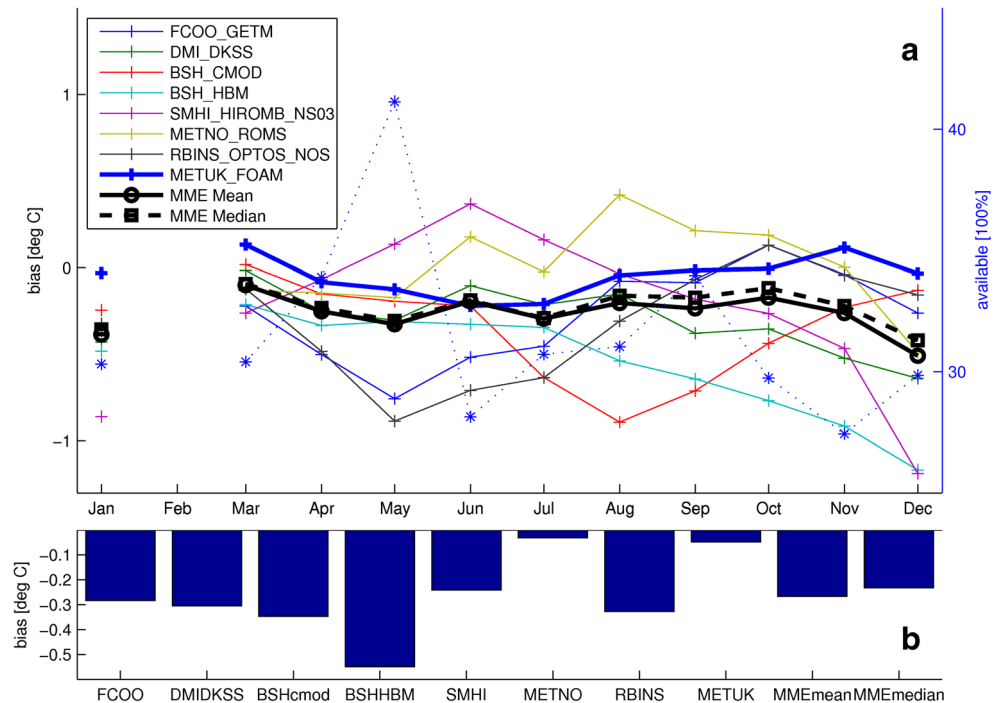
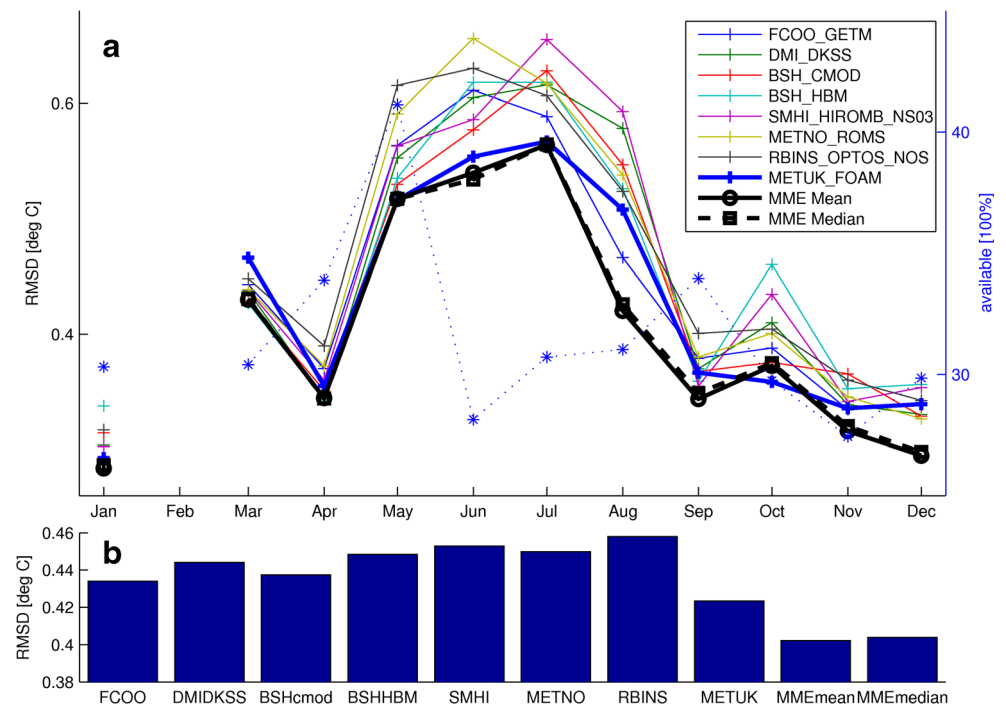


Fig. 9 Monthly mean (a) and annual mean (b) RMSD of SST from the MME mean, MME median, and the ensemble members in the North Sea in 2014. The percentage of available satellite data per month is marked as dotted line



and $0.8\text{ }^{\circ}\text{C}$ (Fig. 11). Some of the forecasts like FCOO_GETM and BSH_CM0D have positive biases in winter but negative biases in summer. The two forecasts from SMHI show an opposed pattern compared to the other forecasts. Data assimilation is applied in these models by using various observations, i.e., in situ and ferry box data and satellite data, which could be a reason that none of these forecasts is very close to the satellite observations used for this study.

Most forecasts have the highest negative biases in July, where the bias from FCOO_GETM even reaches $-2.5\text{ }^{\circ}\text{C}$ (Fig. 11). It indicates that the surface temperature is underestimated by most of the forecasts in July. The MyOcean product (DMI_HBM) has a negative bias in all months. Compared to the ensemble members, the biases of the MME products have less significant changes ranging from slightly above 0 to $-0.6\text{ }^{\circ}\text{C}$ with largest absolute values in July.

Differences in the annual mean biases of the MME products and the forecasts are quite distinct (Fig. 11). The only forecast with slightly positive bias is FMI_HBM_ec while FMI_HBM_hirlam exhibits a slightly negative bias. The biases of the remaining ensemble members vary between -0.2 and $-0.6\text{ }^{\circ}\text{C}$, where the values for the MME products are similar around approximately $-0.32\text{ }^{\circ}\text{C}$.

The RMSD of the MME products and the ensemble members vary strongly with time (Fig. 12). None of the products has the lowest error throughout the whole year. For instance, the lowest error in February and June is calculated for BSH_CM0D, while in May, the error of BSH_HBM is lowest and, in August, the RMSD of SMHI_HIROMB_NS03 has the lowest value. Except for February and between June and August, the MME mean has

the lowest errors with values less than $0.6\text{ }^{\circ}\text{C}$. In addition, a seasonal pattern can be distinguished in the monthly mean RMSDs. Between May and August, the errors of all forecasts are approximately two times higher than the values in the other months, accompanied by a large spread between the errors. The errors of the MME products are higher than the RMSD of some of the ensemble members in these months. This indicates that, if there are large uncertainties among the forecasts, the improvement gained through the ensemble process is decreased. To examine the physical reasons causing these seasonal features, more studies focusing on the atmospheric forcing of each forecast are necessary, which is not part of this study.

Although the MME products do not have the lowest errors throughout the whole year in the Baltic Sea, the MME mean still has the lowest annual mean RMSD of about $0.65\text{ }^{\circ}\text{C}$, which is slightly lower than the value from BSH_CM0D (Fig. 12). It shows that the ensemble process can improve the accuracy of the forecasts.

Figure 13 shows the spatial distribution of RMSD from each ensemble member and the MME products in the Baltic Sea in July 2014. The distribution of regions with high RMSD is different between the individual forecasts and the MME products, whereby the errors seem to increase from the southern part of the Baltic Sea to the North in all plots. This feature is most obvious in the plots showing FCOO_GETM, DMI_DKSS, FMI_HBM_ec, and FMI_HBM_hirlam. Some forecasts, such as BSH_HBM, DMI_HBM, DMI_DKSS, FMI_HBM_hirlam, and FMI_HBM_ec, have large errors along the southern boundary of the Baltic Sea. RMSDs are also high in the Gulf of Finland in all plots. A similar but slightly weaker pattern is

