First of all, we want to thank the reviewer for the positive feedback and constructive comments on our manuscript. All comments have been taken into account, and a list of answers and actions undertaken is given below. Answers are marked in "blue". Snippets of the updated manuscript are cited using "..."

In this study, the authors have presented work investigating the use of Gaussian Process Regression (GPR) to derive spatially continuous elevation changes over glaciers using ICESat-2 observations. This is a highly relevant and interesting topic, and overall the methods seem to be sound. The manuscript is generally well-written and understandable. I have a few general comments, and a number of specific comments/questions for the authors.

General comments

I think the presentation of the pre-processing steps (§4.1) could be re-arranged or re-worded, to make it clear that you are in effect using three different sets of elevation changes for three different "experiments":

1. Long-term trends using observation pairs separated by at least 20±1 (12±1 for Larsen-B) cycles, so long as the number of cycles separating the two observations is a multiple of 4 - this is the main portion of the results presented and analyzed.
2. Annual height changes, using observation pairs separated by 4±1 cycles.
3. Single cycle (short-term) height changes.

As written, it was not immediately clear that you weren't taking the results of each of these different pairings and using them all in the same GPR regression. Hopefully I have understood this correctly, that you have done the regression using three different ways of slicing the data.

The reviewer understood this correctly. We revised section 4.1 to be clearer:

"Height change information was computed for three different time intervals at each study site: (1) long-term trends: To minimize the impact of seasonal variations, surface height changes were computed for measurements with a time separation by multiples of 4 cycles (~1 year). However, the frequent cloud cover at both study sites forced us to add a temporal buffer of +/- 1 cycle to increase the spatial coverage of height change measurements. Average rates of height change were derived for Svalbard for the total observation period (March 2019 – December 2024, Cycle 3-25). For each ground track point, we obtained a height time series cropped to the earliest and latest valid height measurements. Pairs of observations were selected from each time series based on a preferred separation of 20 repeat cycles (~5 years). If no suitable pair was found, pairs within a range of 19 to 21 cycles were also accepted. At Larsen-B, the long-term trends were derived for December 2020 – December 2024 (Cycle 10 – 25), in order to facilitate a direct comparison with the TanDEM-X derived elevation change information (2021-2024, see Section 4.3.1). Here, a preferred separation of 12 cycles (~3 years) and a buffer of

11-13 cycles was applied; (2) annual height changes: Therefore, subsets from January of the starting year to December of the subsequent year were generated. Pairs of valid height measurements with a preferred separation of 4 cycles and a buffer of +/- 1 cycle were used to estimate the height change rates; (3) seasonal height changes: To evaluate the applicability of our GPR-based approach on seasonal temporal scales, we additionally tested input datasets that used height differences derived from only a single ICESat-2 repeat cycle (91 days)."

My second general comment regards a different sort of interpolation. In this, you compare your GPR regression to two other forms of interpolation, Inverse Distance Weighting (IDW) and nearest neighbor (NN). Did you also consider comparing to the hypsometric interpolation often used for altimetry surveys on glaciers (e.g., Arendt et al., 2006; Johnson et al., 2013; Nilsson et al., 2015)? I think that might be a useful exercise for demonstrating the utility/improvement of your method against a common approach in glaciology

Since our major goal was to obtain spatially resolved elevation change information, we did not apply the hypsometric approach. The hypsometric approach leads to physically meaningless discontinuities because interpolation is done for specific elevation ranges.

I would recommend checking the manuscript carefully for consistency in spelling. For example, Mannerfelt vs Mannerfeld; Moršnevbreen (the version used on NPI maps, for example) vs. Morsjnevbreen, Moršnevbreen, or Morsinevbreen; Matérn vs Matern; Doktorbreen vs Dotorbreen; Liestølbreen vs Liestolbreen; Vallåkrabreen vs. Vallakrabreen, etc. Similarly, double-check the in-text citations and the references list, as there are some inconsistencies.

Sorry for the inconsistencies and spelling errors. We hope that we have corrected all issues

Specific Comments

l. 11: are the ± values here uncertainty or standard deviation?

The +/- values represent uncertainties of the average (region-wide) height changes. We adjusted the wording to clarify that we are talking about the regional average values, and we assume that the reader understands the +/- values as uncertainties, since this kind of notation is commonly used.

