Ocean Sci. Discuss., 12, C251–C266, 2015
www.ocean-sci-discuss.net/12/C251/2015/

Ocean Science Discussions

Interactive comment on "Evaluation of numerical models by FerryBox and Fixed Platform in-situ data in the southern North Sea" *by* M. Haller et al.

M. Haller et al.

michael.haller@hzg.de

Received and published: 5 June 2015

Author's Response to referee #3 (Roger Proctor)

The authors would like to thank Roger Proctor for the extensive and constructive review for the benefit of the submitted paper.

The responses to the comment points:

1) The paper has a number of authors, and unfortunately it shows because different sections of the paper are clearly written in different styles with different levels of written English proficiency; it would benefit from a single author (i.e. Siddorn) acting as editor.

The paper has been edited before and now again by a native speaker (i.e. Siddorn), so

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

we hope, that different styles of written English are now less obvious (other reviewers did not complain about low proficiency of the paper).

2) The plethora of statistics is a bit overwhelming and could be trimmed back to the essence of the comparisons, the reporting of the statistics is rather monotonous with little insight into the reasons for the differences.

There are figures of the model-observation differences as well as the presentation of the common basic statistical measures (i.e. bias, stde, rmse, skvar) which underline the findings which can be visually found in the figures. We only examine two different parameters – water temperature and salinity – so we think this should be a clear analysis. In order to present a structured analysis, we arranged the analyses of the three North Sea positions in same way with being of course at risk to present in a more monotonous way. The index of agreement (IOA) as well as the cost function (cf) also contributes to the overall picture, but we omit the analysis of the correlation coefficient – in the text and in the (scatterplot) figures.

3) I struggled with the descriptions of the data extraction from the models, it isn't clear to me exactly what model data is used to compare with the FerryBox observations, what averaging is done (vertical and horizontal) to attempt to compare like-with-like.

FerryBox data have been taken from the HZG FerryBox database. There, data are stored with 10 seconds resolution. The GPS coordinates of the transect between England and Germany have been used with 0.05° resolution. Model data have been taken from HZG model archive (BSHcmod) and from MyOcean database (AMM7). For each transect position, a search radius of 0.02° and a time window of ± 1 hour has been used for nearest neighbor interpolation of model data. For the detailed analysis of three positions in the North Sea (English East coast, Oyster Ground and German Bight) at section 3.2, an internal search routine of HZG database has been applied. There, a search radius of 5 km around the fixed positions has been used for the retrieval of FerryBox data (5 km is the default search radius). The retrieved time series of FerryBox

OSD
12, C251–C266, 2015

data have been filtered for model time steps with a time range of ± 30 minutes. In the next step, the nearest model grid point of BSHcmod and AMM7 has been allocated to that fixed positions.

4) The attempts to explain the differences between models and observations are confusing and sometimes contradictory. In different places in the manuscript vertical mixing in the models is given as a reason for disagreement, but sometimes this is described as underestimating mixing and sometimes as overestimating mixing. But no effort is made to explain properly what the mixing schemes are in the models or how their differences may lead to discrepancies.

Mixing in shallow shelf seas has two main and quite different sources, i.e. bottom friction and wind/wave mixing. The overestimation and underestimation of mixing processes holds for different regions and different reasons. So there is no contradiction apparent. Vertical mixing underestimation is supposed to be a cause for weak (BSHcmod) model performance in terms of water temperatures near the English coast while overestimation of vertical mixing could be a reason for weak simulation of salinity of AMM7 for Oyster Ground (p2). To discuss in detail the mixing schemes would be beyond the scope of this paper; however, we refer to the BSHcmod model description in Dick et al. (2001), where the mixing scheme has been described in detail.

5) It is also a surprising result that AMM7, which includes data assimilation of surface temperatures, does not show better results than BSH.

The statistics for the root mean square error show clearly a difference between the AMM7 (which contains data assimilation of SST) and BSHcmod (which will contain data assimilation in the future). But that holds only for the water temperatures. There, the data assimilation brings a benefit while for salinity SSS is not available for assimilation. Korres et al. (2009) showed in their publication that SSS assimilation would bring a significant benefit for salinity simulation.