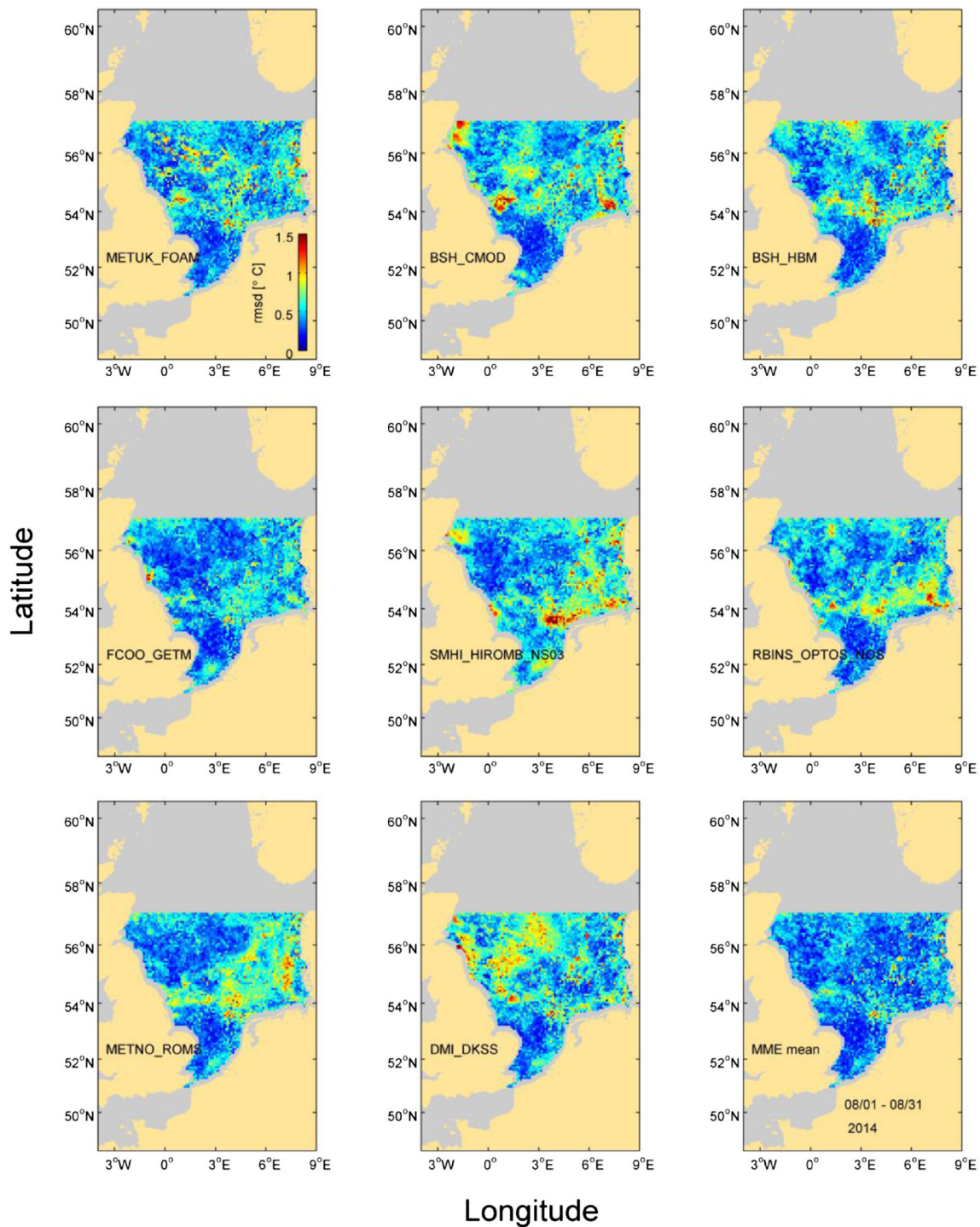


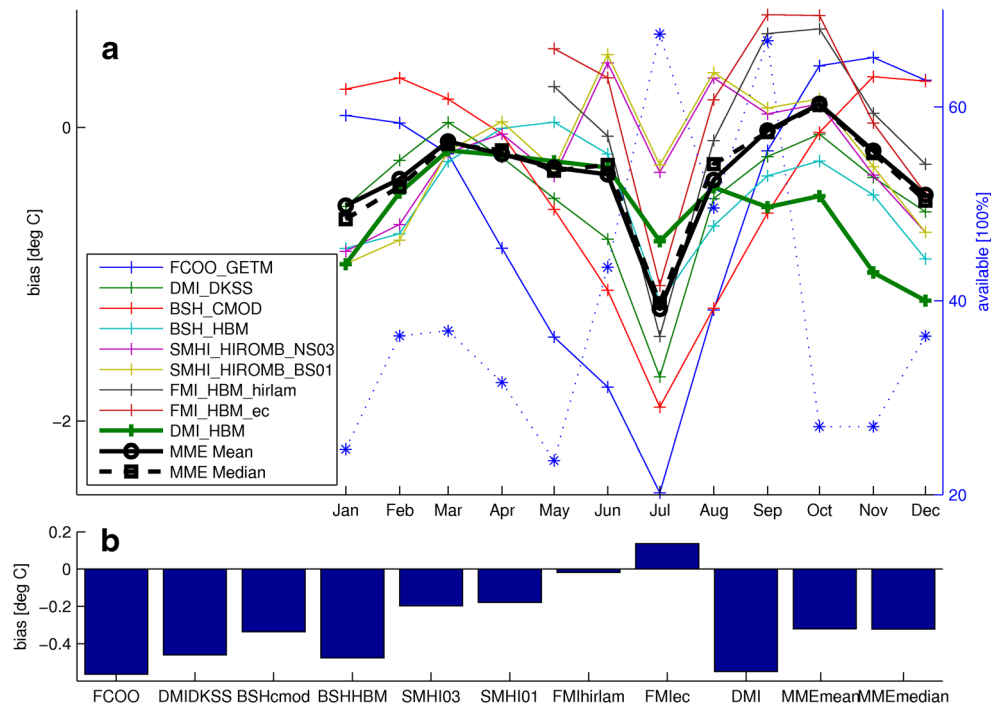
Fig. 10 Spatial distribution of the RMSD from the individual ensemble members and the MME mean (*bottom right*) in the North Sea in August 2014

reflected by the MME products, where the error of the MME mean is lower than the error of the MME median. For the MME mean, very high RMSDs are only shown at the entrance of the Gulf of Finland. In the center of the Baltic Sea, especially in its southern part, the error is mostly less than 0.6 °C (Fig. 13).

In winter, where RMSDs are low, the spatial distribution of errors from the ensemble members is more even with

less extreme values compared to the distribution in July. It has to be noted that in winter, the availability of satellite observations is usually low in the Baltic Sea, especially in the North since this area is often covered by sea ice. During this period, the RMSDs of the ensemble members are usually high which might be related to the low coverage of satellite data.

Fig. 11 Monthly mean (a) and annual mean (b) bias of SST from the MME mean, MME median, and the ensemble members in the Baltic Sea in 2014. The percentage of available satellite data per month is marked as dotted line



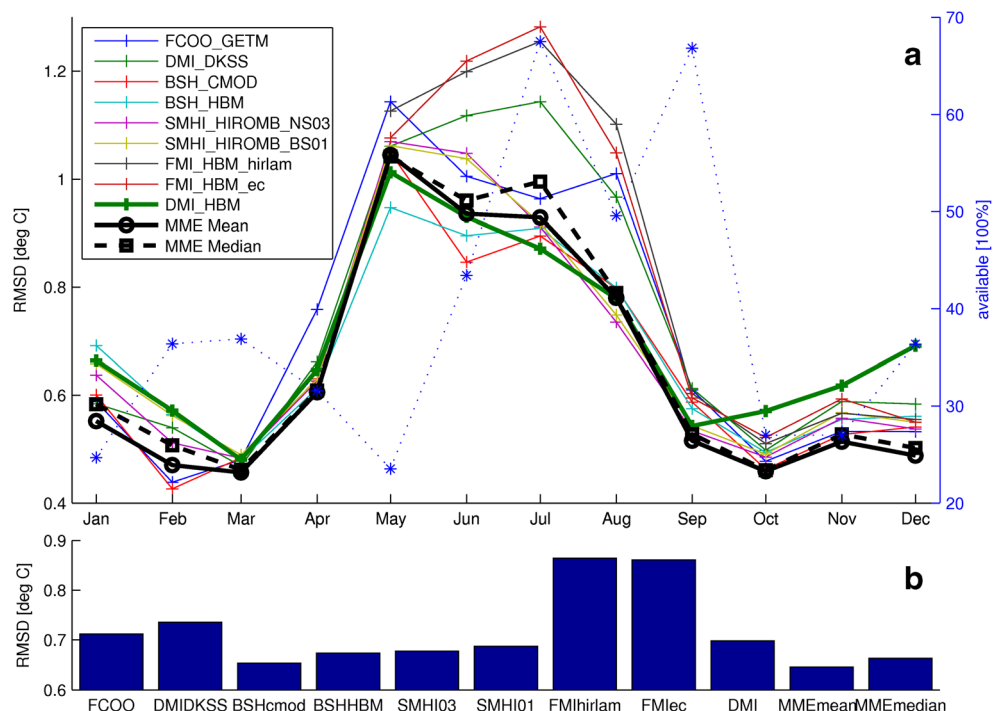
4.2 Seasonal changes of sea surface salinity

4.2.1 North Sea

The annual averages of the MME mean and the standard deviation for SSS are shown in Fig. 14. The extent of freshwater tongues from the river plumes in the southern North Sea is reflected by standard deviations of more than 5

whereas the values increase to 9 close to the river mouths. High standard deviations in these areas indicate that the freshwater inputs from the large rivers are simulated differently by the models. This further leads to the differences in the seaward extensions of the river plumes. These variances are probably caused by the differing methods or data sets that are used for river discharge in the models as highlighted in section 2.1.