" For Larsen-B, GPR-derived average height change rates from 2021–2024 are -0.61 ± 0.02 m a^{-1}, corresponding to a volume loss of 4.55 ± 0.17 km^3 a^{-1} , similar to independent estimates from TanDEM-X. In Svalbard, we observe widespread average thinning (-1.57 ± 0.03 m a^{-1}) and detect clear signals of surging glaciers."

l. 30: say here what the repetition rate (or "cycle", later on) is for ICESat-2.

We added information on the 91-day repetition rate

l. 31-35: could combine these statements into a single statement

We combined both statements:

“Despite its precision, ICESat-2 provides height measurements only at discrete points along its ground tracks, like all altimeters. To overcome this sparse sampling, spatial interpolation is needed to generate continuous fields of height change that are essential for applications such as glacier mass balance assessment and data assimilation (Howat et al., 2008; Nilsson et al., 2015, 2022).”

l. 44: did you explore using slope and aspect as predictors?

We also included slope as a predictor. See extended description and discussion to comment further down.

l. 61: "Spatially distributed uncertainty estimates are provided" - could also mention that this is a feature of kriging

We added this information:

“ Thus, spatially distributed uncertainty estimates are provided that directly quantify the confidence of the interpolation for each location independently, similar to Kriging.”

Fig. 1: label for Morsjnevbreen (Moršnevbreen) could be moved to avoid overlap with Paulabreen

We revised this figure accordingly, removed the overlap, and updated the glacier name.

l. 120: what is the vertical datum for the Svalbard DEM provided by NPI?

I am sorry, but I could not find this information in the metadata: https://data.npolar.no/dataset/dce53a47-c726-4845-85c3-a65b46fe2fea

l. 125: what year(s) do the outlines from Silva et al. correspond to?

We added that the outlines are from 2015.

l. 126-129: was this step not needed for the Svalbard outlines?

Thanks for figuring out this issue. Somehow, we forgot to check the frontal changes at Svalbard, and some glaciers showed considerable frontal changes. In order to be consistent with Larsen-B, we used the min. front position for a recently published dataset of glacier front positions within our study period (2019-2024) (https://zenodo.org/records/19481462). All analyses for Svalbard were re-run. However, the final numbers and conclusions did not really change.

l. 130: previously, you stated December 2024 as your cut-off date - I assume this just means that the data are available until Feb. 2025?

Thanks for pointing this out. In the latest analysis, we used data until December. 2024. We corrected the dates accordingly

l. 133-134: This threshold (2500 m) would remove obvious clouds, but potentially still include lower-lying clouds. Is there a further filtering step applied? Additionally, what is the "fit quality flag"? And, why only use points within glaciers?

We just applied the 2500m threshold here. We added information on the fit quality flag:

“Only points with a fit quality flag of 0, indicating the highest quality of the surface fit, used to generate the surface height timeseries for each point from the different cycles, and within the bounds of the glacier outlines were retained.”

l. 140-145: why the difference in preferred cycle separation for each site?

We used a different time period for Larsen-B to facilitate the direct comparison with the TanDEM-X-derived elevation change information:

“...in order to facilitate a direct comparison with the TanDEM-X derived elevation change information (2021-2024, see Section 4.3.1).

l. 150-167: maybe state the number of points/observations for each filtering step?

We added information on the total number of points for the different filtering steps in Section 4.1. (considering also the suggestion from the other reviewer)

l. 160: is this the absolute deviation from the median of the bin?

Yes per bin. We added this information:

“Outliers were identified per bin as those values exceeding three times the NMAD or having an absolute deviation greater than 5 m a^{-1}. “

l. 171-172: did you use any others? You previously mention slope @ line 153.

Yes, we also tested the slope and mean velocity (see more detailed answer to comment below)

l. 177-182: I understand the point here is that you would like to identify areas impacted by surge by comparing the velocity anomaly to "stable" conditions, but couldn't you also use the contemporary velocity over the time separation of your elevation observations as a predictor for the GPR? I think that Hurkmans et al. (2014) used a mean velocity over several years because that was what was available at the time, but I believe that suitable velocity datasets exist to look at this over a smaller time period (e.g., citations at l. 329-330). Hurkmans et al. 2014 is not included in your references.

We re-considered the approach from Hurkmans et al. (2014) and also tested the mean velocity as an additional predictor. Moreover, we also included a more detailed evaluation of the additional predictor in the revised manuscript. Overall, we concluded that mean velocity and slope did not improve the GPR output (actually, the results were noisier, but the noise averaged out when inspecting regional metrics), but as mentioned in the Introduction, the computational costs increase cubically with the input. Consequently, we decided to skip mean velocity and slope as predictors for further analysis.

