Reviews

We thank the Reviewers for the careful and constructive comments. The suggestions and corrections have greatly improved the quality of this manuscript.

Referee 1: S.M. Hvidegaard

General comments: An important and interesting study. Well written manuscript with few improvements needed. The study still needs some clarifications and justification before publication can be recommended.

Specific comments:

1.1 The manuscript needs more detail regarding the sea-ice/lead classification. Please explain more about the reason for choosing the method of Müller et al 2017 over other procedures using waveform parameters.

In order to clarify the process and to justify the choice of the unsupervised classification, we added the following text to section 2.3.1:

"**Several classification methods have been developed within the last years, which are all based on the analysis of the returned satellite radar echo (e.g. Laxon 2004; Zakharova et. al., 2015; Zygmuntowska et. al., 2013). Most of them use thresholds on one or more parameters of the radar waveforms (e.g. maximum power or backscatter coefficient). In this study, an unsupervised classification approach is applied, which is independent from any training data. This method performed best in a recent study assessing the quality of different classification approaches with respect to very high resolution airborne imagery (Dettmering, et al, 2018).** Briefly summarized…"

1.2 The discussion and use of geoid model to reduce the altimetry derived sea level heights to dynamic ocean topography needs more investigation. As a minimum other geoid models should be included in the analysis e.g the EIGEN, EGM, or Arctic Gravity project geoids or similar.

In fact, the geoid is one important error source for altimetry-derived ocean topography, especially in polar areas where in-situ gravity information is sparse. Within our study, we tested different geoid models but can't find one model superior to all others in the entire study region (see also answer to your next comment). In general, our current study requires high-resolved geoid models with spherical harmonic degrees not smaller than 2190. Thus, we have to use models combining satellite gravimetry data (GRACE/GOCE) with airborne and/or altimetry data sets. Most of the existing combined models use similar "terrestrial" input data sets. For example, the Arctic Gravity project 2008 (ARCGP2008) is part

of EGM2008 (see Pavlis et al. 2012). EGM2008 is used in EIGEN6C4 (over the continents, for higher degrees) and furthermore EIGEN6C4 represents the high spherical harmonic degrees (>719) in OGMOC. However, the incorporated altimetry-derived gravity information changed from SIO/NOAA and DNSC07 for EGM2008 (Pavlis et al., 2012) to DTU10 for EIGEN6C (Shako et al., 2013). In our study, we decided to use the geoid model with the latest combined satellite gravity and altimetry observations, which is the OGMOC. Moreover, this model was developed with special focus on ocean circulation (Fecher T., and Gruber T., 2018).

We added some additional information in section 2.3.3 in order to justify our choice of geoid: "… is applied. **This is one of the latest high resolution global geoid models incorporating the most recent satellite gravimetry and satellite altimetry data sets. Moreover it is optimized for estimating ocean currents and it is assumed to provide the best possible solution for the current application.** More details…"

Shako, R., Förste, C., Abrikosov, O., Bruinsma, S., Marty, J.-C., Lemoine, J.-M., Dahle, C. (2013). EIGEN-6C: A High-Resolution Global Gravity Combination Model Including GOCE Data. Observation of the System Earth from Space - CHAMP, GRACE, GOCE and Future Missions, 155–161. doi:10.1007/978-3-642-32135-1_20