Fig. 12 Monthly mean (a) and annual mean (b) RMSD of SST from the MME mean, MME median, and the ensemble members in the Baltic Sea in 2014. The percentage of available satellite data per month is marked as dotted line



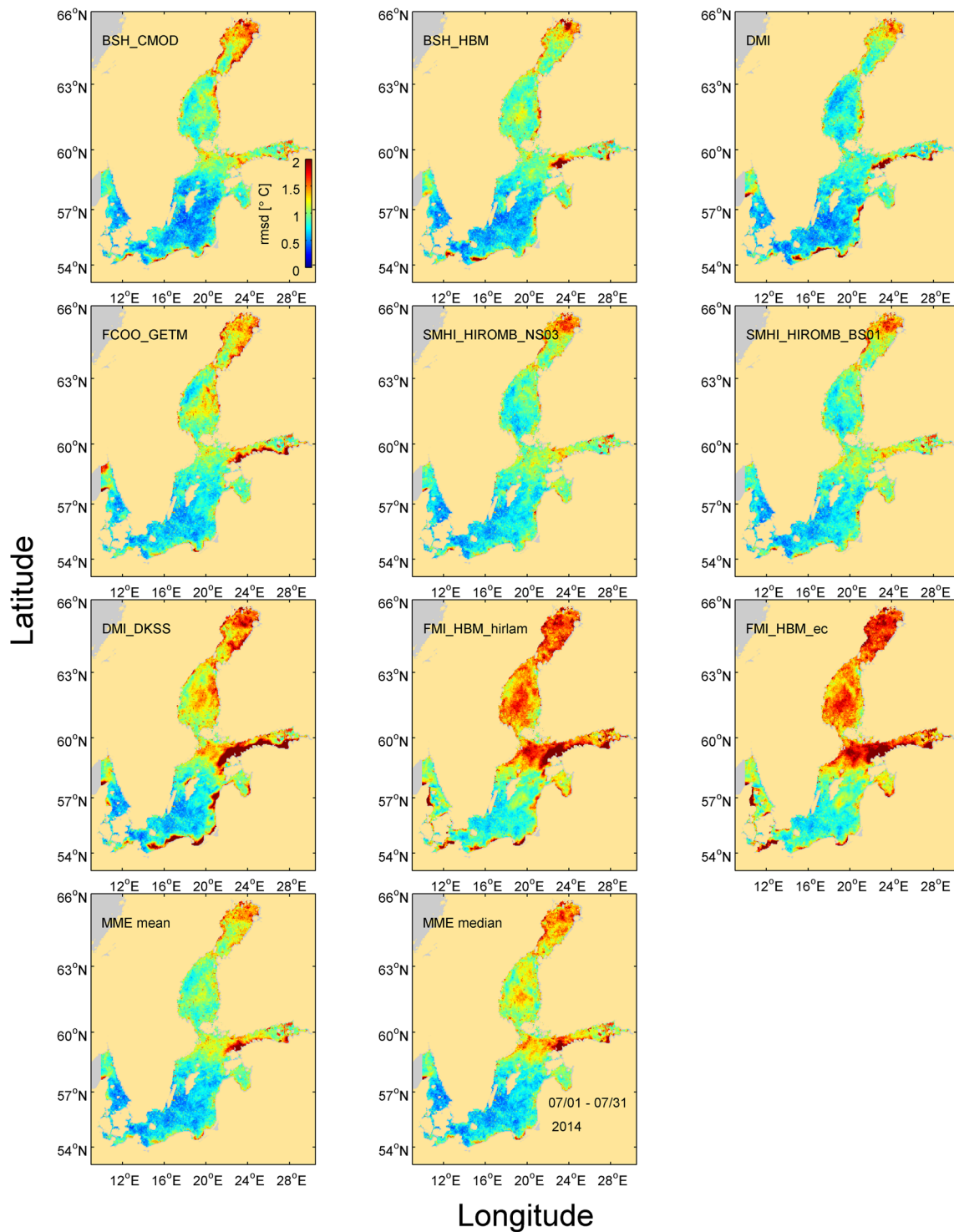


Fig. 13 Spatial distribution of the RMSD from the individual ensemble members and the MME mean (*bottom center*) in the Baltic Sea in July 2014

The large release of freshwater from the Baltic Sea is indicated by large salinity gradients in the Skagerrak and Kattegat. Brackish outflow from the Baltic Sea to the Kattegat and high-saline waters entering the Skagerrak give rise to strong salinity gradients at the surface (Gustafsson 1997a, b; Rodhe 1998). The standard deviation for this region is about 4, but with maximum values in the Northern Skagerrak of almost 9

(Fig. 14). As mentioned in Sect. 2.1, the models cover different domains and some of them do not include the Baltic Sea. The eastern lateral boundaries of those models, and thus the Baltic outflow, are defined as river inputs in the North Sea.

The largest area with average standard deviations greater than 1 is located in the region off the southern Norwegian coast, where the extension of low-salinity water from the coast varies daily.

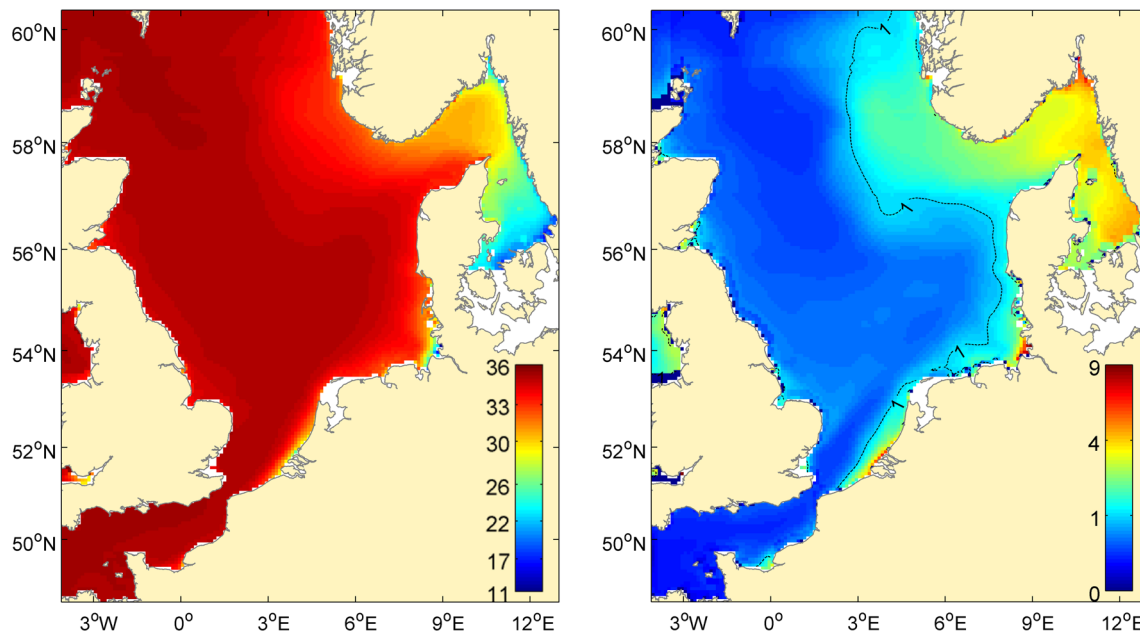


Fig. 14 Temporally averaged MME mean (*left*) and standard deviation (*right*) of SSS in the North Sea in 2014. The *black line* marks the contour line of standard deviation 1, which proceeds approximately along the salinity front of 34

This has already been detected in previous studies (Rodhe 1998; Hordoir et al. 2013). The contour line marking the standard deviation of 1 proceeds approximately along the salinity front of 34. Such patterns have also been found in comparisons of climate models (Bülow et al. 2014). It reflects the difficulties in simulating the low-salinity fronts in the Norwegian Coastal Current.

The daily spatial averages of SSS from the MME products and the individual forecasts in the North Sea are compared in Fig. 15. Systematic offsets can be detected between the time series from the different forecasts. The salinities from METNO_ROMS are much higher than those from the other forecasts, accounting for 35.6. Except for summer time, the MyOcean product (METUK_FOAM) has the lowest salinities of about 34.2. The values from the remaining models vary between 34.5 and 34.9. As mentioned in Sect. 2.1, the Baltic

outflow in the models ROMS and FOAM_AMM is defined as large river input with the eastern boundary located in the Kattegat, whereas the magnitude of river runoff possibly differs between these models. The different extensions of the model grids and different boundary conditions might play an important role in the uncertainty for this region.

Between middle of June and middle of August, the salinities from almost all models, except METNO_ROMS and METUK_FOAM, drop strongly, reaching a spread of 0.8. SMHI_HIROMB_NS03 shows a marked decline with salinities decreasing by 0.7 from about 34.5 to 33.8. In comparison, the values from the other forecasts decrease by about 0.2 and 0.5. Except during summer, the salinities of the MME products are about 34.7, following the seasonal pattern reflected by the ensemble members (Fig. 15). The values of the MME mean also

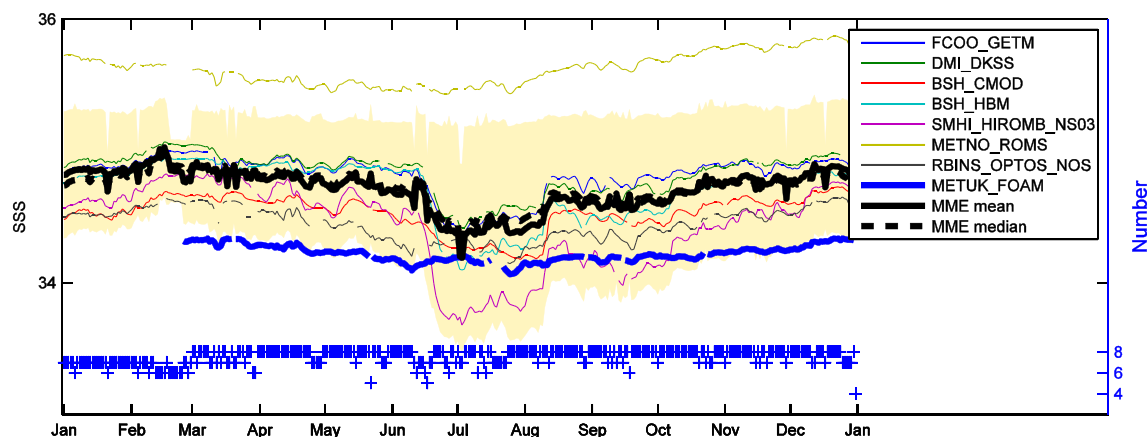


Fig. 15 Daily spatial averages of SSS from the MME mean, MME median (*black lines*), and the individual forecasts in the North Sea in 2014. The ensemble spread (\pm standard deviation) is indicated by the *yellow-shaded field*. The number of forecasts is marked as *plus sign*

depend on the number of forecasts on the current day. If one forecast is missing, the MME mean might change sharply. By the middle of February, for instance, the absence of METNO_ROMS for a few days caused a strong decrease of the MME mean. The MME median is not affected in that extent.

