Review of "A Bayesian approach towards daily pan-Arctic sea ice freeboard estimates from combined CryoSat-2 and Sentinel-3 satellite observations"
By Gregory et al.

General comments

In "A Bayesian approach towards daily pan-Arctic sea ice freeboard estimates from combined CryoSat-2 and Sentinel-3 satellite observations" the authors investigate the use of Bayesian inference to produce daily gridded pan-Arctic radar freeboard estimates. Gaussian Process Regression (GPR) is used to model spatio-temporal covariances between observations made by three ESA's satellite altimetry missions (CryoSat-2, Sentinel-3A, and Sentinel-3B) and to make pan-Arctic predictions of radar freeboard, with uncertainty estimates, on a given day.
This is a novel, interesting and relevant investigation as it attempts, for the first time, to estimate freeboard with a daily temporal resolution based solely on satellite altimetry data. The improved temporal resolution of pan-Arctic freeboard could contribute to our ability to understand physical processes driving sea ice thickness variability on sub-monthly time scales.
The study is generally well structured and the manuscript is clear and pleasant to read. I recommend this paper for publication, however, there are some points that should be addressed by the authors first.

Specific comments

Data

- Why did you choose data between December 2018 and April 2019? By selecting e.g. the following season (2019/20), you could have included in the analysis the months of October and November and make your results representative for an entire Arctic winter season.
- L85-90: Hamming-weighting and zero-padding are both applied to CS2 L0 processing (https://wiki.services.eoportal.org/tiki-download_wiki_attachment.php?attId=4431&page=Cryosat%20Documents&download=y). Please amend this statement and, if CS2 L0 data are processed using GPOD, please state the differences with the official Baseline-D version provided by ESA.

Method

- How do you treat observations from different satellites in the same grid cell acquired on the same day (i.e. co-located in both space and time)? Do you include these as separated inputs or do you feed them as a single averaged estimate to the GPR algorithm? This should be clarified in the manuscript.
- As there is no general "Discussion" section, I add this comment here. Bayesian inference allows to estimate the optimal covariance function hyperparameters based entirely on data as the parameters maximising the log marginal likelihood function. Do you think that the tool you developed could be useful in investigating the spatial and temporal correlation length scales of freeboard measurements? Please add a short paragraph discussing this possibility.

Validation

- How do you think a different grid resolution would affect your results in Section 4, e.g., by using a 25x25 km grid instead? Also, please repeat in the conclusions that the validation presented in Section 4 is based on a 50x50 km grid.
- The results in Table 1 show a slight but systematically lower freeboard mean difference between CS2S3 and S3B compared with CS2S3-CS2 and CS2S3-S3A. While rounding might play a role in this comparison, do you have any idea why CS2S3 tends to best fit S3B data for every month of your analysis?
- I understand the authors' choice of the cross-validation method, however, I think that both section 4.2 and the conclusions should clearly state that the given estimates of prediction error are based only on validation data from regions below 81.5°N and with a sea ice concentration larger than 75%, since these correspond to areas where the absolute uncertainty is usually the lowest (exception made for the Canadian Archipelago and the Fram Strait, as the authors nicely point out in Section 5). Regions above 81.5°N and with ice concentration between 15% and 75% (including the marginal ice zone) are systematically left out of the cross-validation since:
  - only S3 data are used as a validation
  - according to Lawrence et al. (2019a), diffuse waveforms within grid cells with ice concentration lower than 75% are discarded, which means that no freeboard estimates are available from any of the satellites on a given day where ice concentration falls below 75%.
- I would have expected a more significant difference in performance when training the model with CS2 data only, given the lower spatio-temporal coverage when compared with a combined CS2/S3A/S3B training data set. According to your results, a GPR based on CS2 observations alone is able to predict radar freeboard at unobserved locations pretty well (with a 3-4% RMSE increase, from 5.9 to 6.1 cm, when compared to the multi-satellite solution). Do you think this is related to the relatively coarse (50x50 km) grid chosen for your cross-validation? I suggest to add a paragraph in your discussion elaborating on this matter and on the actual advantage of including S3 data in your model training compared with using only CS2 data. In the light of these results, it would also be interesting to discuss the possibility of using data from the three satellites while reducing the number of days used for model training.

Assessment of temporal variability

- This is a nice section highlighting daily variations of regional freeboard estimates and larger discrepancies between CS2S3 predictions and satellite data for sectors like the GIN and the CAA. I suggest to add a couple of statements about the 'Baffin & Hudson' sector. While the average CS2 and S3 freeboard over the entire period agree within 5 mm, they show differences of ~1 cm in December 2018 and March 2019. What do you think might be the reason for this more significant, with respect to other sectors, difference?
- I would rephrase line 253 to reflect that the comparison of mean freeboard estimates over the entire observational period performed in this section is mainly a confirmation of your cross-validation results—the average value of a time series alone does not say a lot about temporal variability. I suggest something like: "… Generally, the mean of the CS2S3 time series lies within 3 mm of CS2 and S3, in line with the results of the cross-validation presented in section 4.2. However, …".

Technical corrections

- L43: according to Lawrence et al. (2019a), the CS2 daily Arctic coverage is lower than 20% up to 82-83°N, not at all latitudes. Also, Tilling et al. (2016) shows Arctic coverage down to a minimum of two days, not one. Please amend this sentence to reflect the content of the cited publications
- L66/378: the DOI provided for Rasmussen and Williams (2006), a book, points to an article by Matthias Seeger with same title. Please correct the reference
- L104: if you want to be consistent with the platform/sensor notation used for the OSI SAF product, this line should perhaps read: "… from the Nimbus-7/SMMR, DMSP/SSM/I, and DMSP/SSMIS, which are …" → (see https://nsidc.org/data/nsidc-0051 for reference)
- L108: you probably mean OSI-403-c? The 403-b product has been superseded and did not include AMSR-2 data
- L138: "For now…" → "For now, …"
- L190: "corresponds" → "correspond"
- L291: add comma after "Greenland" → "… and the Greenland, Iceland and Norwegian Seas, …"
- L301: I suggest not use "K" in the final statement → "… and the fact that the covariance structure can take any form, so long as the covariance matrix is symmetric, positive, and semi-definite, means …"
- Figure 1: please state which day the sea ice concentration, type (FYI/MYI boundary) and radar freeboard refer to in the example
- Figure 3: please add the grid resolution (25x25 km) and the day which the radar freeboard estimates and uncertainty correspond to
- Figure 6: if the benchmark time series is not explained in the caption, please add a reference to the section 5
- Figure 7: please write the name of the sectors in full and provide the abbreviations, when used in the text and/or in Figure 6, in parentheses