We added the following test in the Methods Section:

“Following the approach of Hurkman et al (2014), we also tested glacier surface velocity information as a predictor for the multi-annual datasets. Therefore, we used the average ice flow speeds throughout our observation periods, derived from the ITS_LIVE (Gardner et al., 2025) yearly ice flow composites. Moreover, surface slope information, derived from the reference DEMs, was also tested as a predictor. In order to evaluate the contribution of each predictor (height, glacier IDs, velocity, slope) on the resulting GPR output, we randomly shuffled a specific predictor across all locations on the glaciers and applied the trained GPR model (with all predictors for the ICESat-2 measurements, without shuffling). This approach is similar to the Permutation Feature Importance (PFI) (Breimann, 2001) method used for evaluating machine learning models. The results are exemplarily shown for the region around Doktorbreen and Liestølbreen, Svalbard, in Figure 2. Shuffling of the height and glacier ID predictors led to strong noise in the GPR output (Fig. 2a and d). Whereas shuffling the velocity and slope predictors (Fig. 2b and c) had a less pronounced impact on the GPR output. Consequently, the importance of the latter predictors is lower than for surface height and glacier ID information. Additionally, the experiments with different sets of predictors were carried out using artificial data voids of 7 000 m radius (see Section 4.3.2). The revealed offsets to the actually measured ICESat-2 values showed only small differences for the different predictor sets (mean offsets -0.03 to -0.07 m/a, standard deviations 1.37 to 1.49 m/a). Considering the PFI and artificial data voids analyses and the fact that the computational costs of GPR scale cubically (Section 1) with the input, we selected surface height and glacier IDs as predictors for further analyses.”

l. 185: how was this interpolation done?

We used linear interpolation, which was just mentioned in the next sentence. We also included it here:

“….we calculated the mean height change within 100 m elevation bins, linearly interpolated to all observed elevations, and then removed the linear fit from the height change data.”

l. 190: suggest using space rather than period to indicate thousands (e.g., 11 000)

We revised all numbers accordingly.

l. 197-200: what implication does this have for the smaller glaciers included in the study area?

The correlation length was selected as the best compromise regarding the different glacier sizes (in particular relevant for Svalbard). As mentioned, for different “average” glacier geometries, a larger or smaller correlation length might be more meaningful. We revised this section to be clearer. Moreover, we included a more detailed description of the impact of the correlation length on the GPR output (including a Figure showing different results)

“Different kernels and combinations of kernel functions, as well as correlation lengths for elevation, were tested using the multi-year height change information. The correlation lengths for x and y coordinates were initially estimated using a semivariogram and fitting a spherical variogram model to the “flattened” height change data. At both study sites, similar correlation lengths of ~10 km were obtained. We also tested smaller and larger correlation lengths, as illustrated exemplarily for the region around Doktorbreen and Liestølbreen in Svalbard in Figure 3. A correlation length of 1 km led to rather patchy results (Fig. 3a), whereas a larger correlation length led to some spatial leakage of glaciers with a pronounced elevation change signal to nearby glaciers, even separated by ice-free areas (Fig. 3c). Different kernel functions and combination were also tested (Matérn 5/2 ARD, combination of individual Matérn 5/2 ISO kernel in each dimension, combination of Matérn 5/2 ISO and 3rd order Polynomial kernel, 3rd order Polynomial kernel). The output using the different kernel functions is exemplarily shown in Fig. 4. The Matérn 5/2 ARD kernel showed the most meaningful results, since the combination of four individual Matérn 5/2 ISO kernels or Matérn 5/2 ISO and 3rd order Polynomial kernel led, e.g., to strong spatial leakage of the strong surface lowering in the upper reaches of Liestølbreen to the neighboring glacier to the South (Dobrowolskibreen). The pure 3rd order Polynomial kernel showed a very blurred and averaged output, which is not meaningful at all. The combination of individual kernel functions for each input dimension led to high-frequency noise and discontinuities in the output. Consequently, the selected model and parameter are suitable for glaciers of a kilometer to tens of kilometers in size. However, for smaller mountain glaciers (e.g., like in the Tropical Andes) or large icecaps (e.g., Canadian Arctic), different parameters might be more suitable."

l. 223: wouldn't it be better to do this comparison/correction with the original point measurements, rather than the GPR outputs?

It was done using the point measurements. We rephrased this sentence to be clearer:

An elevation-dependent offset of the TDX measurements was revealed by comparing the TDX and the original ICESat-2 point measurements.

l. 236: I assume this is meant to be Nilsson et al. 2015, which is included in the references list, rather than 2016?