In order to explain the severe drop in June and August, temporally averaged MME mean and standard deviation over this period are compared to the corresponding statistics for May (Fig. 16). In May, low-salinity surface water flowing out the Baltic Sea is localized close to the Norwegian coast where standard deviation is high. Between 24 June and 8

August, the low-salinity front extends to the center of the North Sea. The region with high standard deviation is expanded accordingly. This extension in summer is presented by many studies, e.g., (Gustafsson 1997b; Rodhe 1998). The extension of the low-salinity water leads to the severe drop of salinity in June and August shown in Fig. 15. The increase of ensemble spread (Fig. 15) is caused by the enlargement of the area with high standard deviations (Fig. 16). The comparison between the two periods indicates that large errors might occur in the simulated extension of the low-salinity water in summer.

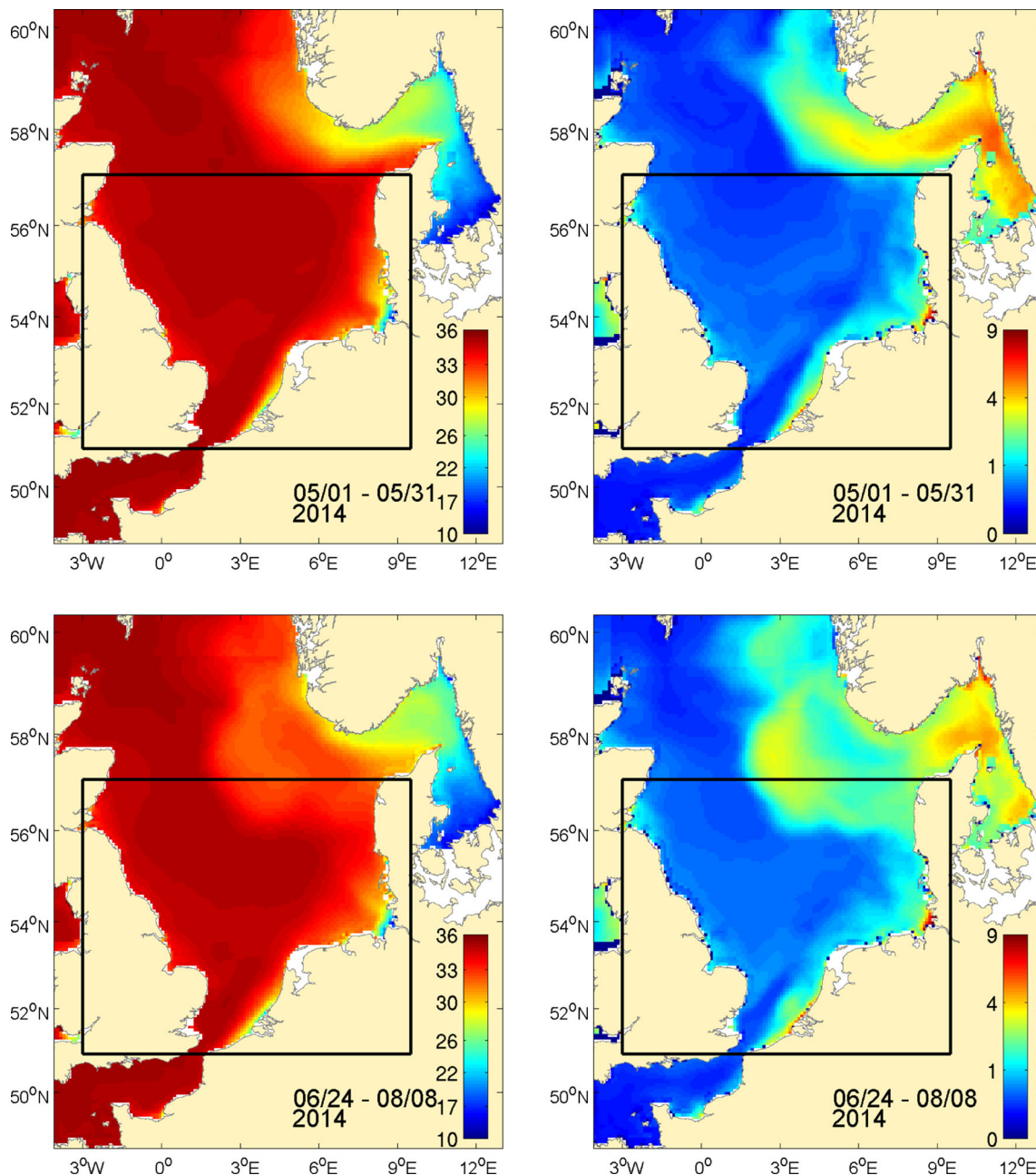


Fig. 16 Temporally averaged ensemble mean (*left*) and standard deviation (*right*) of SSS for the period with small uncertainties (*top*) and for the period with large uncertainties (*bottom*). The region, which is used for the spatial average (Fig. 15), is marked by the *black frame*

4.2.2 Baltic Sea

The annual averages over 2014 of the MME mean and the standard deviation for SSS are displayed in Fig. 17. The largest deviations occur in the Kattegat and Skagerrak as well as in the Vistula Lagoon and Curonian Lagoon in the South with values higher than 2 throughout these areas. High uncertainties in the lagoons are probably caused by the different bathymetries used for the models. Deviations in the Gulf of Finland and Gulf of Bothnia are slightly less than 1. These values are accompanied by salinities below 4 indicating that the relative uncertainties are high in these areas.

The daily spatial averages of SSS of the MME products and the individual forecasts are displayed in Fig. 18. The values of all forecasts vary almost simultaneously with time between 7 and 8.5. FCOO_GETM has the greatest offset to the MME products. Obvious discontinuities are shown in the time series from DMI_DKSS. The MME mean and MME median reflect the main features shown by the individual forecasts. Obvious differences between MME mean and median can be observed during the first half of the year, but in July, the time series converge. The spread of the ensemble members is quite stable during the whole year and no seasonal pattern can be distinguished (Fig. 18).

4.3 Regional pattern in forecast deviation for sea surface current

For the evaluation of SSC and surface transports (see Sect. 4.4), some transects are chosen representing the main inflow and outflow areas of the North Sea and Baltic Sea. The groups of transects are listed in Table 3. Evaluation of SSC is primarily

done by determining the final displacement, hence the distance between the end points of the MME PVD and the PVD of each forecast yielding a matrix with distances (km) for each day and each transect (as illustrated in Fig. 7). The temporal mean of final displacements reflects the mean differences in deviation from the MME at each transect for each forecast.

In Fig. 19, an example matrix for the North Sea, represented by FOAM_AMM, and the temporal mean of all corresponding final displacements in these regions are shown. Vertical white lines in the matrix indicate no data on that day. For a promising statistical evaluation, a nearly gapless data set is essential and therefore a constant data flow is quite important. The greatest displacements in the matrix occur in the Norwegian Coastal Current, region III. This pattern is also reflected in the temporal mean displacements of all forecasts, indicating high uncertainties in SSC magnitude and direction, which is possibly caused by different boundary conditions of the models. Regions I and II are characterized by generally lower values in both plots. In region III, METNO_ROMS exhibits the greatest displacements at most transects, while this forecast exhibits lower values in the other areas.

In regions I and II, FCOO_GETM, SMHI_HIROMB_NS03, and BSH_CMOD have also the lowest displacements at most transects. It should be mentioned that at Tr1, Tr2, and Tr15, the forecasts contributing to the MME are only METUK_FOAM and METNO_ROMS. Nevertheless, the displacement of both forecasts is comparatively high at Tr2 in region III.

An example matrix for the Baltic Sea (BSH_CMOD) is shown in Fig. 20, accompanied by the temporal mean final displacement of all forecasts. High values are distributed evenly among the transects, indicating no strong regional differences. But, there seems to be a seasonal component which is reflected

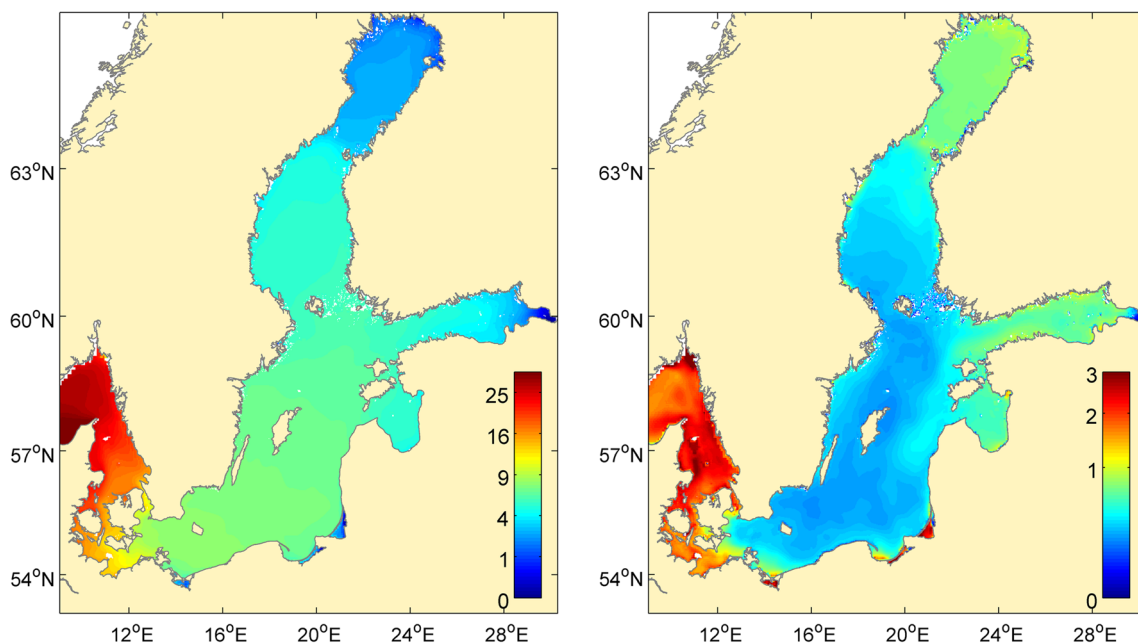


Fig. 17 Temporally averaged MME mean (*left*) and standard deviation (*right*) of SSS in the Baltic Sea in 2014

