Point-to-Point Responses to RC1

Comment 1: *This paper describes an interesting and potentially very useful methodology for realistic interpolation of sub-glacial topography (one of many potential applications). Overall I think that this is solid and important contribution to the field. The methodology is quite complex, however, with many steps involved. The text is a bit dense with undefined jargon, and I feel the authors could do a much better job at explaining these steps, and particularly in explaining basic concepts. For example, I was never sure what the authors meant by "distance" between two training images, and that set me at a big disadvantage in trying to comprehend the rest of the methodology. Another example: the authors never define how the values of MDS1 or MDS2, key parameters in the methodology, are determined. There are many more such examples noted in my marked-up pdf file.*

Response 1:

First of all, we would like to thank Dr. John Goff for the review and detailed feedback. The comments play a constructive role in improving the quality of our manuscript. With the aim of better explaining the basic concepts and components, we revise numerous sentences in our paper (revisions are highlighted in red). Furthermore, comments in the supplement file are responded point-by-point. The modifications are listed in the following.

Comment 2: *I hate sounding like the aggrieved reviewer, but really, the authors scant mention of my own paper on the conditional simulation of nearly the same data set had me at a loss. The two methods are extremely different, but the ultimate product and goals are identical in trying to produce a realistically rough surface conditioned on existing radar soundings and accounting for a high degree of spatial heterogeneity. Of particular note, my method spent a considerable effort on ensuring the continuity of fjord-like channels beneath the glacier, which are obviously very important factors in flow simulation and likewise are poorly reproduced by standard interpolation schemes like kriging. How does this method perform in that measure? I suspect it actually does quite well – that the highest probability deglaciated terrain training images do a good job in conditioning the data interpolation to that geometry. But the authors do not explore that property. The authors also did not do a good enough job distinguishing the superiority of their method over SGSIM. The latter images actually looked quite good.*

Response 2:

Comparing a new method with existing works is an important component in the scientific research. With the aim of accurately describing the study performed by Goff et al. (2014), we rewrite the sentence. The modified one is shown below:

"Goff et al. (2014) conducted a conditional simulation of Thwaites Glacier. To improve the modeling quality, the channelized structures and the abrupt between lowland and highland are individually handled. The method has the advantage to ensure the continuity of fjord-like channels beneath the glacier"

Next, we examine our multiple-point statistics (MPS) realization, sequential Gaussian simulation (SGSIM) realization and kriging realization in the channelized region. The maps are shown below. Based on the visual inspection, it is clear that the MPS realization exhibits comparable structures to the deglaciated areas in Arctic and Antarctica.

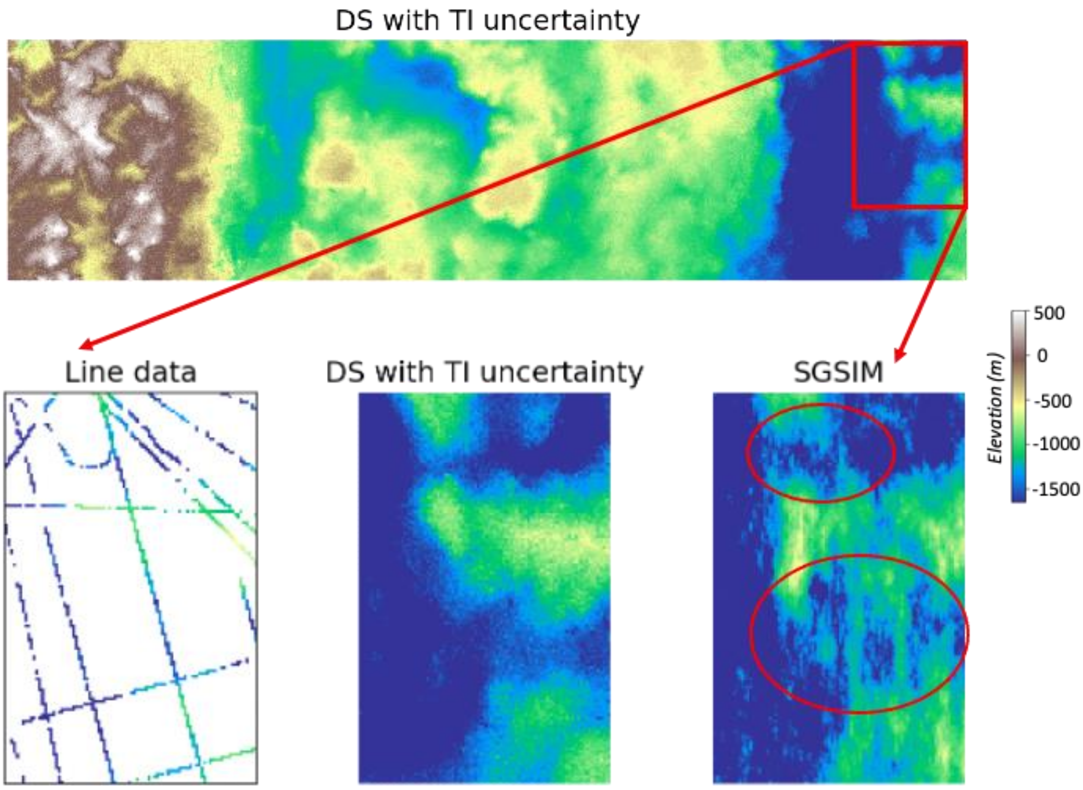
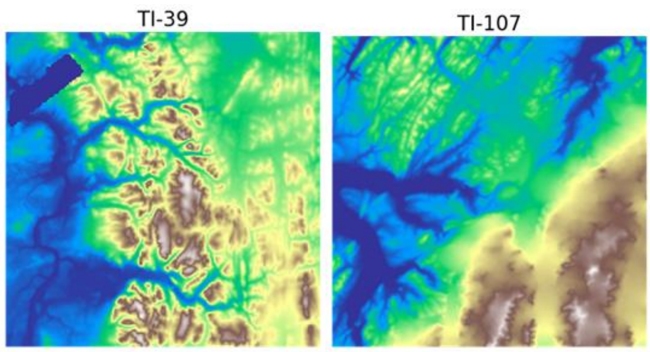


Figure 14. Comparison of SGSIM and proposed DS with uncertain TI selection in the local sparse lines area. Red circles highlight the areas where SGSIM failed to simulate meaningful channels.



Training images provided to our MPS simulation in the channelized area

Comment 3: *As noted in my returned pdf, the figures and captions could use some work. A few of the issues: A lot of the training images were just reduced from larger versions, meaning that the annotation was too small to read. On several images the color white is used both to indicate areas of no data and Z values >500 m. This ambiguity needs to be resolved. Many of the captions were far too brief and failed to explain what is going on in the figure.*

Response 3:

Thanks for these suggestions. We added more captions and relevant contents in our manuscript to better explain the figures. We also revised the figures to separate areas with non-data (non-study area).

