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 “...”

This manuscript by Seehaus and co-authors presents an application of Gaussian Process Regression (GPR) to interpolate ICESat-2 derived glacier surface elevation change (dh/dt) across two glacierized regions characterized by complex terrain and heterogeneous glacier dynamics: the Larsen-B embayment on the Antarctic Peninsula and central southern Svalbard.

The topic is timely and relevant for the cryosphere community. Satellite altimetry provides highly accurate but spatially sparse measurements, and developing robust approaches to reconstruct spatially continuous elevation change fields remains an important challenge. The use of a probabilistic interpolation framework, such as GPR, is well motivated, and the manuscript demonstrates that the approach can reproduce spatial patterns of glacier thinning and thickening, including signals associated with glacier surge activity.

Overall, the manuscript is clearly written and well structured. The use of two contrasting study regions strengthens the evaluation of the approach, and the sensitivity experiments provide useful insight into the effects of data gaps and terrain complexity. The figures effectively illustrate the spatial patterns of elevation change and interpolation uncertainty. However, several methodological aspects would benefit from clarification or further discussion.

Major Comments:

1. Justification of GRP Kernel

The manuscript adopts a Matérn 5/2 kernel with correlation lengths derived from semivariogram analysis. While this choice appears reasonable, the justification for the specific kernel and parameter values remains somewhat qualitative. It would strengthen the study's methodological rigor if the authors clarified whether kernel hyperparameters were optimized using marginal likelihood within the GP framework or fixed based on the semivariogram analysis. In particular, expanding the discussion of alternative kernel testing and kernel/correlation-length combinations would strengthen the justification for choosing this kernel.

We revised this section (4.2) and included a more detailed description of the impact of the correlation length and kernel selection on the GPR output (including Figures showing different results). The correlation lengths were not optimized and kept constant; only the hyperparameters were optimized.

“GP hyperparameters were optimized using maximum likelihood estimation, and a fixed observation noise variance of log(0.24) was used.”

“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.”

2. Use of Glacier ID as a Predictor

Including the glacier ID as a predictor variable is an interesting approach to reduce spatial leakage between adjacent glaciers. While the examples in the manuscript demonstrate that this approach reduces leakage in practice, the inclusion of this variable also raises some questions. Glacier ID is a categorical variable, rather than a continuous physical variable, like the other predictors. Treating the ID as a numerical feature in the covariance kernel may introduce artificial relationships between glaciers. While this may not be a major concern, it would be helpful for the authors to discuss the statistical implications of this choice and whether alternative approaches were considered.

In order to account for the fact that the glacier ID is a categorical (in our case, it is actually an integer), we applied a correlation length of “1” to minimize cross-correlations between different glaciers due to similar IDs. We added this information:

“ Height change patterns between neighboring glaciers can vary strongly, e.g., glacier surge. To account for the individual glacier dynamics and heterogeneous coverage of the glaciers by measurements, the glacier IDs (integer values) were included as a predictor, with a correlation length of 1 to minimize cross-correlations between different glaciers.”

Moreover, we included the analysis of different predictor sets on the GPR output, where we conclude that glacier ID and height are the most important predictors and thus used for further analysis.

“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."

3. Spatial Correlation Length for Smaller Glaciers

The spatial correlation length used in the model (~11 km) appears relatively large compared to the characteristic scale of many glaciers in the study regions, particularly in Svalbard, where surge-related signals and terminus dynamics may occur over much shorter distances. A brief discussion of how this correlation length affects the model's ability to capture sharp spatial gradients in dh/dt would be useful. In particular, the authors may wish to comment on whether the chosen correlation length could lead to smoothing of localized elevation-change signals.

See the answer to the comment above regarding the different kernels.

4. Artificial Gap Experiments

The validation strategy combines comparisons with TanDEM-X elevation change fields and artificial data-gap experiments. These tests are helpful and provide useful insight into the interpolation performance. The artificial gaps used in the sensitivity analysis are circular. whereas real ICESat-2 sampling gaps are typically elongated and aligned with satellite ground tracks. The authors may wish to briefly discuss whether this difference could influence the interpretation of the gap-filling experiments.

We understand the reviewer's concerns. However, our intention was to simulate gaps due to, e.g., cloud cover, and thus a circular "gap/void" seems to be more realistic to us. We added this information in section 4.3.2.

"In order to simulate data voids due to cloud cover, we cropped circular gaps of increasing radii…."

The real ICESat-2 sampling gaps are actually rhombic. Thus, a circular gap is at least a rough approximation, which is rather easy to implement and to scale.

Minor Comments:

1. The detailed discussion of the pre-processing of the ICESat-2 data is well-structured and informative. However, it is difficult to determine exactly how many ICESat-2 observations were used in each region. The addition of a summary sentence with this information (or a table in section 4.1) would aid the readability of the manuscript.

According to the reviewer's suggestion, we added the number of ICESat-2 observations in Section 4.1

2. When comparing GPR results with the ATL15 product, the authors should note that ATL15 is a coarser-resolution product that applies spatial smoothing for regional elevation change estimates.

We accounted for the different spatial resolutions. In Section 5.2.4. We state at the beginning that we resampled our observations to the ATL15 resolution of 1 km.

"Figure 11 illustrates the long-term elevation change rates using ATL15 and the difference between our GPR-derived fields (resampled to 1 km resolution) at both study sites. "

3. Several typographic errors are present throughout the manuscript. Some are noted here, but the manuscript would benefit from detailed proofreading.
   We carried out a detailed proofreading after the revision of the paper

   Figure 1 (inconsistent capitalization in the legends)
   We revised the legends accordingly.
   L. 233 & 390 (study sides rather than study sites)
   "study side" is the right wording; it refers to a specific location or place where a study or research is conducted.
   4.2 Heading (Gaussing rather than Gaussian)
   Corrected
   L. 396 (acounts rather than accounts)

   Corrected