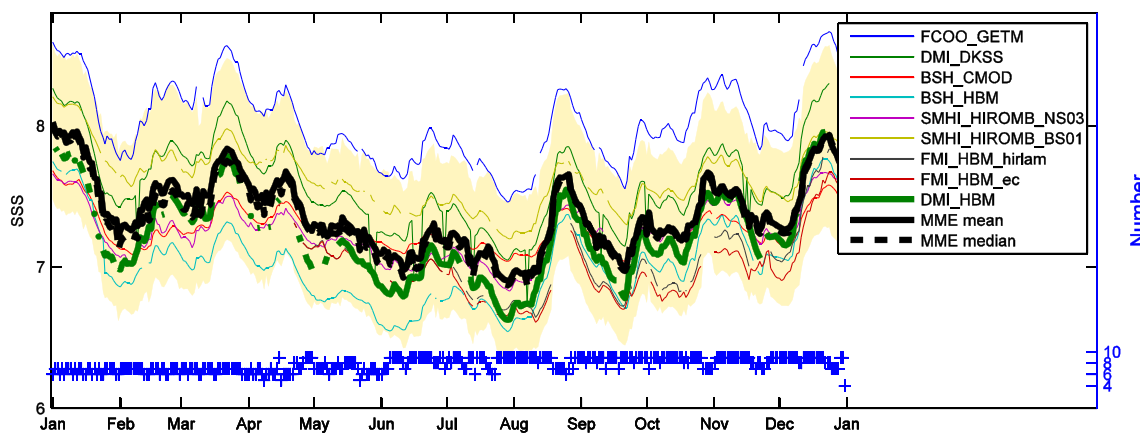


Fig. 18 Daily spatial averages of SSS from the MME mean, MME median (*black lines*), and the individual forecasts in the Baltic Sea in 2014, except for the forecast from MSI. The ensemble spread is indicated by the *yellow-shaded field*. The number of forecasts is marked as *blue plus signs*

by comparatively higher displacements at most transects during winter months, mainly in region V. This period is followed by distinct lower values in spring. Compared to the other forecasts, BSH_CMOD is in the normal range of displacements. In contrast, DMI_HBM exhibits the highest values at most transects in both regions, the Central Baltic Sea and the Gulf of Finland. Forecasts with lowest values differ between transects. The greatest range of displacements occurs at Tr50 varying between 4.8 km (SMHI_HIROMB_NS03) and 10 km (DMI_HBM). The displacements in the Baltic Sea, where maximum values range from 9 to 11 km, are low in comparison to the North Sea with maximum displacements varying between 10 and 30 km depending on the region. This is due to generally higher SSC in the North Sea than in the Baltic Sea.

General difficulties in comparing SSC in region IV, the Skagerrak and Kattegat, occur due to different model resolutions. In these highly dynamic areas, higher resolution of all forecast models would be necessary to obtain more convincing results for comparison of SSC. Therefore, this region is not evaluated here.

The relative deviation of SSC magnitude gives more information about the spread between the forecasts (Fig. 21). Higher deviations between North Sea forecasts occur in region III, thus at those transects located in the outflow area of the Baltic Sea. Low deviations occur at transects situated in the central North Sea, German Bight, and English Channel where

SSC are often highest. Regarding the deviation between Baltic Sea forecasts, comparatively high values appear at all transects, indicating a strong spread between the forecasts in the whole area. The same pattern is reflected in Fig. 20, where transects with strong differences in final displacements are correlated with high deviation in Fig. 21, e.g., Tr4, Tr13, Tr9, Tr21, and Tr22. In contrast, Tr30, Tr49, and Tr50, characterized by large spreads in Fig. 20, exhibit low relative deviations in Fig. 21 which might be due to the fact that the deviation in current magnitude is lower compared to the deviation of current components. Nevertheless, the relative deviation displayed in Fig. 21 gives more information about the real spread between the models, while Fig. 20 gives the impression that the spread at the Baltic Sea transects is very small. The reason is that the statistics are dependent on the absolute SSC values, which are mostly lower in the Baltic Sea, resulting in lower displacements of PVDs.

4.4 Regional pattern in model deviation for water transport

The percentage occurrence of the CV categories are shown as bar plot at each transect in the North Sea (a) and the Baltic Sea (b) in Fig. 22. The distribution shows that the best category 1 appears most with more than 65 % at most transects while category 3 appears less frequent. This indicates that the

Table 3 Regions and groups of transects defined for the evaluation of sea surface currents and surface transports

Number	Name of region	Number of transect
I	Inflow North Sea	Tr1, Tr4, Tr7, Tr10
II	English Channel, German Bight	Tr15, Tr14, Tr13, Tr12, Tr11, Tr19, Tr20
III	Norwegian Coastal Current	Tr23, Tr22, Tr21, Tr9, Tr6, Tr5, Tr2
IV	Skagerrak, Kattegat	Tr28, Tr27, Tr26, Tr25, Tr24
V	Gulf of Finland	Tr46, Tr45, Tr44, Tr43
VI	Central Baltic Sea	Tr52, Tr50, Tr49, Tr41, Tr36, Tr32, Tr30, Tr29
VII	Gulf of Riga	Tr38, Tr42

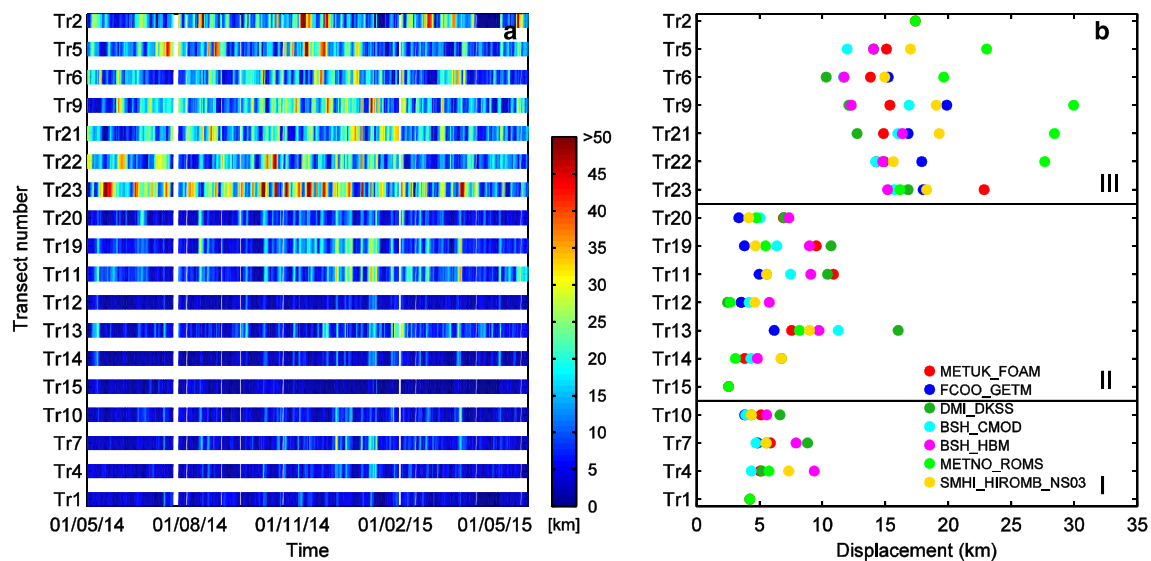


Fig. 19 An example matrix showing the daily distance between the end points of the PVD of the MME and those of FOAM_AMM in the North Sea (a) and the temporal mean of final displacements for each forecast (b)

are displayed for regions I, II, and III (see Fig. 7 for transect locations) for the time period 01.05.2014–31.05.2015

transport data are mostly consistent at all transects. In the Transition Area and the Straits, Tr23–Tr29, category 1 appears with more than 80 %. Also, transects situated in the Norwegian Coastal Current show considerably consistent results (Tr2, Tr8, Tr9). High agreement between transport data also exists at Tr11 and Tr13 located in the English Channel. However, in the lower central North Sea, there are some transects, i.e., Tr10 and Tr12, with higher uncertainties. Tr10 is situated in a region where water masses from the North Atlantic, coming down the British coast, change toward the east, which is visible in the mean circulation of the North Sea (Backhaus 1989). High

uncertainties in daily transport data across this transect might arise due to different model results of currents. In the Baltic Sea, there is less agreement at Tr31, Tr38, Tr39, and Tr42. Those short transects are located between the mainland and small islands, where different bathymetries might have a strong influence on the model results.

The mean correlation between the forecasts (R_{mod}) and the mean correlation between the MME and the forecasts (R_{MME}) are calculated for the regions defined in Table 3 and displayed in Fig. 23. R_{mod} is by definition always lower than R_{MME} , with most of the correlations ranging between 0.8 and 1.0.