Page 2 Line 43:

Sentence: The objective of geostatistical simulation is to generate multiple realizations of phenomena that reproduce the spatial variability of observations, as modeled by variogram or spatial covariance and can be used to quantify uncertainty.

Comment: Not the only objective. Also to generate >>realistically rough<< fields.

Response:

. The target of geostatistical simulation is twofold: realistic structure and uncertainty. Therefore, the sentence in our paper is revised as:

"The objective of two-point geostatistical simulation is to not only reproduce structures modeled by variogram or covariance but also generate multiple realizations to express uncertainty".

Page 2 Line 45:

Sentence: Thwaites Glacier has previously been simulated by Goff et al. (2014), though only one realization was generated.

Comment: A lot more could be said here! My algorithm optimized for continuity along channel forms. Does this do equally as well? I also incorporated highly non-stationary statistical behavior. And I only generated one realization for the paper, but could have generated infinite others.

Response:

. We have modified the description as follows:

"Goff et al. (2014) conducted a conditional simulation of Thwaites Glacier with a geostatistical framework. To improve the modeling quality, the channelized structures and the abrupt between lowland and highland are individually handled. The method has the advantage to ensure the continuity of fjord-like channels beneath the glacier, which is an important factor in flow simulation."

Page 2 Line 29:

Sentence: However, spatial variation over very large areas is inherently non-stationary.

Comment: This term is more appropriate for temporal fields. For spatial fields, the better term is statistically heterogeneous.

Response:

Thanks for this suggestion. In this paper, one of our focused problems is spatial non-stationarity, not heterogeneity. Heterogeneity and non-stationarity are different, at least as is traditional used in geostatistics, for example in subsurface applications. Heterogeneity is opposite to homogeneity, which mean no spatial variation. In non-stationary, the spatial correlation structure itself varies in space, which need not be the case for heterogeneity. This is a common problem when modeling very large area such as the West Antarctica. We cannot use the same TIs for non-stationary (see Figure 11) for non-stationary, otherwise the model will not honor the spatial correlations between the local subareas. Therefore, we will keep “non-stationary”

Page 2 Line 56:

Sentence: Regardless, all these approaches are limited in expressing non-stationary in terms of a mean or covariance function only.

Comment: Again - my method did incorporate a high degree of statistical heterogeneity. A little credit is due!

Response:

Thanks for this comment. We have revised this paragraph in the paper revision and highlighted the contribution of the reviewer’s paper.

Page 3 Line 74:

Sentence: We review three categories of approaches to build non-stationary geospatial models using MPS.

Comment: This paragraph is too detailed for introduction. Includes concepts not yet introduced (e.g., "ad hoc weighting".

Response:

We appreciate the reviewer for this advice. In order to keep concise, we added more explanations and reorganize several sentences. The modified paragraph is as follows:

“We briefly review three categories of approaches to build non-stationary geospatial models using MPS. The first way is to divide non-stationary TI or simulation grid into several stationary subareas. Each stationary simulation area has its specified stationary TI (Honarkhah and Caers, 2012; Strebelle, 2002; Wu et al., 2008; Zhou et al., 2014). But the zonation brings difficulties to create a smooth transition between subareas. Therefore, the second third way is most commonly used. It incorporates spatially continuous non-stationary maps (named as “auxiliary variables”) with weighted aggregation or so-called “ad-hoc weighting” (Chugunova and Hu, 2008; Mariethoz et al., 2010; Oriani et al., 2014; Zuo et al., 2020). Such auxiliary variables determine which TI patterns should fill which location in the simulation domain in a spatially smooth manner. The limitation is that the ad-hoc weights do not scale to the complexity of bed topography. The determination of weights is also subjective. More importantly, auxiliary variables are very difficult to obtain in subglacial topographic modeling. Another challenge in the non-stationary modeling is how to choose training images (Tahmasebi, 2018). This is particularly important as the MPS modeling relies on the spatial information provided by the training images. Different from the above two methods, Hoffimann et al. (2019) introduced an approach to generate time-series training images to model the spatial and temporal evolutions of geomorphology, which is similar to Pirot et al. (2014, 2015). A training image transitional model in time was proposed to reproduce the nonstationary geomorphologic evolutions. However, in subglacial topographic modelling, there are no available training images because subglacial topographic measurements are only made along flight lines. Satellite-based observations from deglaciated areas in the Arctic offer a potential source of training imagery. However, training images retrieved from the Arctic would be logically non-stationary due to the natural variability of the landscape. Furthermore, the Arctic provides a vast amount of deglaciated topographic data, which presents a significant computational burden on MPS simulation algorithms. We therefore will need a strategy to explicitly specify which training images or patterns should go where when filling the radar line gaps"

Page 3 Line 76:

Sentence: Then MPS then uses different divided TI patterns to fill different locations in the simulation domain.

Comment: delete 'then'

Response:

Thanks for the careful check. The redundant word is removed.

Page 3 Line 77:

Sentence: But the zonation in either the simulation domain or training images can make it difficult to maintain smooth transitions between the modeling patterns.

Comment: However,

Response:

Thanks for the careful check. We apply the word 'However' instead of 'But'.

"However, one limitation of the partitioning strategy is discontinuity. The zonation brings a difficulty to create a smooth transition between subareas."

Page 3 Line 96:

Sentence: In this paper, we generalize a geospatial modeling framework to fill irregular geophysical data gaps in large areas.

Comment: Again - too much detail in this paragraph for introduction.

Response: In order to better explain our work, we have shortened this paragraph as follows:

"In this paper, we generalize a geospatial modeling framework to fill irregular geophysical data gaps in large areas. We will use MPS to address the non-stationary topographic modeling by probabilistically selecting non-stationary training images. We first collect a large amount of TIs from the deglaciated areas in the Arctic and Antarctica. Then we will develop a probability aggregation method to estimate each TI's probability of being assigned to different local radar lines. Such probabilistic TI selection scheme will avoid the use of auxiliary variables with arbitrary ad-hoc weightings. We will demonstrate our method using the entire Amundsen Sea Embayment in West Antarctica. This region has alternating areas of sparse and dense measurements with a variety of radar line orientations. We show that the training image sampling process accommodates a range of data configurations. It will generate realistic non-stationary topographic realizations that reflect the subglacial topographic uncertainty in ASE. In addition, we will use the topographic realizations to model subglacial hydrologic flow. The impact of topographic uncertainty on hydrologic uncertainty is further investigated. ”

Page 4 Line 101:

Sentence: The posterior TI probability will be calculated using kernel density estimation conditioned to the actual radar line observations.