6) It is pointed out in the text that the discrepancy between AMM7 and the MARNET

OSD
12, C251–C266, 2015

observations is also surprising given better agreements in other (MyOcean) studies, but no effort is made to resolve these points.

“The bias in the AMM7 seems higher than one would expect in an assimilating model and is slightly at odds with other results (e.g. O’Dea et al. (2012) which finds a bias of around 0.1 K). This may at least be partly explained by the inability of the model to represent shallow water processes and river flows, which are particularly important in this region, and may also be partly explained by observation errors (with the potential issues arising due to bias correcting the Ferrybox data to a single point on its transect).”

7) I am puzzled by the choice of locations for comparisons . . . point P2 is described as “marks the TD/TC meeting point in the Oyster Ground area” (caption figure 1) and this caption describes route TC (Copenhagen-Bergen, in green) but no route appears on the figure and no mention of TC occurs in the text.

This description of figure 1 was from a previous version of the manuscript and was still present manuscript by mistake. Now, the caption is as follows:

"Figure 1: FerryBox routes and crossing points in the North Sea. Contour lines indicate the bathymetry. The blue line marks the TD route Cuxhaven-Immingham and the red lines indicate the LB route England-Norway-Germany. Specific analysis points of FerryBox routes are indicated by black points and labelled p1, p2, p3, respectively. p1 is situated at the English East Coast. p2 marks the analysis point in the Oyster Ground area. At p3, the MARNET station Deutsche Bucht is located."

8) I found the introduction section rather German-centric in its choice of references. Many well-known UK, Belgian, Dutch and Norwegian references are ignored in preference to some rather obscure (or less well known) German references. We restructured the overview section about efforts towards the simulation of North Sea hydrodynamics to take the European models into account. The section has been rewritten as follows:

“Since the 1980s, baroclinic 3D-models have been developed to prognostically model

OSD

12, C251–C266, 2015

water temperature and salinity variations in the North Sea. All countries around the North Sea have been contributing to this effort, i.e. Denmark (Vested et al., 1992), Norway (Svendsen et al., 1996), UK (Proctor and James, 1996), Belgium (Delhez and Martin, 1992; Luyten et al., 1996), The Netherlands (de Kok, 1997) and Germany (Backhaus, 1985; Dick et al., 2001). For the present study, two different hydrodynamic models, BSHcmod and FOAM AMM7 NEMO, were used. These models are commonly applied, also for e.g. ecosystem modelling (Edwards et al., 2012; Maar et al., 2011) and predicting wave-tide-current interactions (Pleskachevsky et al., 2009) in the North Sea."

9) The model sections need improving, it wasn't made clear until deep in the results section that the BSH model is a nested grid with different fixed resolutions, there is an implication that the grid varies BETWEEN 900m and 5km.

In the submitted version of the paper, the different grid resolutions have been made clear, but maybe the text has been missing the information, that these grid resolutions are fixed. So, the section for description of the BSHcmod model grid characteristics has been rewritten, to make it clearer:

"The model is based on the Reynolds-averaged Navier-Stokes equations which are discretized on a geographical Arakawa-C grid and on adaptive vertical coordinates. A two-way nesting approach is applied with a coarse resolution grid (5km grid spacing) in the North and Baltic Sea and a fine resolution grid (900 m grid spacing) in the German Bight and the western part of the Baltic Sea (focus region)."

10) The surface and lateral boundary conditions need to be clearer, e.g. surface waves are mentioned – are these used at all?

In BSHcmod, surface waves are not explicitly taken into account in the applied version.

11) Which rivers have daily averaged data, and what kind of climatology is applied to the others (and how many?).

OSD
12, C251–C266, 2015

For BSHcmod, daily runoff data are included from the Rhine, Eider, Elbe, Weser and Ems. They are provided by the Bundesanstalt für Gewässerkunde (BfG). For remaining rivers, a monthly climatology is commonly applied.

12) AMM7 is said to assimilate SoO data . . . where? Is this likely to impact on results in the southern North Sea?