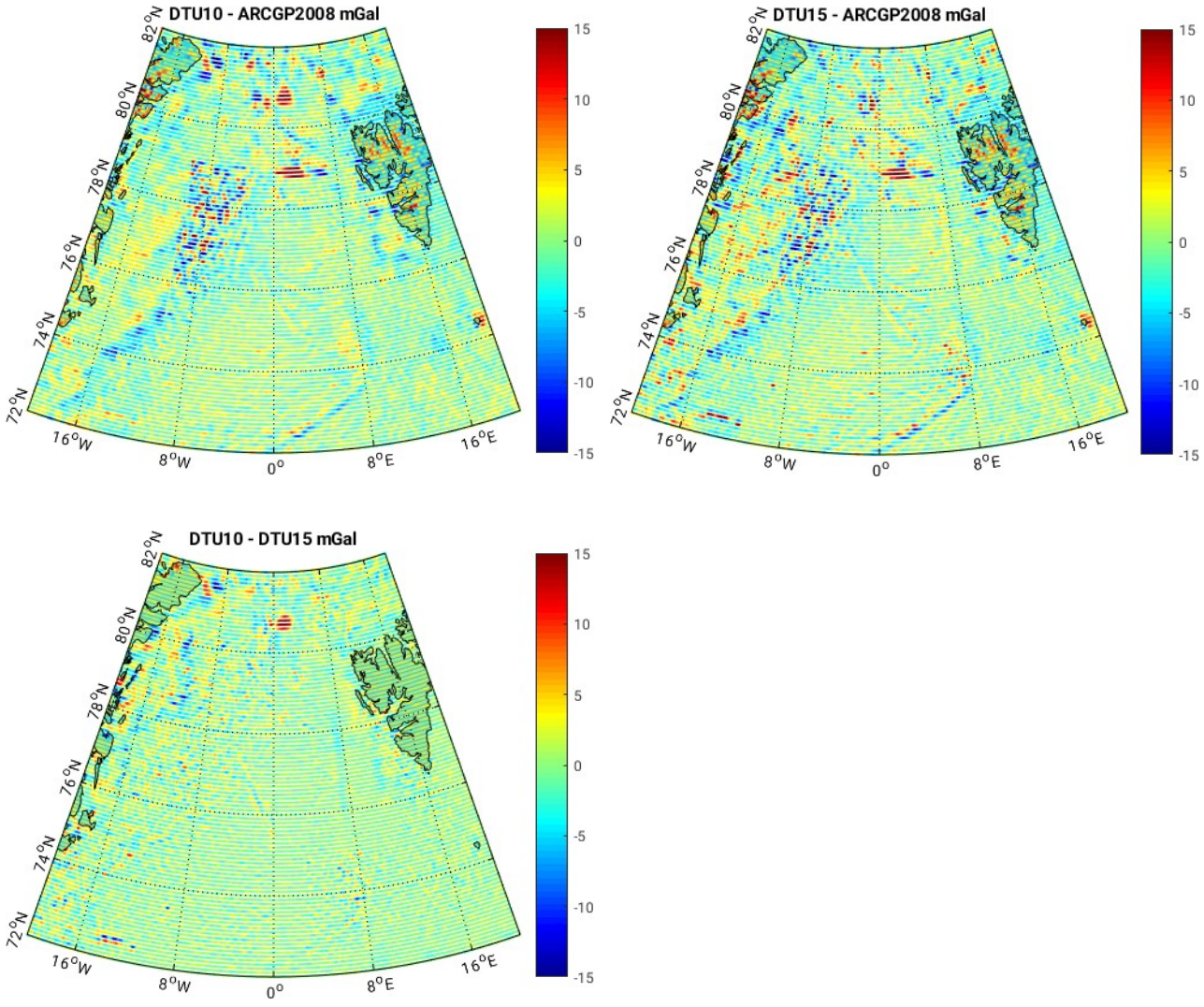
1.3 Also, elaborate more in the statement in line15 p.17. I am not convinced that the pattern is dues to the geoid model but rather the issued with sea ice present in the Envisat footprint and melt-ponds on top of the sea ice in Summer.

Concerning the pattern in the Fram Srait region, we don't think it is directly related to the influence of melt ponds on the altimetry measurements used in this study, since it is constantly present during the entire investigation period (i.e. also in winter when no melt ponds exist). It does not show any temporal variability, indicating a static error source, i.e. the underlying geoid. However, we agree that melt ponds could have an impact on the altimetry data used in the higher harmonic degrees in the geoid model. Hence, the geoid will be less accurate in regions with many melt ponds degrading the altimetry measurements. This is especially true for older altimetry-derived gravity data sets.
The latest altimetry-derived gravity information included in a global geoid model (EIGEN, OGMOC) is the DTU10 (Andersen O. B., 2010). Unfortunately, up-to-now, the most actual data set DTU15 (Andersen O. B., 2017) has not been included in any global geoid model. A comparison between DTU10 and DTU15 reveals clear discrepancies between the two data sets (see Figure below). In particular, a difference of about 15 mgal is visible in the northern Fram Strait region at about 80.8°N and 2°E. This is the location of the artifact we see in the DOT comparison.