Comment: example of too much detail/jargon for intro

Response: With the intention of improving the quality, we remove several the technical details and jargon in the introduction section.

Page 4 Line 103:

Sentence: We will demonstrate our method using the entire Amundsen Sea Embayment in West Antarctica.

Comment: (ASE)

Response:

We are grateful for the careful inspection. The ignored abbreviation is added.

“We will demonstrate our method using the entire Amundsen Sea Embayment (ASE) in West Antarctica.”

Page 4 Line 113:

Sentence: The data is gridded at a 500-meter resolution (Figure 1).

Comment: are

Response:

Thanks for the check. This sentence is refined as:

“The data are gridded at a 500-meter spatial resolution with the nearest neighbor strategy (Figure 1).”

Page 4 Line 115:

Sentence: The swath bathymetry data (Arndt et al., 2013) and subglacial swath radar data (Holschuh et al., 2020) (provide some training imagery.

Comment: delete

Response:

We check this sentence and the new one is in the following.

“The swath bathymetry data (Arndt et al., 2013) and subglacial swath radar data (Holschuh et al., 2020) provide some training imagery.”

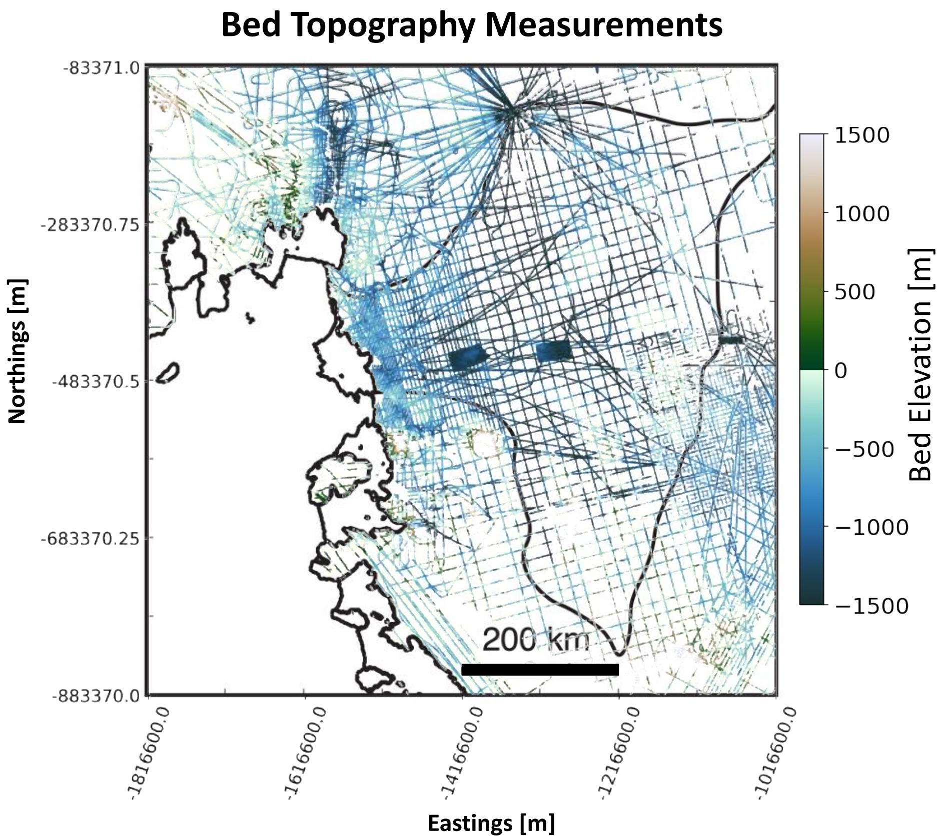
Page 5 Line 126:

Sentence: Figure 1.

Comment: What are the X and Y coordinates?

Response:

We have added coordinates to the Figure 1.



Page 6 Line 132:

Sentence: Figure 2. (a, b) Geographical locations of the 166 training images

Comment: It's not really clear where these 166 patches are located, exactly.

Response:

Again, sorry, we do not have the exact coordinate locations in our dataset.

Page 6 Line 132:

Sentence: (c) example of the 166 training images

Comment: These are in need of horizontal scale.

Response:

Horizontal scale is added to the revision.

Page 7 Line 149:

Sentence: The algorithm used in this work is Direct Sampling (DS) (Mariethoz et al., 2010b; Mariethoz and Renard, 2010), which will be introduced in section 0.

Comment: ??

Response:

The sentence is rephrased as follows:

“The algorithm used in this work is Direct Sampling (DS) (Mariethoz et al., 2010b; Mariethoz and Renard, 2010), which will be introduced in the following section.”

Page 8 Line 174:

Sentence: Based on the explanation above, there are mainly three important parameters in DS

Comment: Enumerate (1), (2), and (3) rather than "one", "another", and "third".

Response:

We rewrote this paragraph, and the new version is in the following:

“Based on the explanation above, there are mainly three key parameters within DS. (1) the number of conditioning points n. In a continuous simulation scenario, $n \geq 30$ is suggested to extract complex patterns from TI as well as the simulation grid (Bruna et al., 2019; Meerschman et al., 2013). (2) the distance threshold t. Because a conditioning pattern of big size is applied, it is possible that there is no completely matching structure in TI. Therefore, the program would accept a training pattern whose distance with the conditioning pattern is lower than t. When many suitable patterns exist in TI, the first pattern found by the searching program is suggested. The value of t has a significant influence on the DS performance. A small value could improve the modeling quality while bring a computational burden. In the most cases, $t = 0.1$ is generally recognized as the upper bound (Meerschman et al., 2013; Zuo et al, 2020). (3) the fraction of scanned TI f. Repeated morphological structures can be common in TI. With the aim of saving time, we can scan only a fraction of TI. For example, $f = 0.1$ implies that the computer only inspects 10% TI. According to the investigation conducted by Mariethoz and Caers (2014)., a recommended value of f ranges from 0.1 to 0.5.”

Page 8 Line 180:

Sentence: *t* = 0.1 is generally recognized as the upper bound in the most cases (Meerschman et al., 2013; Zuo et al, 2020)

Comment: Is t dimensionless? If so, how is it normalized to the dimensions of the grid? A value of t = 0.1 doesn't have any meaning to me.