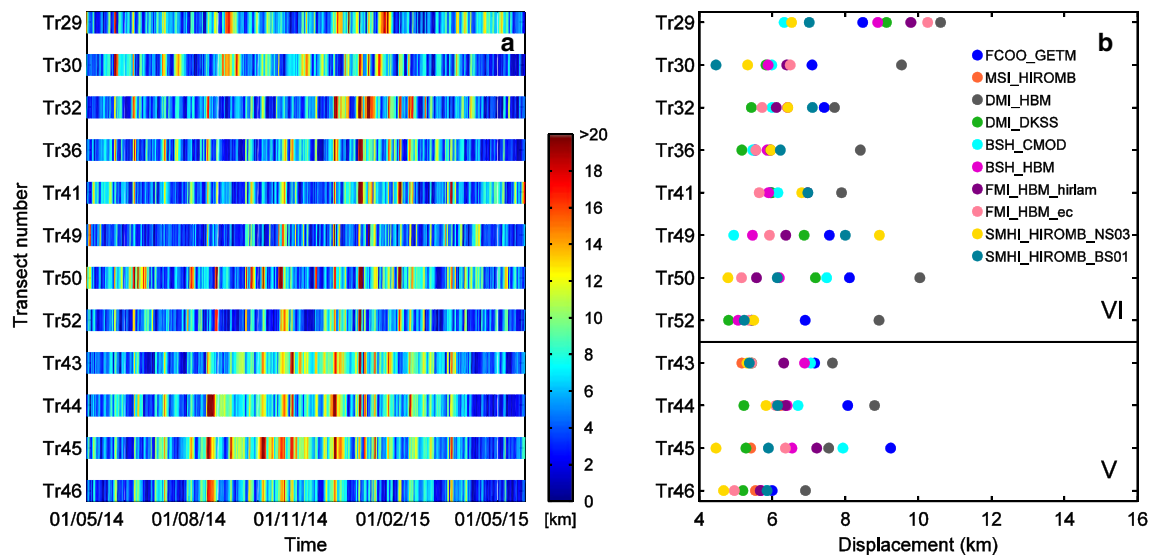


Fig. 20 An example matrix showing the daily distance between the end points of the PVD of the MME and those of BSH_CM0D in the Baltic Sea (a) and the temporal mean of final displacements for each forecast (b)

are displayed for regions V and VI (see Fig. 7 for transect locations) for the time period 01.05.2014–31.05.2015

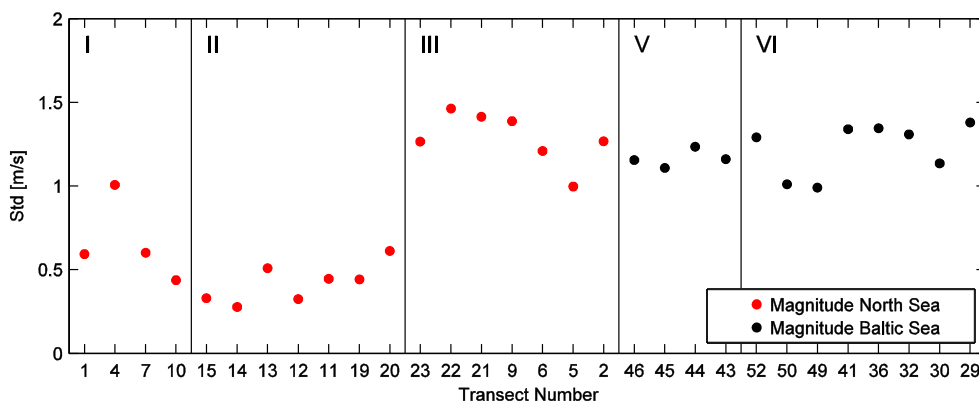


Fig. 21 The relative deviation of SSC magnitude between the forecasts is displayed for transects in regions I, II, III, V, and VI

Strong differing values of R_{mod} and R_{MME} occur in the northern part of regions I and III (Tr1, Tr2, and Tr5) and in the western part of region II (Tr14 and Tr15). Some irregularities between the products might evolve because these transects are located close to the boundaries of some model domains. At the remaining transects in regions I–III, correlations are stable with more than 0.85 for R_{mod} . Similar to the CV statistics, the correlations have the highest values in region IV (Tr24–Tr28) and in the Gulf of Finland (Tr43–Tr46). Tr38 and Tr42 in region VII have already been detected in the CV statistics and are characterized by very low correlation. Time series of transports at these transects reveal that FCOO_GETM has comparatively strong differences in transport patterns.

The normalized RMSD between each product and the MME median (Fig. 24) displays the mean error independent of the absolute transport values, which differ severely over the

whole study area. Transects with high uncertainties, already detected in the CV statistics (Fig. 22) and the correlations (Fig. 23), are accompanied by high mean error for at least one product in Fig. 24. For instance, low correlations at the boundaries (Tr1, Tr2, Tr5, Tr14, Tr15) are characterized by high RMSD and a low number of products (only 3) contributing to the MME. At Tr7, Tr10, and Tr18 in regions I and II, the high RMSD of DMI_HBM has low effect on the correlation at those transects, probably because the MME is calculated with six products. This indicates that the number of contributions to the MME is important. DMI_HBM has high RMSD at most transects, which are covered by the model. The transport values and thus current values of that model are often higher compared to the other products. This is also reflected in the PVD statistics, where the final displacements of DMI_HBM are highest in regions V and VI (Figs. 20 and 24). The closest products to

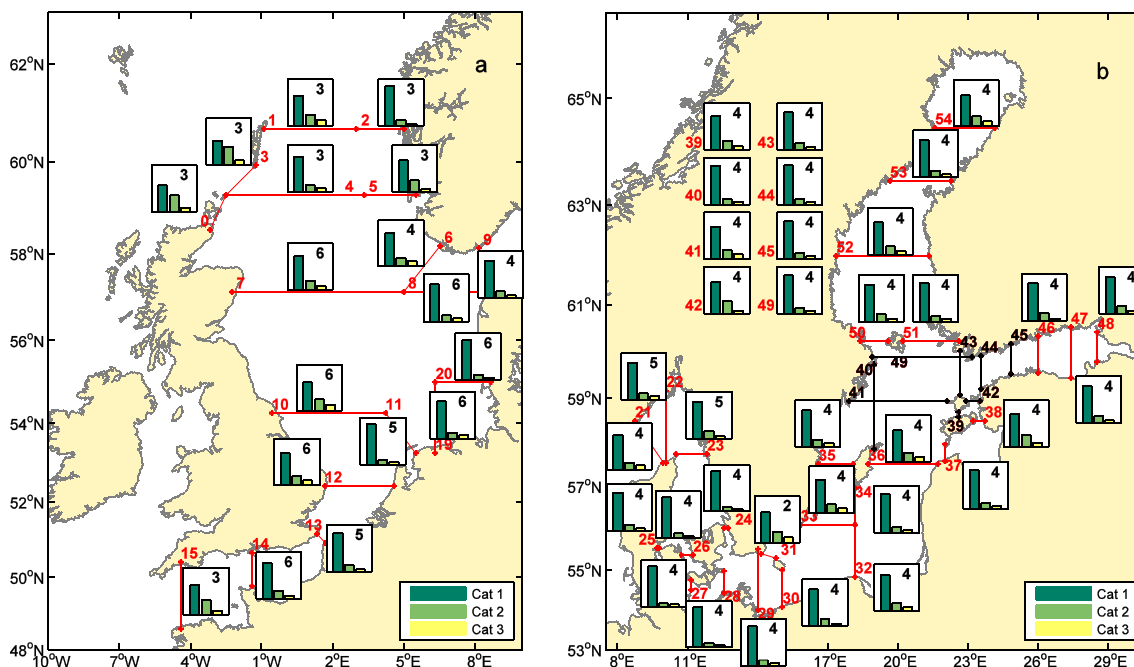


Fig. 22 Percentage occurrence of CV categories for the time period 01.04.2013–31.05.2015 in the North Sea (a) and the Baltic Sea (b). The height of each box marks 100 %. The number of partners providing data is indicated by the number in each box

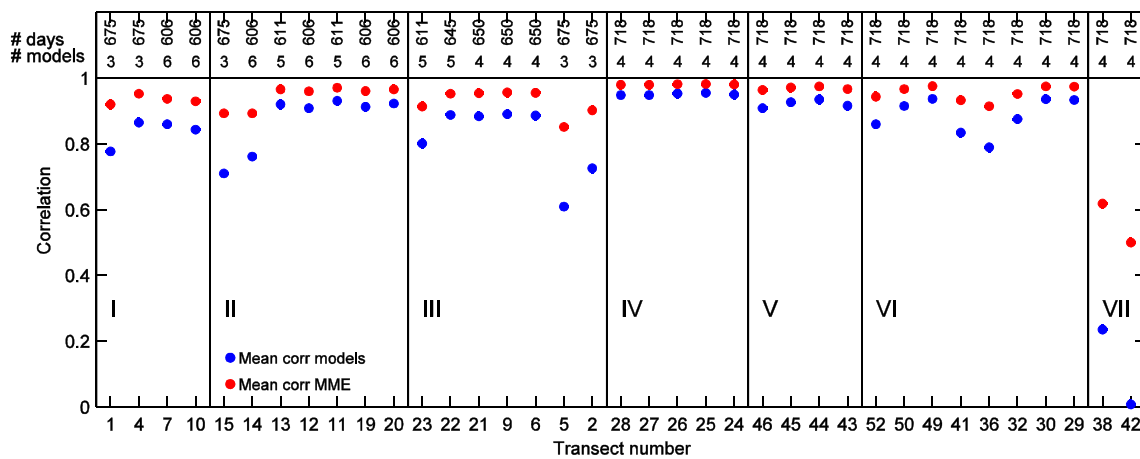


Fig. 23 Mean correlation between the forecasts (blue) and between the MME and the forecasts (red) at transects in regions I–VII for time series covering the period 01.04.2013–31.05.2015. The number of data points and products involved in the statistics are also indicated