Apparently, this artifact is not included in ARCGP. However, this model shows a lot of bathymetry related features not visible in both DTU data sets (e.g. in the East Greenland Current).

Andersen, O. B., The DTU10 Gravity field and Mean sea surface (2010), Second international symposium of the gravity field of the Earth (IGFS2), Fairbanks, Alaska.

Andersen, O.B. & Knudsen, P & Kenyon, S & Factor, J.K. & Holmes, S. (2017). Global gravity field from recent satellites (DTU15) - Arctic improvements. First Break. 35. 37-40. 10.3997/1365-2397.2017022.



Comparison of DTU10, DTU15 and ARCGP2008 geoid model in terms of gravity anomalies. The comparison DTU10 (which is included in EIGEN6C4 and OGMOC) vs. DTU15 shows a significant pattern in the central Fram Strait. A similar comparison DTU10 vs. ARCGP2008 displays also a similar pattern in the same region, in contrast to DTU15 vs. ARCGP2008. This means that DTU10 suffers from bad or insufficient processed altimetry data. DTU15 and ARCGP2008 don't include this artifact. However, DTU15 vs. ARCGP2008 exhibits more obvious differences in the sea-ice zone and at the Greenland shelf edge, than plotted in DTU10 vs. DTU15.

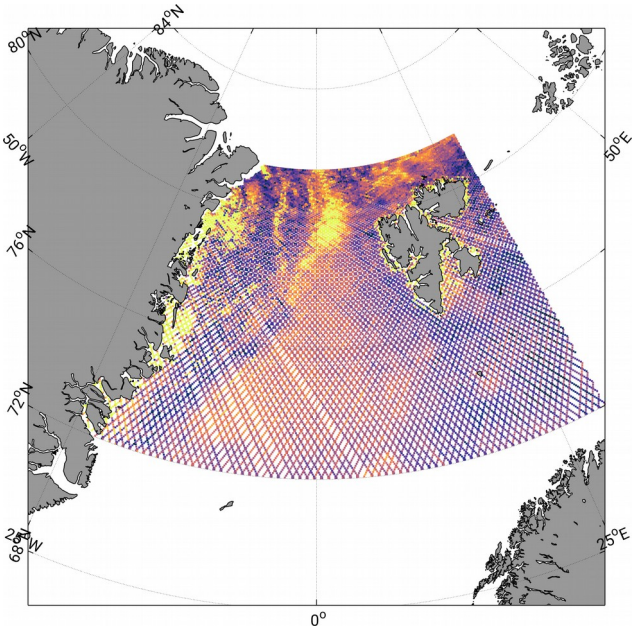
1.4 The discussion of seasonal cycle/mean annual amplitude needs to include more about the uneven sampling from Envisat Altimetry in Summer. How does this affect the study and conclusions?

We refer to Figure 4b of the paper. The figure shows a comparison between the original gridded FESOM and the same data set but interpolated to the altimetry tracks and epochs. Hence, all differences are due to the data sampling and show the impact of the uneven data distribution of Envisat. The comparison indicates an amplitude difference of about ~1cm in the annual signal. It seems that the along-track data sampling leads to an overestimation of the annual amplitude. The effect might be related to alias periods stemming from the repeat cycle of Envisat or from the missing information in specific regions (with sea ice cover) where the amplitude differs from open ocean areas.
However, even if the estimated amplitudes are impacted by the uneven altimetry sampling, the comparison of altimetry and FESOM are not affected by this, since both data sets have exactly the same spatial and temporal sampling (as FESOM is interpolated to the altimetry tracks).