Response:

t is a user-defined threshold to determine whether a training pattern is accepted. Thus, t is dimensionless. In this paper, we apply a normalized Euclidean distance to measure the similarity between two patterns. Figure 3 provides a simplified example to explain the pattern searching program. It should be noted that the maximum and minimum values in TI are 99 and 0, respectively. As Figure 3(b) displays, a conditioning pattern $P^1_{n=3} = (86,80,37)$ with three known points is checked. Then, the computer launches a random searching procedure. As shown in Figure 3(c), a pattern $P^2_{n=3} = (21,7,30)$ is created. The distance metric is as follows:

$$dis(P^1_{n=3}, P^2_{n=3}) = \sqrt{\frac{1}{3}\left(\frac{(86-21)^2}{(99-0)^2} + \frac{(80-7)^2}{(99-0)^2} + \frac{(37-30)^2}{(99-0)^2}\right)} \approx 0.57$$

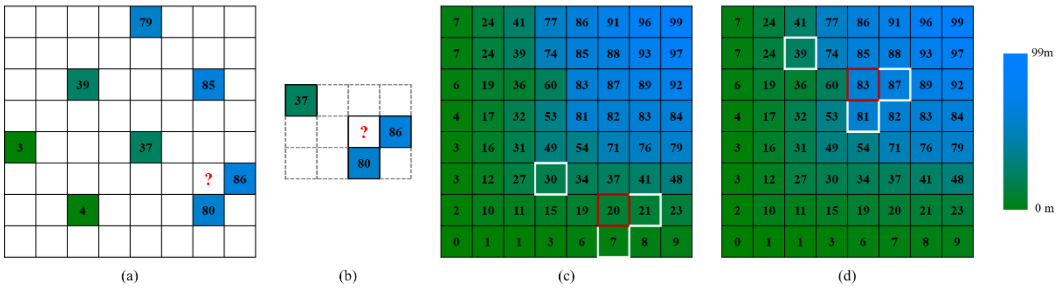


Figure 3. A conceptual example of the DS point simulation. (a) Radar lines on the simulation grid; (b) Three known points (value: 37, 80, 86) constitute a conditioning data pattern; (c) A mismatch pattern in TI; (d) A similar pattern in TI.

Because the distance is much larger than the threshold, the program has to test another point in TI. As Figure 3(d), a pattern $P^3_{n=3} = (87,81,39)$ is found. The distance between $P^1_{n=3}$ and $P^3_{n=3}$ is 0.01. Consequently, $P^3_{n=3}$ is output as the searching result. The program assigns the value 83 to the simulating point.

The detailed explanation about the pattern distance can be found in Mariethoz et al, 2010 and Zuo et al, 2020.

Page 8 Line 181:

Sentence: The third main DS parameter f is the fraction of scanned TI.

Comment: This needs more explanation. fraction of what?

Response:

Thanks for this comment. We add more description in the new paragraph. The explanation is listed below:

"(3) the fraction of scanned TI f. Repeated morphological structures can be common in TI. With the aim of saving time, we can scan only a fraction of TI. For example, $f = 0.1$ implies that the computer only inspects 10% TI. According to the investigation conducted by Mariethoz and Caers (2014)., a recommended value of f ranges from 0.1 to 0.5."

Page 8 Line 181:

Sentence: With the intention of saving time and avoid verbatim copy, an recommended value of f is between 0.1 and 0.5 (Mariethoz and Caers, 2014)

Comment: This is lazy. You can avoid verbatim copy by summarizing.

Response: The redundant words are removed. We exhibit the modified sentence below.

"According to the investigation conducted by Mariethoz and Caers (2014), a recommended value of f ranges from 0.1 to 0.5."

Page 9 Line 193:

Sentence: Then, like other MPS approaches such as SNESIM (Strebelle, 2002) and DISPAT (Honarkhah and Caers, 2010), we extract the spatial patterns from each TI with a fixed template. We then use the classical agglomerative hierarchical clustering (Romary et al., 2015) to divide the spatial patterns of each TI into a finite number of groups. The group number in agglomerative hierarchical clustering is determined by a distance threshold (between the clustered groups). We referred to the commonly used distance threshold in DS approach to set it as 0.1 (Meerschman et al., 2013) of the maximum pattern distances of the TI. The TI with more complex spatial patterns will therefore have more clustered groups. The medoid pattern of each group is taken as the representative pattern of that group

Comment: find this text to be incomprehensible. far too much unexplained jargon. Figure 4 provides no help given the lack of explanation. You really need to match up this explanation with the figure.

Response:

In order to better explain our method, we have re-written this paragraph by removing un-necessary jargons:

“Then, like other MPS approaches (Honarkhah and Caers, 2010; Strebelle, 2002), we extract all the spatial patterns from each TI with a fixed template. Next, we use the classical agglomerative hierarchical clustering (Romary et al., 2015) method to divide the spatial patterns of each TI into a finite number of groups. The number of groups is determined by a distance threshold between the clustered groups in agglomerative hierarchical clustering. As mentioned in Section 3.1.2, we set the distance threshold as 0.1 since it is commonly used to distinguish two patterns in DS (Meerschman et al., 2013). Therefore, TI with more complex spatial patterns will have more clustered groups, thus more representative patterns. The medoid pattern of each group is taken as the representative pattern of the TI.

Moreover, the diversified caption of Figure 4 is shown in the following.