Assimilation of data of SoO is done in AMM7 for the North Sea:

"It is worth noting that although a number of Ships of opportunity (SoO) data were assimilated into the system, including reasonable data density in the southern North Sea, the Ferrybox data being used in this study was not available for assimilation and so was not included."

13) Satellite SST is also said to be assimilated, so how is this done?

In the description of AMM7, a detailed section about satellite data assimilation is inserted:

"The system assimilates observations using an Optimal Interpolation scheme (Martin et al., 2007), with updates described in Storkey et al 2010 and adaptations to allow it to address the particular requirements for shelf applications (O'Dea et al., 2012). (Siddorn et al., 2007). The assimilation system uses First Guess at Appropriate Time (FGAT) scheme to calculate model/observation differences (innovations) which are converted to model increments using an iterative method. A daily analysis window is used, with the model being rerun for the same day with an Incremental Analysis Update (IAU) scheme to update the model state using these increments. Only SST data are assimilated. Temperature and salinity profile assimilation along with sea surface height assimilation are technically more challenging in the shelf environment and will be implemented as future developments to the system. Data assimilated include in-situ data and level 2 satellite SST data provided by the Global High-Resolution Sea Surface Temperature project (GHRSST). In-situ data are obtained from a variety of sources

OSD
12, C251–C266, 2015

and include measurements taken by ships, moored buoys, and drifters. Satellite observations are obtained from the Advanced Microwave Scanning Radiometer-Earth observing system (AMSRE), the Advanced Along-Track Scanning Radiometer (AATSR), and the Advanced Very High Resolution Radiometer (AVHRR) instruments on board the NOAA and MetOp satellites. Also assimilated are data from the geostationary Spinning Enhanced Visible and Infrared Imager (SEVIRI). All data are quality controlled and a bias correction scheme, based on comparisons to in-situ and AATSR data, is applied to the AMSRE, AVHRR, and SEVIRI observations. A full description of the satellite data types and the scheme used to bias correct them can be found in (Donlon et al., 2012)."

14) And what model data is actually used? These models are forecast models, are you using part of the forecast, initial conditions, or a specific hindcast?

For both model data sets, the analyses at forecast time step +0 have been applied. This information has been added in the text.

Technical corrections:

- Abstract: line 14 ... "Statistical errors differ between the models and the measured parameters, as the root mean square error (rmse) accounts for BSHcmod v4 to 0.92 K, for AMM7 only to 0.44 K. For salinity, BSHcmod is slightly better than AMM7 (0.98 and 1.1 psu, respectively)." Is poor English and needs rewriting

Analyses of statistical errors differ between the models and between the measured parameters, as root mean square error (rmse) of water temperatures amounts to 0.92 K (BSHcmod v4) and 0.44 K (AMM7), while for salinity the performance of BSHcmod is slightly better (0.98 and 1.1, respectively).

- P357, line 10: "salinity near the coasts is only about 15–25" – near all coasts??

That holds for the salinity near the coasts of the southern North Sea, where freshwater inflow from several rivers is mixing with sea water. To avoid confusion, we changed to:

OSD
12, C251–C266, 2015

“Because of freshwater inflow from several major rivers in the Southern North Sea (e.g. Rhine, Maas, Elbe), salinity near that riverine influenced coasts could be only about 15-25. In central parts of the North Sea the salinity amounts to 35.”

- P357, line 13: “several factors like bathymetry, density distribution and wind stress” – I would add tides to this list.

Text has been changed to:

” Besides the freshwater inflow, the North Sea is also strongly influenced by the tides and the residual circulation. This is strongly governed by several factors like bathymetry, density distribution and wind stress.”

- P358, line 8: “installed on ships of opportunity (SoO), as well as on fixed onshore stations near harbours, river banks or estuaries.” Examples of these other systems?

Helmholtz-Zentrum Geesthacht operates a FerryBox at the Elbe river estuary in a fixed onshore container that pumps water from the sea into the system. Another FerryBox is located at harbour of Helgoland island in the German Bight. Text has been inserted:. . .

” (e.g. at Cuxhaven harbour located at the Elbe river estuary). . .”

- P358, line 13: “without limitation of energy”, I think you mean “without power supply issues”?