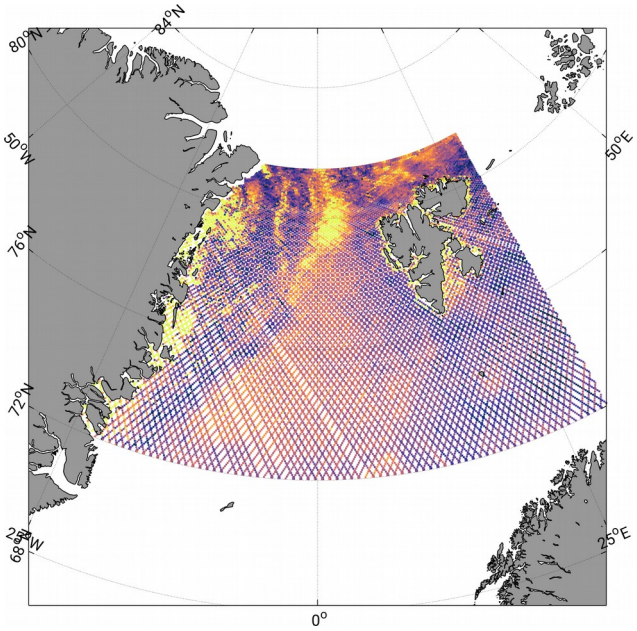
We changed some text passages and included some additional discussion on this in the manuscript (Section 3.1):
"**It can be clearly observed that the three days period is not confirmed by the original data set. Moreover, higher discrepancies can be found in the short periodic domain, which can be addressed to more variability due to more input information. However, all other dominant periods are caught by both data sets. The obtained amplitudes show good agreement in all periods, except for the annual signal. Here, the irregularly sampled profile data overestimates the amplitude by about 1 cm. This might be related to alias effects from remaining tidal influences due to the repeat cycle of Envisat (see section 4 for more details).**"

1.5 The figures 6, 8, and 9 all suffer from the fact that the color intensity is dependent on the change in spacing between satellite tracks with latitude. I suggest to try alternative plots. Also, the plots need titles and numbering (a,b,..) and the phase difference does not make sense with 0-360 degrees (change to +-180 degrees)

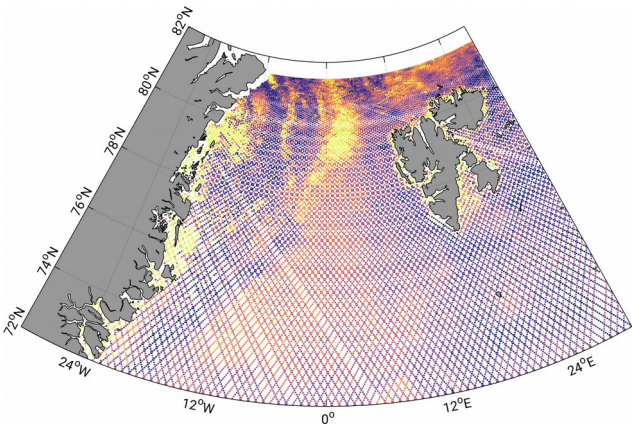
The general problem of the graphical track separation in the higher northern latitudes is due to the turning latitude of Envisat (81.5°N). The tracks are very close together and therefore it seems that the colors are overlapping. A change of the projection wouldn't help, because it doesn't change the circumstance of the very close adjacent tracks. Please find attached some other projection examples (e.g. azimuthal, conic), which can be used in polar areas. Cylindrical projections like the Mercator projection are not suited for high latitudes due to a strong distortion. The Albers Equal-Area projection is used in the present paper.



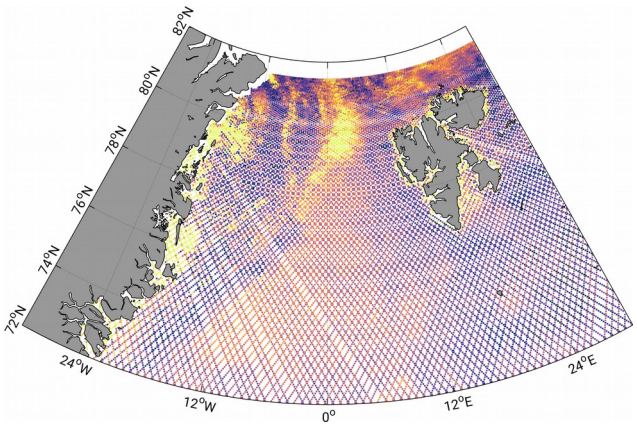
Stereographic (azimuthal)



Azimuthal equal-area (azimuthal)



Lambert Conformal Conic



Albers Equal-Area Conic

We agree that a labeling of the individual sub-plots improves the readability of the paper. We also added the numbering to the caption.
However, please note that the "phase" in Figure 6 is given in days (instead of degree or radiant). We plotted day of year (DOY) ranging from 1 to 365 (Fig 6bd) and days from 0 to 182.5 in case of phase differences. We now add "**in days**" in the caption to avoid any possibility of confusion.