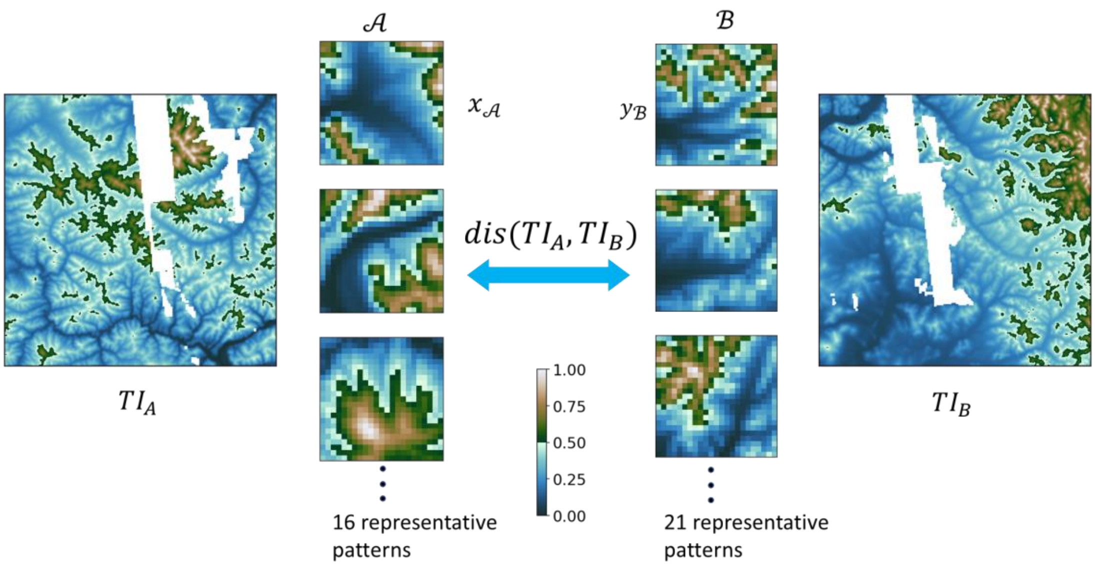


Figure 4. Calculating the distance between any two training images (TI_A and TI_B) using modified Hausdorff distance. There are three key steps: (1) Extracts training patterns with a fixed template. (2) The representatives are selected by a hierarchical clustering method. In this example, the computer found 16 important patterns from TI_A and 21 patterns are from TI_B. The number of representatives is dependent on the complexity of morphology. (3) Calculates the modified Hausdorff distance between two pattern sets. The output distance becomes an indicator of similarity between two TIs.

Page 9 Line 199:

Sentence: Figure 4 shows a few representative patterns. The distance used in the clustering is the normalized Euclidean distance.

Comment: This needs to be defined.

Response: Thanks for this suggestion. The definition of the normalized Euclidean distance is an important concept in our paper. With the purpose of making the paper concise, we provide a reference here:

“In this case, we apply the normalized Euclidean distance (Mariethoz et al. 2010) as the metric.”

Page 9 Line 202:

Sentence: Figure 4. Calculating the distance between any two training images using modified Hausdorff distance.

Comment: This caption is inadequate to explain what is going on in the image. I really can't figure out what is happening. What are the individual patterns in columns A and B? Also, the annotations in the A and B column are far too small to read.

Response:

We have added more detailed explanations for this caption.

Page 9 Line 204:

Sentence: After clustering and medoid selection, training images are now represented expressed by a set of representative patterns

Comment: repetitive

Response:

This sentence is improved as follows:

“After clustering and medoid selection, TIs are expressed by a set of representative patterns.”

Page 10 Line 215:

Sentence: MDS projects high-dimensional objects into a 2D cartesian space, where the difference between points in that space approximates the Hausdorff distance.

Comment: Really having trouble getting to this point. I just don't have any sense of how "distance" is being defined here. Also, are MDS1 and MDS2 defined in Figure 5?

Response:

We are sorry that the insufficient explanation brings confusion. First, we improve our description to express the definition of distance. The new content is shown in the following:

“We define the distance between any two training images as the difference between their representative patterns. A small distance indicates that two TIs have similar morphological structures.”

Then, we explain the main idea of MDS in detail in Line 215. Similar to the principal component analysis and other dimension reduction techniques, MDS1 and MDS2 are coordinates calculated from projection on the principal vectors.

“Once a distance is defined, we can visualize the metric space in low-dimensional Cartesian space using multi-dimensional scaling or MDS (Scheidt et al., 2018). The main idea of MDS is to project objects from a high-dimensional space into a 2D cartesian space, to visualize the similarity between all the TIs. Figure 5 show the projection of 166 training images in 2D, each dot represents a TI. Similar training images map close to each other in the MDS scatterplot.”

In addition, the caption of Figure 5 is improved:

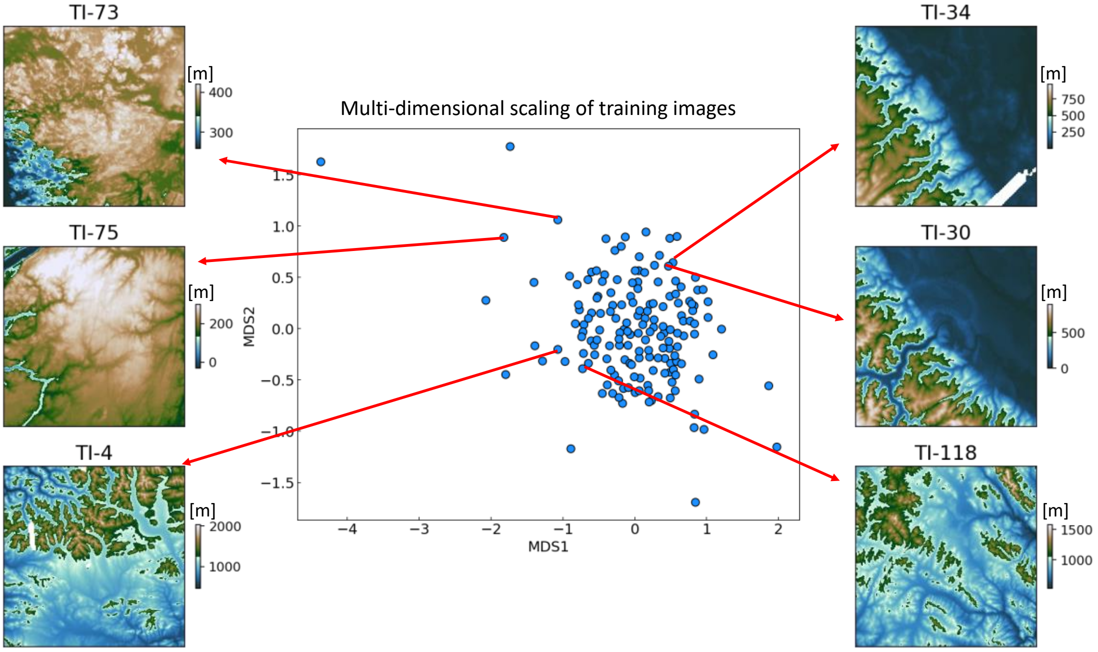


Figure 5. Visualization of the metric space using multi-dimensional scaling (MDS) into a two-dimensional cartesian space. Each dot on the plot represents a TI. It shows TIs with similar morphology are close in this metric space.

Page 10 Line 227:

Sentence: Direct sampling, by construction, avoids any artifact boundary, because the data template is not aware of the subareas.

Comment: aware?

Response:

We refine this sentence as follows:

“Direct sampling, by construction, avoids any artifact boundary between the radar line subareas, because the data template is not limited by subareas borders.

Page 10 Line 230:

Sentence: Training images of two adjacent areas are not necessarily independent.

Comment: Explain why this is important

Response:

The main reason is that there is spatial correlation between neighboring areas. In Figure 6, the area A2 has the similar morphology to the area A3. In comparison, the structures in A1 and A4 are considerably different. In order to create a satisfactory transition, TI selection of A2 should not be a complete independent process.

We added relevant explanations in our paper:

"Training images of two adjacent subareas are not necessarily independent because of spatial correlations between the subareas"

Page 10 Line 231:

Sentence: Our approach is to model the posterior distribution of each area through a probability aggregation problem.