Inside of a ship, FerryBox systems get their energy supply directly from the ship‘s engines. There is plenty of energy available, so limitation of energy (and space, by the way) is no issue for these FerryBoxes, in contrast to observations systems deployed on buoys, piles, moorings, etc. “energy” has been replaced by “power supply”.

- P359, line 22: “Satellite imagery is somewhat limited regarding the time resolution and restricted to certain parameters. Moreover, satellite coverage is limited in coastal regions and in the vicinity of land” I don’t believe this to be the case, so better justification for this statement needed.

OSD
12, C251–C266, 2015

Satellite coverage is given only several times a day at best. Moreover, salinity cannot be measured. Satellite SST has been found to be consistent to fixed station and FerryBox measurements (Grayek et al., 2011). However, due to cloud cover, satellite products of instrument like MERIS contain gaps. This is also mentioned by Petersen et al. (2008).

- P360, line 2: “The aim of this study . . .” say “The aim of the present study. . .” to avoid confusion with the previous references Wehde and Petersen.

“this” has been changed to “the present”.

- P360, line 6: “and identify the limitations and weaknesses of the operational models AMM7 and BSHcmod v4.” Beyond the obvious that both models show less good agreement with the observations near coasts I think more effort is needed to explain these differences.

Besides the indeed obvious poor performance near the coasts, the comparison of both models for water temperatures shows a clear benefit of satellite SST assimilated into the AMM7. It explains the less rmse for AMM7 (0.44 K) opposite to BSHcmod (0.92 K). Salinity observation data (satellite or in-situ) are not available. Thus, performance of both models is in the same range in general. Interestingly, the daily runoff data for the German rivers does not improve the BSHcmod significantly.

- P360, line 25: “in case of severe errors” meaning what?

In order to shorten this section, this has been removed. But severe errors occur for instance, when e.g. the pumping is not working.

- P361, line 15: Is all the technical detail really needed? A reference will do I think.

We believe that besides the details of the applied models (which have been demanded to be presented in more detail by all reviewers) also some details of the measurement devices should be briefly outlined. A reference is included but that alone is not sufficient. Yet, the technical details have been shortened.

OSD
12, C251–C266, 2015

- P362, line 3: “MARNET has a long tradition of monitoring” how long?

Some time series in the German Bight date back to 1872, but the unmanned light ships measurements started in 1984. More details can be found in the reference BSH, 2014.

- P363, line 12: “extrapolated from the lowest pressure level height” how high? Does it matter?

For getting model wind data in 10 m height, it is a common method to extrapolate wind data from the lowest pressure levels, which is e.g. for the COSMO-EU forecast model (which is run by the DWD) in approx. 20 m height.

- P363, line 28: “hybrid s-sigma terrain following coordinates are applied with 50 equally spaced levels” not sure I understand what is meant by this ... the purpose of using hybrid s-sigma is to allow some variability between levels, please explain.

The sentence has been changed to:

“To get the correct vertical resolution of the terrain, hybrid s-sigma terrain following coordinates are applied with 50 levels (interpolated onto 24 geopotential levels for data distribution).”

- P364, line 16: “the bias can then described as the mean difference between simulations and observations” is incorrect given the following equation which shows the bias to be the difference between the mean of the observations and the mean of the simulation.

The text has been changed to:

“the bias can then described as the difference between the mean of simulations and the mean of observations”

- P365, line 20: “A value of 0.5 means that the model is on average 0.5 times the standard deviation off the observations. “ What does this mean??

OSD
12, C251–C266, 2015

The cost function serves as a measure, how far the model results are off the observed variability (i.e. the standard deviation) of the analyzed parameter. Text has been changed to:

"A cf value of 0.5 means that the model error is on average 0.5 times the standard deviation of observations."

- P367, line 15-20: repeated text

The repeated text has been removed.

P368 ... A number of things need to be made clearer about how the data are extracted and compared. One factor not explored is how well the models represent the tides because even 30-minute discrepancies in tidal phase could influence the temperature/salinity values extracted from the models.