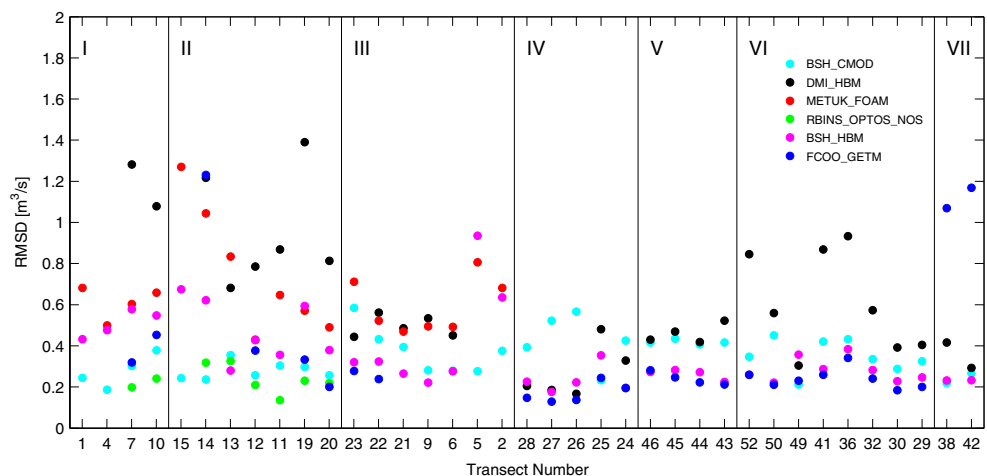
the MME median are FCOO_GETM and BSH_HBM at most transects in regions IV–VI and RBINS_OPTOS_NOS in regions I and II. The high RMSD of FCOO_GETM in region VII, already detected in Fig. 23, is caused by opposed transport patterns in the data.

5 Summary

A description of a new MME for SST, SSS, SSC, and TRA has been provided, and the contributing individual ocean forecasting models have been presented. The models are characterized by differences in numerical schemes, parametrizations, boundary conditions, forcing fields, and spatial resolutions. The processes of the MME system have been described, and some examples of the daily products including ensemble statistics such as standard deviation, ensemble mean, and ensemble median have been shown. In this study, the uncertainty between the forecasts is mainly expressed by the standard deviation and individual methods of comparisons, such as CV index or RMSD.

In the daily products, high standard deviation for SSS was detected mainly in the Skagerrak and Kattegat, thus the Transition Area between North Sea and Baltic Sea. SSC are evaluated by standard deviation, stability, and angular difference between the MME mean and the nominal MyOcean product. Regions with low current strength in the MME mean are accompanied by high relative uncertainty between the forecasts. However, as expected, the standard deviation is high in regions with stronger currents, since it scales with the absolute current values. Areas of high and low standard deviation vary also with the tides. The PVDs, calculated from the SSC time series, revealed a variety of patterns typical for different regions of the study area ranging from tidally dominated in the North Sea to density- and wind-driven currents in the Transition Area and the Baltic Sea. Large disagreements in the time series and corresponding PVDs of the forecasts are caused by variations in current amplitude and phase. These differences are possibly related to the different boundary conditions and turbulence schemes of the models. The deviation of transport data is expressed by the variation coefficient which was found

Fig. 24 The normalized RMSDs between the time series of each product and the MME median covering the period 01.04.2013–31.05.2015 are displayed for each region I–VII



to be critical when the MME mean is close to zero. Nevertheless, it gives information about the variability between the products which appears to be low in most parts of the study area. Transports are calculated using residual currents while the SSC, characterized by high uncertainties, include tides. This consolidates the assumption that the boundary conditions, and thus tidal constituents, play an important role in the forecast uncertainties. Further comparisons should be performed using residual currents to evaluate the causes for high deviation between forecasts.

Spatio-temporal statistics have been calculated yielding information about possible seasonal deviation patterns or regional differences between the forecasts, including information about forecast deviations from the MME or from observations. The region with high forecast uncertainty for SSS and SSC was found to be the highly dynamic Transition Area and the Norwegian Coastal Current, where large differences in PVD displacements occur (Fig. 18). This pattern is not fully reflected in the transport data since CV statistics and correlations have comparatively good results at most transects in this area. As mentioned above, high uncertainties in SSC, as reflected in the PVDs, might therefore be due to differences in boundary conditions and tidal constituents of the models, while transports are calculated from residual currents. The major cause for high standard deviation in SSC and SSS in this area are difficulties in simulating the frontal structures and movements of the low-salinity water of the Baltic outflow. The vertical coordinate systems and turbulence schemes of the individual models are different causing varying distributions of density and mixed layer depths, both having a strong effect on the surface parameters. In addition, there are two models which cover only the North Sea where the eastern boundary is located in the Kattegat. This might complicate a correct simulation in this area. High uncertainties in SSS between the individual forecasts, simulating the salt plume in the Baltic outflow area, are displayed in Fig. 16. Regions close to river mouths are also prone to high forecast uncertainty for SSS due to different data sets for river runoff used by the forecasting models.

Regarding forecast inter-comparisons, no forecast could be revealed which deviates most from the others in the whole study area for all parameters. The amount of deviation of each forecast for SSC is dependent on the area. In region III, the Norwegian Coastal Current, METNO_ROMS has the highest PVD displacements at most transects, while in the Baltic Sea, DMI_HBM has the highest values at most transects in regions V and VI. This is also displayed in the spatio-temporal statistic of TRA where DMI_HBM exhibits higher deviation from the median at most transects in regions I–III, V, and VI. Transects, where products with opposed or extremely differing transport pattern are included in the MME, are clearly detectable in the mean correlation and the deviation from the median (i.e., region VII). As the number of ensemble members in the MME of TRA is relatively low at most transects, products with a

strongly differing pattern also have a strong impact on the MME (Figs. 23 and 24). The same effect is displayed in the spatial mean of SSS in the North Sea, where METNO_ROMS has clearly higher values than the other forecast throughout the whole year. On those days, when this forecast is missing, the MME mean varies sharply and the standard deviation drops significantly. In the Baltic Sea, FCOO_GETM has the highest deviation in SSS from the MME mean although the spread between the forecasts varies little during the whole analysis period.

A comparison of SST forecast to satellite observations showed that the biases and RMSD of the MME mean, the MME median, and METUK_FOAM are lowest in the North Sea compared to the other forecasts. In addition, a distinct seasonal pattern has been detected, characterized by high spread between the forecasts during summer. Although the availability of satellite data varies strongly between the months, there seems to be no clear link between the monthly mean RMSD and the availability of satellite observations. In order to fully understand the mechanisms influencing the forecast uncertainties in SST including seasonal features, the atmospheric forcing of each forecast needs to be taken into account. Regarding the annual mean bias, the lowest errors and thus values close to zero appear for METUK_FOAM and METNO_ROMS, which both imply data assimilation. These results show that the ensemble process can improve the accuracy of the forecasts. Similar seasonal patterns with higher values for RMSD and bias and larger spread between the forecasts during summer can be detected in the Baltic Sea. The forecasts of the models applying data assimilation, SMHI_HIROMB_NS03 and SMHI_HIROMB_BS01, also exhibit comparatively low errors.

This study has demonstrated that the MME is a useful tool to evaluate the spread, based on uncertainty measures, between individual forecasts for different parameters. The comparison of SST forecast to satellite observations showed that the combined MME products provided better results than most individual forecasts. However, the low number of MME members (i.e., 3 to 4 for TRA), and thus a non-representative spread, seems to have an impact on the results. Thus, a large number of ensemble members are quite important for a qualitatively promising MME. In this study, both mean and median have been taken into account for evaluation. As no weighting is applied on the individual forecasts, the resulting MME mean is prone to outliers, especially in a low-member MME. In contrast, the MME median is less impacted by outliers. Therefore, the spatio-temporal statistics for TRA have been calculated using the median, due to the low number of products.

In this study, it was not intended to undertake validation for each forecast or to provide the best overall estimate with the MME for all parameters. The latter simply cannot be realized due to the lack of reliable in situ data with

reasonable coverage of the North Sea and Baltic Sea. The comparison with satellite observations has given the opportunity to evaluate the real spread between the forecasts. Comparisons of SSS and SSC forecasts to in situ data would yield important information about RMSD. In the future, it could be useful to assess weighting to the individual forecasts based on either model resolution or based on model performance as a result of model validation done by each institute. At present, a MME of sea bottom salinity and sea bottom temperature is being developed, which can give the opportunity to study stratification issues, especially in the Transition Area between North Sea and Baltic Sea.