Comment: ?

Response:

We have explained the probability aggregation problem in section 3.3.

Page 11 Line 241:

Sentence: $TI(A_i)$ is a discrete random variable that has 166 possible outcomes.

Comment: corresponding to the 166 specific training images chosen for this application, correct? The way it is written it sounds as if this is true for the general case.

Response:

We would like to thank the reviewer for his careful reading. Here, the number of possible outcomes is equal to the number of candidate TIs. Therefore, the corrected sentence is shown below:

"$TI(A_i)$ is a discrete random variable that has 166 possible outcomes (number of candidates TIs). To obtain the posterior distribution,

Page 12 Line 262:

Sentence: For example, if data of region A_i is highly correlated with data in region A_j, then they are likely redundant with respect to the training image selection.

Comment: are

Response: Thanks for this correction. The refined sentence is in the following:

"For example, if data of region A_i are highly correlated with data in region A_j, then they are likely redundant with respect to the training image selection."

Page 12 Line 268:

Sentence: A direct estimate of $P(TI(A_i)|d_{Ai})$ is challenging because the d_{Ai} are very high-dimensional.

Comment: What is meant by this?

Response:

The radar measure d_{Ai} is shown in the figure below. It is clear that each subarea contains many data points. For instance, there are 7982 known point in the region A2. In other words, d_{A2} is a vector of size 7982. Therefore, we describe the radar data d_{Ai} as a high-dimensional variable.

To better describe our method, the paragraph is improved in the following:

"A direct estimate of $P(TI(A_i)|d_{A_i})$ is challenging because the d_{A_i} are very high-dimensional. For example, there are 7982 radar measurements in subarea A2."

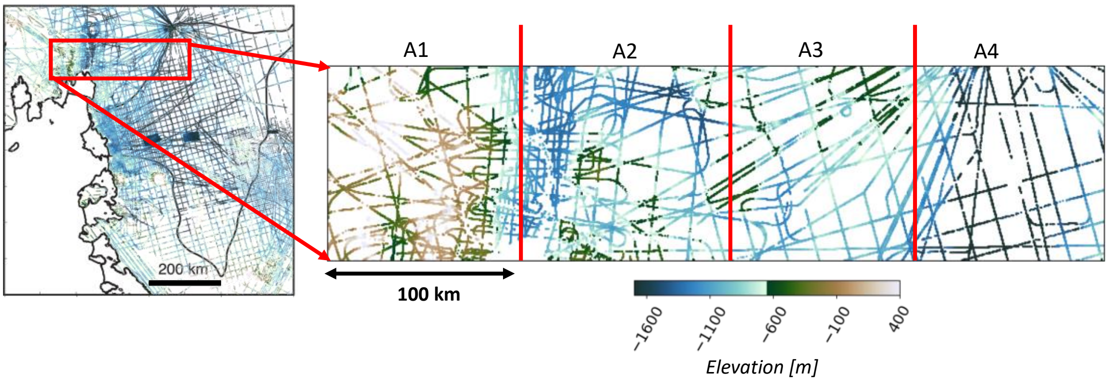


Figure 6. A subset of Pine Island Glacier is used to illustrate the methodology. Apparently, A2 and A3 share the similar morphology. Thus, our program assigns comparable TI to these two areas. In contrast, there is a considerable difference between A1 and A4.

Page 12 Line 269:

Sentence: We turn this high-dimensional problem into a low-dimensional as follows.

Comment: a low dimensional what?

Response:

As mentioned above, it is challenging to directly estimate the probability $P(TI(A_i), d_{Ai})$ because d_{Ai} if very high dimension. With the goal of efficiently finding the suitable TI, we convert a high-dimensional problem into a low-dimensional space.

"We turn this high-dimensional problem into a low-dimensional space as follows."

Page 12 Line 270:

Sentence: we find those training images that constitute a set of most probable training image,

Comment: images

Response:

Thank you for this check. The corrected sentence is shown below.

"With the aim of efficiently calculate the conditional probability, we replace the radar data d_{Ai} of big size with the most probable training images $\widehat{TI}$ of low dimension."

Page 12 Line 270:

Sentence: those images closest to the radar line data in that area

Comment: by what measure

Response:

The measure of "close" and distance is conducted according to the morphological consistency or similarity between radar data and TIs. Accordingly, we rewrite this sentence:

"those images closest to the radar lines in that area in terms of morphological similarities

Page 12 Line 270:

Sentence: Term this set as $\widehat{TI}$.

Comment: awkward - rephrase

Response:

We are gratitude for this suggestion. We rephrased it as follows:

"We term this set as $\widehat{TI}$.

Page 12 Line 276:

Sentence: $\{dis(I_{TI}(\widehat{TI}), d_{A_i})\}$

Comment: define

Response:

We have defined this formular in detail:

"… $\underset{\widehat{TI}}{\mathrm{argmin}}\left\{dis\left(\mathbb{I}_{\boldsymbol{TI}}(\widehat{TI}), d_{A_i}\right)\right\}$. where $\mathbb{I}_{\boldsymbol{TI}}$ is an indicator function which returns $\widehat{TI}$, a n-size subset of $\boldsymbol{TI}$. $\boldsymbol{TI} = \left[TI^{(1)}, TI^{(2)},\ \ldots, TI^{(166)}\right]$ and is the total set of training images."

Page 13 Line 289:

Sentence: Figure 7. Illustration of measuring the distance between training image and radar line data

Comment: Again - this caption is not sufficient to explain what is going on in the figure. Also, annotations are too small to read on the smaller panels.

Response:

Thanks for this suggestion. The new caption of Figure 7 is in the following.