How the data are extracted and compared is more explained in point 3 of major corrections. As the window for temporal interpolation amounts to $\pm$ 1 hour, it is supposed to influence the tidal phase and this would be part of the explanation why salinity performance directly at the coasts is less good than at the central part of the transect. However, we analyzed the spectral densities of MARNET, BSHcmod and AMM7 in the German Bight and found out, that the tides are well represented in the models.

- P370, line 6: "The differences have been marked according to the double stde of the FerryBox data which has been described above" what? Where?

In order to distinguish between significant and non-significant model-observation differences, we calculated the standard deviation of error from the FerryBox data evaluation in section 2.7. We assume that model differences that are greater than the twofold stde of the FerryBox data are significant. In the figures, this threshold is marked by the use of different colours. It is also written in the caption.

The mentioned sentence has been modified for clearance as follows:

OSD

12, C251–C266, 2015

"The differences have been marked in the figure according to the double stde of the FerryBox data which has been described in the previous section 2.7."

- P370, line 12: spelling "spatial"

Has been corrected.

- P371, line 6: "the differences as well as the according measures bias, stde and skvar show average values," what is meant here?

This sentence was indeed not clear enough. The average statistical values are -0.02 (bias), 0.72 (stde) and 0.92 (rmse). The text has been rewritten:

"It should be noted that the TD transect crosses the southern North Sea approximately along the transition zone between the stratified and well mixed regions of the southern North Sea and therefore small errors in the position of the seasonal front will cause biases in this region. In the central part of the transect, stde and rmse range around the average values of 0.72 K and 0.92 K, respectively, while in the German Bight (east of 7 °E) both amount to 1 K. Near the English coast, a local maximum of 0.8 K (stde) and 1.1 K (rmse) is visible, together with a local minimum of bias and skvar, amounting to -0.8 K and 0.8, respectively. On the central parts of the transects, bias ranges around $\pm$ 0.3, while skvar ranges around 0.9 and 1.0. In the German Bight, bias is slightly more variable; skvar reveals an overestimation of simulated water temperature variability near the German coast."

- P371, line27: "one could argue" one needs to demonstrate, not argue.

The sentence has been rewritten in the following way:

"Having the general good agreement of AMM7 to observations in mind, it leads to the assumption that the coarse resolution of 7 km is not sufficient enough to reflect the highly variable temperature field in this complex area."

- P375, line 17: "in some cases it may be right for the wrong reasons. Agreement

OSD

12, C251–C266, 2015

is achieved when observed salinity happens to be in same range than tidally varying model values; otherwise there is no agreement." Rewrite this!

The section of English East coast position analysis for salinity has been rephrased:

"In the time period of 2009-2012, observations range between 30 and 35, resulting in a mean value of 33.03. Some low salinity events occur below 30, mainly in winter months (not shown). These low salinity events are not entirely reproduced by BSHcmod in 2010 and 2011, resulting in high positive differences. The BSHcmod v4 salinity values range around 33.67 and does not capture the high variability seen in the observations, with the variation mainly showing oscillatory changes as would be expected from water mass movements due to tidal fluctuations in the English coastal waters. AMM7 results starting in April 2011 give salinity values between 30 and 34 (mean at 32.39), with a bias of -0.99. The mean FerryBox salinity for the AMM7 period is 33.38. The skvar for AMM7 is 1.19, which is better than for BSHcmod (0.45). But the IOA is slightly higher for BSHcmod (0.52) than for AMM7 (0.39). Agreement is achieved when observed salinity happens to be in same range than tidally varying model values; otherwise there is no agreement. Concluding, BSHcmod v4 results show too salty water, AMM7 results are too fresh. This is also pictured by the different sign of the cost function (cf) results (negative for BSHcmod v4, positive for AMM7)."

- P376, line 9: "getting the annual cycle and the amplitude in the correct phase in their respective time period resulting in mainly good agreement" how can you tell this from Figure 10?

The figure shows the differences, not the water temperatures itself. It had turned out before, that the picturing of the differences has more benefit instead of showing the annual cycle of water temperatures of models and observations. However, the sentence seems to be misleading, so it was rephrased:

"BSHcmod and AMM7 simulations of water temperatures are in line with observations for most times. Differences in upper left panel show peaks mainly in summer seasons,

OSD

12, C251–C266, 2015

ranging to ± 1.8 K at maximum. The according annual cycles agree to each other (not shown). Agreement is also apparent in statistical measures, shown in Figure 10 upper left). The statistical measures are . . .”