The reviewer is right. We meant Nilsson et al. 2015

l. 245: why nearest-neighbor and not bilinear?

We just implemented two classical approaches out of a variety of different approaches. Inverse Distance Weighting (IDW) and Bilinear interpolation are more comparable as they both consider multiple points and distances for interpolation, unlike Nearest Neighbor, which only uses the closest point. Thus, we selected IDW and NN to test two more "different" approaches.

Fig. 4c), elsewhere: I don't think the units of variance should be m/a.

Thanks for spotting this issue. We adjusted to (m/a)^2.

l. 279: I think this suggests that your extreme change rate filter (described at l. 148) should be higher, if observed thinning reaches -30 m/a?

Thanks for figuring out this issue. The reviewer is right. We carefully checked the ICESat-2 measurements and reevaluated the filter. At the Hektoria-Green-Evans Glacier system, a few measurements up to ~ 36 m/a were filtered out in the yearly fields from 2022-2023 and 2023-2024. Now, we are using a 40m/a threshold for the yearly and 1-cycle datasets. For the long-term dataset, the 30 m/a filter is adequate. We re-run the corresponding analysis. However, numbers did not change a lot, except for some extreme surface lowering values on Hektoria-Green-Evans Glaicer system:

"For each dataset, observations with extreme change rates (> ±30 m/year for the long-term and > ±40 m/year for the yearly and seasonal datasets) were removed to reduce outliers."

l. 308-311: indicate which glaciers match which pattern/style of propagation

We updated Figure 1 to also indicate the surge type.

l. 381-382: this also seems to be the case for Hektoria Glacier, but at much lower elevations - why might that be?

We are sorry, but we do not understand which part of Hektoria Glacier is meant. Anyway, the coverage by ICESat-2 samples changed throughout the review process (see answer to comment on the 30 m/a filter)

Fig. 7: could also include a panel showing a histogram of the offsets/differences, with mean/median, nmad, etc.

We included a histogram in the Figure showing the distribution of the offsets and mean and median values.

l. 395: "offsets peak" rather than "offset peak"

corrected

l. 399: remove "revealed" (repeated)

corrected

l. 402: volume rather than volumen

corrected

l. 415-417: how does this compare to the difference between the GPR output and the observed dh/dt values from the "main" result? That is, if you look at the offset between the GPR output and the original gridded ICESat-2 observations, is the difference observed here similar or greater than the difference between the GPR output and the original observations?

GPR is reproducing the original data where it got measurements. Thus, there is no offset, making the suggested comparison not possible

l. 432: I don't see these plots in the Supplement - only the plots for the annual and single cycle GPR outputs.

We are sorry. Somehow, the plots were not included in the supplement. We updated the supplement. Now, the plots should be included.

Fig. 9: suggest using different symbols as well as color here

According to the reviewer's advice, we revised the color and symbols of the plots. We hope the figure is now clearer.

l. 449: remove figure caption from cross-reference

Thanks for pointing out this issue. Somehow, the cross-reference option included the figure caption

References

Arendt, A. A., Echelmeyer, K. A., Harrison, W. D., Lingle, C. S., Zirnheld, S. L., Valentine, V. B., Ritchie, J. B., and Druckenmiller, M.: Updated estimates of glacier volume changes in the western Chugach Mountains, Alaska, and a comparison of regional extrapolation methods, Journal of Geophysical Research, 111, https://doi.org/10.1029/2005JF000436, 2006.

Hurkmans, R. T. W. L., Bamber, J. L., Davis, C. H., Joughin, I. R., Khvorostovsky, K. S., Smith, B. S., and Schoen, N.: Time-evolving mass loss of the Greenland Ice Sheet from satellite altimetry, The Cryosphere, 8, 1725–1740, https://doi.org/10.5194/tc-8-1725-2014, 2014.

Johnson, A. J., Larsen, C. F., Murphy, N., Arendt, A. A., and Zirnheld, S. L.: Mass balance in the Glacier Bay area of Alaska, USA, and British Columbia, Canada, 1995-2011, using airborne laser altimetry, Journal of Glaciology, 59, 632–648, https://doi.org/10.3189/2013JoG12J101, 2013.

Nilsson, J., Sørensen, L. S., Barletta, V. R., and Forsberg, R.: Mass changes in Arctic ice caps and glaciers: Implications of regionalizing elevation changes, The Cryosphere, 9, 139–150, https://doi.org/10.5194/tc-9-139-2015, 2015.