1.6 The discussion lacks an evaluation of the different sampling of the model versus the altimetry. How does this affect the conclusions? What are the advantages and disadvantages of the choices in section 3.

The impact of the different data sampling is partly discussed in section 3.1. See also our answer to your comment 1.4. The direct comparisons within this paper are not affected by the different sampling since the FESOM has been interpolated to the observation points of Envisat in order to allow for a consistent comparison. However, we agree that in principle, the better sampling of FESOM is an advantage when using model data instead of altimetry observations.

We added the following paragraph on the sampling of altimetry data in section 4 (see also comment of Rev2):
"Due to its measurement geometry, satellite altimetry has a high along-track resolution, but data are scattered in time and space. In addition, in polar regions, an irregular sampling due missing data caused by sea-ice coverage must be taken into account. This can significantly influence the estimation of annual sea level variability as tests with simulated data with different sampling revealed (see section 3.1). However, an interpolation of the data set (as it is done by the majority of other studies (e.g. Kwok and Morison (2015), Armitage et al. (2016), Farrell et al. (2012)) could be avoided in order to conserve more high-frequency observations and spectral content."

Technical corrections:

We have considered all comments and changed the text accordingly.

1.7 P 2 l 19 "in spite of the difficult..."
We changed the text, accordingly.

1.8 P 2 l 23 "impression of the model accuracy ..."
We changed the text, accordingly.

1.9 P 2 l 30 "ocean model grids with local refinements in the region of complex ..."
We changed the text, accordingly.

1.10 P 2 l 32 "spatial representation for other regions ..."
We changed the text, accordingly.

1.11P 2 l 33 "It includes, besides..."
We changed the text, accordingly.

1.12 P 2 l 34 "salinity), a sea-ice ..."
We changed the text, accordingly.

1.13 P 3 l 13 "In addition, more ..."
We changed the text, accordingly.

1.14 P 3 l 16 “Figure 1. The study …”
We changed the text, accordingly.

1.15 P 4 Figure1 also add mean sea-ice extent to the plot e.g. from NSIDC or

Our feeling is that including the sea ice concentration in the existing figure will overload the plot. Hence, we provide a second plot showing a mean sea-ice concentration based on monthly NSIDC sea-ice concentration grids within the investigation period (2003-2009).
In addition, please note that Figure 1 has been updated based on a comment of Reviewer 2.

1.16 P 5 l 10 “that the model performed well in simulating …”
We changed the text, accordingly.

1.17 P 6 l 20 “to develop Arctic and …”
We changed the text, accordingly.

1.18 P 6 l 25 … various missions. BUT only one mission is used here; Envisat
This is absolutely correct. Since we don’t combine different missions here, this correction is not necessary for the inter-mission calibration. However, it corrects for a known bias in the Enivsat ranges (of almost half a meter) and thereby improves the absolute level of the altimeter SSH. To clarify this, we changed the text to:
“Furthermore, a mean range bias correction, computed by a …, is included to eliminate a known constant offset in the Envisat range measurements.”

1.20 P 6 l 27 What is meant by “long-temporal”?
In this context “long-temporal” means more than 20 years. We clarified the sentence in the manuscript.

1.21 P 7 l 3 “the high resolution Optimal Geoid …”
We changed the text, accordingly.

1.22 P 9 l 7-19 and Figure 4 could be placed in supplementary material
Based on your points 1.4 and 1.6, we decided to keep this figure and to even enlarge the discussion, since the lower plot of figure 4 nicely demonstrates the influence of different sampling of input data to the amplitude estimation (see also our comments above).

1.23 P 9 l 19 “to more variability due to more …”
We changed the text, accordingly.

1.24 P 14 l 10 “and phases as obtained by …” or rephrase sentence
We changed the text, accordingly.

1.25 P 16 l 5 “,which act as kind of …” rephrase sentence
We changed the text, accordingly.

1.26 P 16 l 12 “areas are significantly more noisy than …”
We changed the text, accordingly.

1.27 P 17 l 23 “The present paper shows basic…”
We changed the text, accordingly.