Acknowledgments The authors would like to thank Inge Menzenhauer-Schumacher who designed and created the webpages for the MME on the NOOS and BOOS website. Many thanks also to Simon Jandt who supported the comparison of satellite data with the MATLAB CalVal Toolbox. E.W. Blockley and A. Sellar would like to thank Rachel Furner, who has helped with providing the NOOS transects data for the Met Office.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alvera-Azcárate A, Troupin C, Barth A, Beckers J-M (2011) Comparison between satellite and in situ sea surface temperature data in the Western Mediterranean Sea. *Ocean Dyn* 61:767–778
- Aribas A, Robertson KB, Mylne KR (2005) Test of a poor Man's ensemble prediction system for short-range probability forecasting. *Mon Weather Rev* 133:1825–1839
- Atger F (1999) The skill of ensemble prediction systems. *Mon Weather Rev* 127:1941–1953
- Axell L (2013) BSRA-15: a baltic Sea reanalysis 1990-2004 vol SE-60176 swedish meteorological and hydrological institute. Norrköping, Sweden
- Backhaus JO (1989) The North Sea and the climate. *Dana* 8:69–82
- Becker J (2007) BMA for ECOOP. Technical Report. Deltares, Delft
- Berg P (2012) Mixing in HBM vol 12-03. Danish Meteorological Institute, Copenhagen, Denmark
- Berg P, Poulsen JW (2012) Implementation details for HBM vol 12-11. Danish Meteorological Institute, Copenhagen
- Brown CE (1998) Coefficient of variation. In: Applied multivariate statistics in geohydrology and related sciences. Springer Berlin Heidelberg, New York, pp 155-157
- Brüning T et al. (2014) Operational Ocean Forecasting for German Coastal Waters Die Küste 81
- Büchmann B, Hansen C, Söderkvist J (2011) Improvement of hydrodynamic forecasting of danish waters: impact of low-frequency north atlantic barotropic variations. *Ocean Dyn* 61:1611–1617
- Buizza R, Richardson DS, Palmer TN (2003) Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *Q J R Meteorol Soc* 129:1269–1288
- Bülöw K et al. (2014) Comparison of three regional coupled ocean atmosphere models for the North Sea under today's and future climate conditions (KLIWAS Schriftenreihe; KLIWAS-27/2014) KLIWAS Schriftenreihe 27
- Burchard H, Bolding K, Villarreal MR (1999) GOTM – a general ocean turbulence model. Theory, applications and test cases. European Commission. <http://www.gotm.net/pages/documentation/manual/stable/pdf/a4.pdf>. Accessed 28 May 2014
- Burchard H, Janssen F, Bolding K, Umlauf L, Rennau H (2009) Model simulations of dense bottom currents in the western Baltic Sea. *Cont Shelf Res* 29:205–220
- Burchard H, Bolding K, Umlauf L (2010) GETM Source Code and Test Case Documentation Version pre 2.0.x. <http://www.getm.eu/data/getm/doc/getm-doc-devel.pdf>. Accessed 28 May 2014
- Daley R (1991) Atmospheric data analysis, Chapter 3. Cambridge Atmospheric and Space Sciences Series. Cambridge University Press
- Dick S, Kleine E (2008) Janssen F A new operational circulation model for the North Sea and Baltic Sea using a novel vertical co-ordinate setup and first results. In: Dahlin H, Bell MJ, Flemming NC, Petersson SE (eds) Coastal to global operational oceanography: achievements and challenges. Proceedings of the Fifth International Conference on EuroGOOS, Exeter, pp 225–231
- Dick S, Kleine E, Müller-Navarra S, Klein H, Komo H (2001) The operational circulation model of BSH (BSH cmod). Model description and validation vol 29. Bundesamt für Seeschifffahrt und Hydrographie, Hamburg
- Doblas-Reyes FJ, Hagedorn R, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting - II. Calibration and combination. *Tellus* 57A:234–252
- Ebel M, Becker J (2010) Operational river and coastal water level forecast using Bayesian Model Averaging. Deltares
- Ebert EE (2001) Ability of a poor Man's ensemble to predict the probability and distribution of precipitation. *Mon Weather Rev* 129:2461–2480
- Emery WJ, Thomson RE (2001) Data analysis methods in physical oceanography, 2nd edition. Elsevier
- Evans RE, Harrison MSJ, Graham RJ, Maylne KR (2000) Joint medium-range ensembles from the Met. Office and ECMWF systems. *Mon Weather Rev* 128:3104–3127
- Feistel R, Nausch G, Wasmund N (2008) State and Evolution of the Baltic Sea, 1952–2005: A Detailed 50-Year Survey of Meteorology and Climate, Physics, Chemistry, Biology, and Marine Environment. Wiley
- Funkquist L, Kleine E (2007) An introduction to HIROMB, an operational baroclinic model for the Baltic Sea vol SE-60176. Swedish Meteorological and Hydrological Institute, Norrköping
- Gustafsson B (1997a) Dynamics of the seas and straits between the Baltic and North Seas: a process-oriented oceanographic study. Göteborg University
- Gustafsson B (1997b) Interaction between Baltic Sea and North Sea. *Deutsche Hydrographische Zeitschrift* 49:165–183
- Hofmeister R, Beckers J-M, Burchard H (2011) Realistic modelling of the exceptional inflows into the central Baltic Sea in 2003 using terrain-following coordinates. *Ocean Model* 39:233–247. doi:10.1016/j.ocemod.2011.04.007
- Hordoir R, Dieterich C, Basu C, Dietze H, Meier HEM (2013) Freshwater outflow of the Baltic Sea and transport in the norwegian current: a statistical correlation analysis based on a numerical experiment. *Cont Shelf Res* 64:1–9
- Houtekamer PL, Lefaiivre L, Derome J, Ritchie H, Mitchell HL (1996) A system simulation approach to ensemble prediction. *Mon Weather Rev* 124:1225–1242
- Jandt S, Lagema P, Janssen F (2014) A comprehensive validation toolbox for regional ocean models - Outline, implementation and application to the Baltic Sea. In: Geophysical Research Abstracts: EGU General Assembly 2014, held 27 April - 2 May Paper presented at the EGU General Assembly 2014, Vienna, Austria. <http://meetingorganizer.copernicus.org/EGU2014/EGU2014-11198.pdf>

- Kleine E (1994) A class of hybrid vertical coordinates for ocean circulation modelling. Bundesamt für Seeschifffahrt und Hydrographie, Hamburg
- Krishnamurti TN et al. (1999) Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble vol 285
- Lagemaa P (2012) Operational Forecasting in Estonian Marine Waters. TTU Press
- Lagemaa P, Janssen F, Jandt S, Kalev K (2013) General Validation Framework for the Baltic Sea. . Paper presented at the GODAE OceanView Symposium 2013, Baltimore, 4 - 6 November 2013. <http://www.godae.org/~godae-data/Symposium/GOV-posters/S3.3-08-Lagemaa.pdf>
- Lenartz F, Moure B, Barth A, Beckers J-M, Vandenbulcke L, Rixen M (2010) Enhanced ocean temperature forecast skills through 3-D super-ensemble multi-model fusion Geophysical Research Letters 37 doi:10.1029/2010GL044591
- Leppäranta M, Myrberg K (2009) Physical Oceanography of the Baltic Sea. Geophysical Sciences. Springer
- Logutov OG, Robinson AR (2005) Multi-model fusion and error parameter estimation. Q J R Meteorol Soc 131:3397–3408. doi:10.1256/qj.05.99
- Luyten PJ (2014) COHERENS — A Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas: User Documentation. Operational Directorate Natural Environment, Royal Belgian Institute of Natural Sciences
- Luyten PJ, Jones JE, Proctor R, Tabor A, Tett P, Wild-Allen K (1999) COHERENS - A Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas
- Madec G (2008) NEMO ocean engine. Institut Pierre-Simon Laplace (IPSL), France
- Martin M et al (2012) Group for high resolution Sea surface temperature (GHRSSST) analysis fields inter-comparisons. Part 1: a GHRSSST multi-product ensemble (GMPE). Deep-Sea Res II 77-80:21–30
- Molteni F, Buizza R, Palmer TN, Petroliagis T (1996) The ECMWF ensemble prediction system: methodology and validation. QJR Meteorol Soc 122:73–119
- Moure B, Chiggiato J, Lenartz F, Rixen M (2012) Uncertainty forecast from 3-D super-ensemble multi-model combination: validation and calibration. Ocean Dyn 62:283–294. doi:10.1007/s10236-011-0504-6
- Mylne KR, Evans RE, Clark RT (2002) Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting. Q J R Meteorol Soc 128:361–384
- O’Dea EJ et al (2012) An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European North-West shelf. J Oper Oceanogr 5:3–17
- Otto L, Zimmerman JTF, Furnes GK, Mork M, Saetre R, Becker G (1990) Review of the physical oceanography of the north Sea netherlands. J Sea Res 26:161–238
- Palmer TN et al (2004) Development of a european multi-model ensemble system for seasonal to inter-annual prediction (DEMETER). Bull Am Meteorol Soc 85:853–872. doi:10.1175/BAMS-85-6-853
- Pérez B et al (2012) ENSURF: multi-model sea level forecast – implementation and validation results for the IBIROOS and Western Mediterranean regions. Ocean Sci 8:211–226
- Poulsen JW, Berg P (2012) More details on HBM - general modelling theory and survey of recent studies. Danish Meteorological Institute
- Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005) Using bayesian model averaging to calibrate forecast ensembles. Mon Weather Rev 133:1155–1174
- Richardson DS (2001) Ensembles using multiple models and analyses. Q J R Meteorol Soc 127:1847–1864
- Rodhe J (1998) The Baltic and North Seas: a process-oriented review of the physical oceanography. In: Robinson AR, Brink KH (eds) The sea, vol 11. Wiley, New York, pp 699–732
- Shchepetkin AF, McWilliams JC (2005) The regional ocean modeling system: a split-explicit, free-surface, topography following coordinates ocean model. Ocean Model 9:347–404
- Song Y, Haidvogel D (1994) A semi-implicit ocean circulation model using a generalized topography-following coordinate system. J Comput Phys 115:228–244
- Toth Z, Kalnay E (1993) Ensemble Forecasting at NMC: The Generation of Perturbations. Bull Am Meteorol Soc 74:2317–2330
- Weisheimer A et al. (2009) ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs vol 36
- Ziehmann C (2000) Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. Tellus 52A:280–299