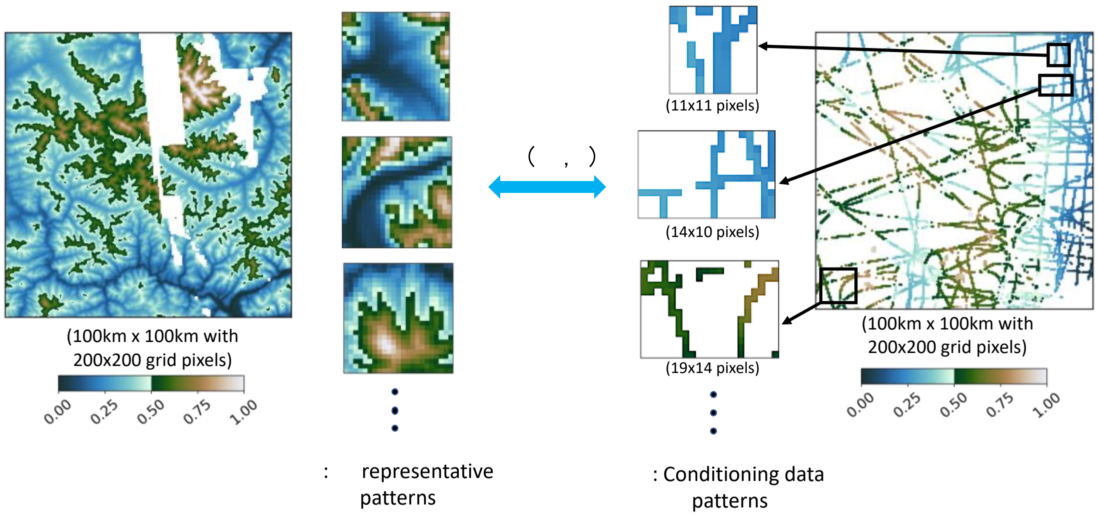


Figure 1. Illustration of measuring the distance between training image ($\boldsymbol{TI}$) and radar lines data ($\boldsymbol{d}$) in subarea A_1. We first extract a group of radar data patterns $\mathcal{D}$ from the simulation grid with flexible sized templates. Then the Hausdorff distances between the representative patterns $\mathcal{A}$ and radar patterns $\mathcal{D}$ are individually computed. Representative pattern $x_{\mathcal{A}}$ has a fixed size of 23x23 pixels, while the size of conditioning data pattern $y_{\mathcal{D}}$ varies.

Page 13 Line 291:

Sentence: We use a particle swarm optimization (PSO) to minimize the distance function $dis\left(I_{TI}(\widehat{TI}), d_{A_i}\right)$.

Comment: What is this?

Response:

Based on the explanation in Section 3.4.1, one important step in our method is to find TIs that have the minimum distance with radar measurements. This procedure can be mathematically defined as follows:

$$\{dis(I_{TI}(\widehat{TI}), d_{A_i})\}$$

Particle swarm optimization (PSO) is a widely used computational method to solve this optimization problem. The core idea is to iteratively improve a candidate solution with regard to an evaluation function. Compared with other optimization techniques, such as gradient descend and genetic algorithm, the advantages of PSO include less parameterizations, easy implementation and fast convergence with acceptable accuracy. Therefore, we choose PSO as a preferred optimizer for our initial TI selection.

In order to facilitate the reading, we add a PSO review paper in the new version:

"As a heuristic optimization approach, PSO has its specific advantages in requiring less parameterizations, easy implementation, and fast convergence with good accuracy (Rezaee Jordehi and Jasni, 2013; Sengupta et al., 2019)"

PSO is elaborated in the Appendix section.

Page 15 Line 306:

Sentence: We therefore consider a Gaussian kernel density estimation (KDE) to predict the probability to each TI.

Comment: What is this?

Response:

Kernel density estimation (KDE) is a statistical method to estimate a probability density function using only samples drawn from it. As shown in Figure 9 in our paper, PSO selects 3 images according to the similarity between radar data in the region A1 and 166 candidate TIs. In the metric space, three selected TIs are highlighted by the red while other images are expressed by the blue. Next, we estimate the prior probability distribution $P(TI(A_i)|\widehat{TI})$ on the basis of $\widehat{TI}$. Figure 9(b) displays the resulting probabilities computed by KDE. It is worth noting that each dot represents a candidate TI in the metric space. Apparently, our program assigns large weights to the images close to the three selected TI.

The technical detail about KDE is elaborated by Scheidt, Li and Caers in their book "Quantifying Uncertainty in Subsurface Systems" Section 3.3.2. Therefore, we add the reference in our paper:

"We therefore consider a Gaussian kernel density estimation (KDE) (Scheidt et al., 2018) to predict the probability that a training image TI is assigned to a subarea A_i."

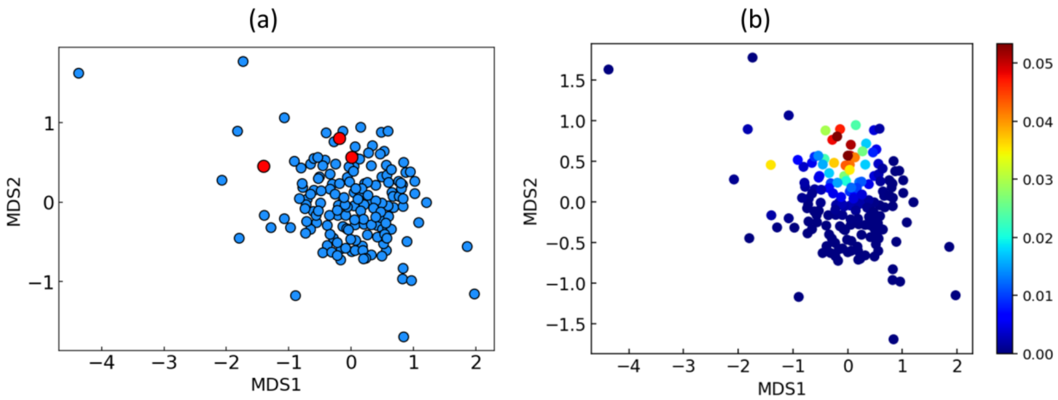


Figure 9. Probability computation based on the selected TIs. (a) Estimated $\widehat{TI}$ for subarea A1 in MDS space. The red dots are $\widehat{TI}$ while blue points represent other TIs. (b) Prior probability $P(TI(A_1)|d_{A_1})$ of each TI. Our kernel density estimation gives a high possibility to images close to $\widehat{TI}$.

Page 15 Line 313:

Sentence: We choose the optimal bandwidth by Silverman's rule of thumb (Silverman, 1981).

Comment: What is this?

Response:

As mentioned above, we adopt kernel density estimation (KDE) estimate a probability density function according to TIs selected by PSO. In KDE computation, only one parameter is the bandwidth of kernel. A small value of bandwidth leads to spurious data artifacts and an under-smooth result. By comparison, over-smooth is created by a large bandwidth. With the aim of facilitate practical applications, it is necessary to find an optimal and adaptive bandwidth in our case. Silverman's rule of thumb is a commonly used method to calculate the bandwidth when a Gaussian kernel is applied. The detailed process is explained by Silverman in his book "Density estimation for statistics and data analysis".

The related sentence in our manuscript is changed into:

"We calculate the optimal bandwidth h by following Silverman's rule of thumb (Silverman, 1981).

Page 16 Line 328:

Sentence: $K_E(\cdot)$ is the Epanechnikov kernel function.