- P377, line 6: “It could be argued, that a second drop has been there, but at a shifted position which could not be detected by the FerryBox.” Pure speculation!

The sentence has been removed.

- P377, line 10: “It is known for example that the AMM7model underestimates flushing in the German Bight” reference?

Unfortunately, there is no peer-reviewed reference available for this finding. Met Office report from 2013 suggests a model residual flow error to be responsible for the flushing underestimation.

- P377, line 16: “freshwater eddies far away from the coasts” what are these? Where do they come from?

The term “freshwater eddy” seems misleading here. Instead we introduce “low-saline water body” as it was also used in Petersen et al., 2011. These low-saline water bodies are coming from the Rhine or from one of the rivers, which enter at the German and Dutch coast. The text has been changed to:

“Moreover, long-persisting low-saline water masses seem to cover only small scales in space and could be missed either by the model or the FerryBox traveling along the route, resulting in higher discrepancies between model and FerryBox. In this context, the different special features of model and FerryBox should be reminded. Whereas the FerryBox samples data of spots along a track, the model covers an area of several kilometres.”

- P378, line 26: “It is not clear why both models should both predict freshening in the summer 2011 which in fact did not occur.” Perhaps a bit more investigation needed??

OSD
12, C251–C266, 2015

In both years, an episode of enhanced river runoff with low-saline water entering the North Sea has been detected by FerryBox and MARNET observations as well as river gauges at Elbe river. The salinity drop in 2011 happened in April, amounting to 30 at the MARNET station. Both models could not predict the low-saline water correctly. In summer, the simulated salinity drops to below 32 while observations show values of around 33. This holds not only for the MARNET position but for the transect in the German Bight east of 6° E. Salinity observations of MARNET are available for two depths – 6m and 30m. In summer, differences between the surface and the bottom show a (thermal induced) stratification, also apparent in salinity. However, the stratification is not stable throughout the summer. Several episodes of mixing are reflected by very low temperature difference between surface and bottom (<0.2 K). An assumption could be that both models overestimate the thermal stratification (and, thus, underestimate vertical mixing) in summer, leading to fresher water masses. This holds for 2010 and 2011.

The text in manuscript has been rewritten containing now parts of the above section.

- P379, line 13: “The spectral densities of each time series are located in the same range”. It would indeed be surprising if they didn’t! what is not discussed it the SLOPE of these spectra, especially in the case of BSH.

The slope of AMM7 spectral densities as well as the MARNET observation spectral densities is found to be in the order of -2/3 as it is common for the inertial range of the energy cascade. However, the spectra of BSHcmod show a descent of -1.2. The overall calculation of the spectra is the same for all three data sets, but the time periods of the models are different and BSHcmod is run only once a day (while AMM7 is run four times). Still, we cannot explain these differences fully. The text has been slightly adapted and extended, even if we are aware, that there are further analyses probably necessary:

“For extending the analysis of temporal variation in observations and model simula-

OSD
12, C251–C266, 2015

tions, the spectral densities for water temperatures of MARNET measurements and BSHcmod v4 and AMM7 simulations have been evaluated. In Figure 12 the power spectra of the three time series are shown. The frequency is shown in s-1. No smoothing has been applied to the spectra. In all spectra, the density peak at the diurnal cycle is present, at the BSHcmod v4 model simulations more sharpened than at MARNET observations and AMM7 simulations. Also the tide peak at 12.43 h is recognizable in every spectrum. The slope of MARNET and AMM7 is found to be in the order of the -2/3 descent in the inertial range of the energy cascade, while for BSHcmod the slope amounts to -1.2. To resolve this discrepancy, further analyses have to be made."

- P381, line 24: Is the Schulz-Stellenfleth & Stanev reference the correct one? (seems to deal with water levels not temperature and salinity).

The Schulz-Stellenfleth reference has been in the text by mistake and has been removed.

Interactive comment on Ocean Sci. Discuss., 12, 355, 2015.