Comment: This needs some kind of description

Response:

Epanechnikov kernel is a kernel function that is of quadratic form. The expression of Epanechnikov kernel is defined as follows:

$K(u) = \frac{3}{4}(1 - u^2)$ for $|u| \leq 1$

Explaining the Epanechnikov function in detail exceeds the scope of this paper. We add a reference with the goal to provide mathematical procedure.

"$K_E(\cdot)$ is the Epanechnikov kernel function (Fouedjio, 2020)."

Page 19 Line 328:

Sentence: Figure 13a shows one realization of the simulated result.

Comment: Can you combine Figures 12 and 13 so that the comparison can be made directly rather than flipping back and forth between the two figures?

Response:

Thanks for this suggestion. We have combined the Figure 12 and 13 together. The new figure is shown below.

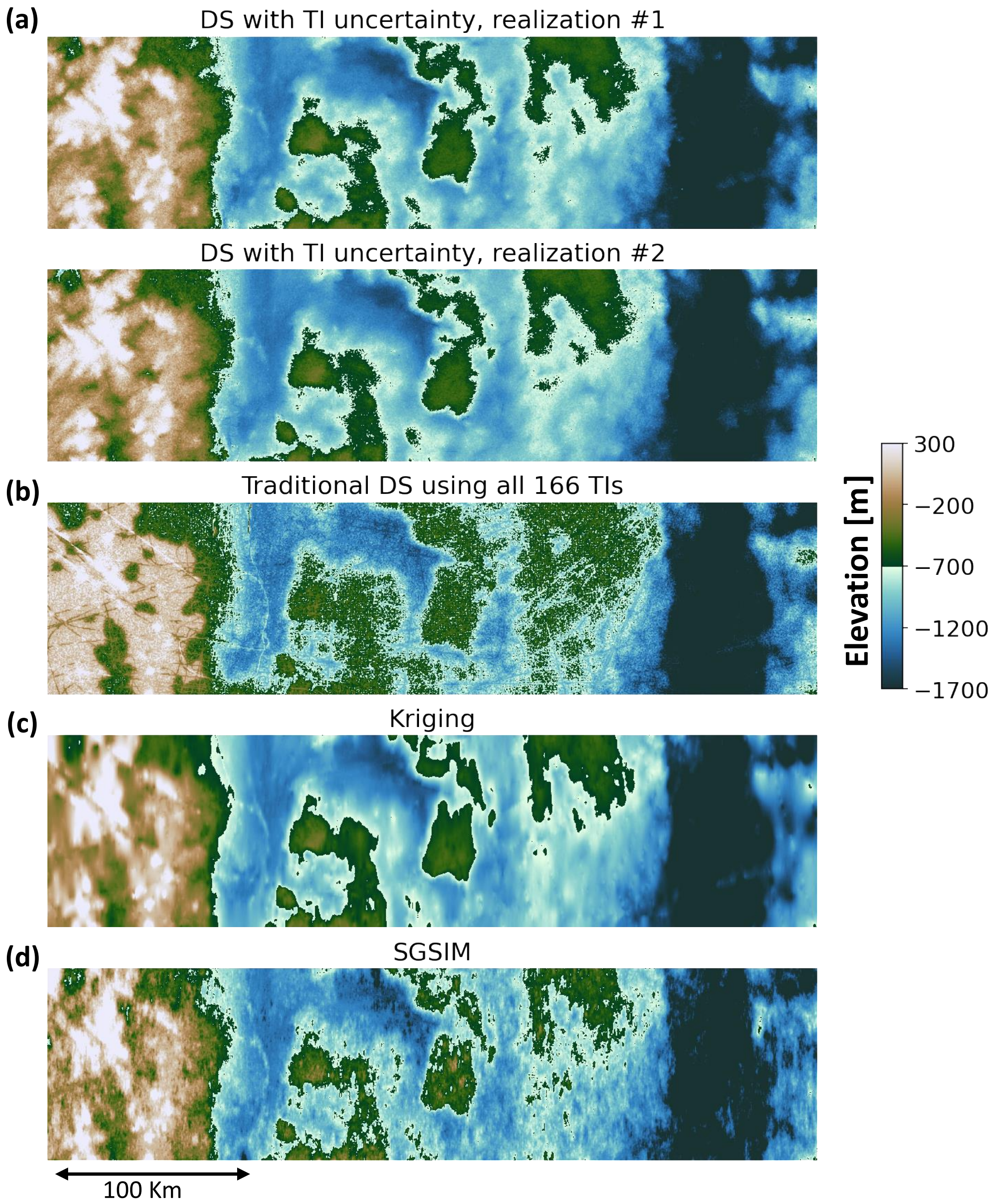


Figure 12. (a) Two realizations of DS simulated topographical models by filling the radar line gaps. Model realization number corresponds to the TI realization number in Error! Reference source not found.**. (b) line gaps filling by traditional DS using all the 166 TIs (without TI sampling). (c) and (d) line gaps filling using kriging and SGSIM.**

Page 20 Line 377:

Sentence: After all, kriging is a deterministic modeling approach.

Comment: Delete

Response:

We are grateful to the reviewer for this careful check. We remove the words. The new sentence is displayed below:

"Besides, Kriging is a deterministic modeling approach. It cannot quantify the spatial uncertainty."

Page 20 Line 378:

Sentence: Our SGSIM approach uses local ordinary kriging; this way non-stationarity is addressed by limited the neighborhood of spatial inference

Comment: limiting

Response:

The improved sentence is in the following:

"SGSIM uses local ordinary kriging where non-stationarity is addressed by limiting the neighborhood of spatial inference."

Page 20 Line 379:

Sentence: The limitation of SGSIM, an approach based on spatial covariances, lies on its limitations in capturing complex morphological features

Comment: only one "limitation" should be used in this sentence

Response:

Thank you. We modified this sentence as follows:

"As a covariance-based approach, the limitation of SGSIM lies on its ability to capture morphologically complex structures."

Page 20 Line 380:

Sentence: especially when the radar line data are very sparse (see the circle highlighted on Figure 13c)

Comment: You need to be clear about what is "wrong" with the circled regions. It is not obvious that the SGSIM is doing anything undesirable here.

Response:

The contrast between our program and existing methods is an important component in the quality evaluation section. At first, it is necessary to define the simulation target. One key contribution of our method is that 166 images from the deglaciated area in Arctic and Antarctica becomes training images. Therefore, a competitive method should (1) reproduce morphological structures from TIs, (2) honor the radar observations in the simulation grid, (3) create multiple realizations to express spatial uncertainty, and (4) save the running time.

With the aim of highlight the advantages, we separately compare our realizations with SGSIM and kriging models in the area with sparse radar data. It shows clearly the difference between SGSIM and our proposed DS approach.

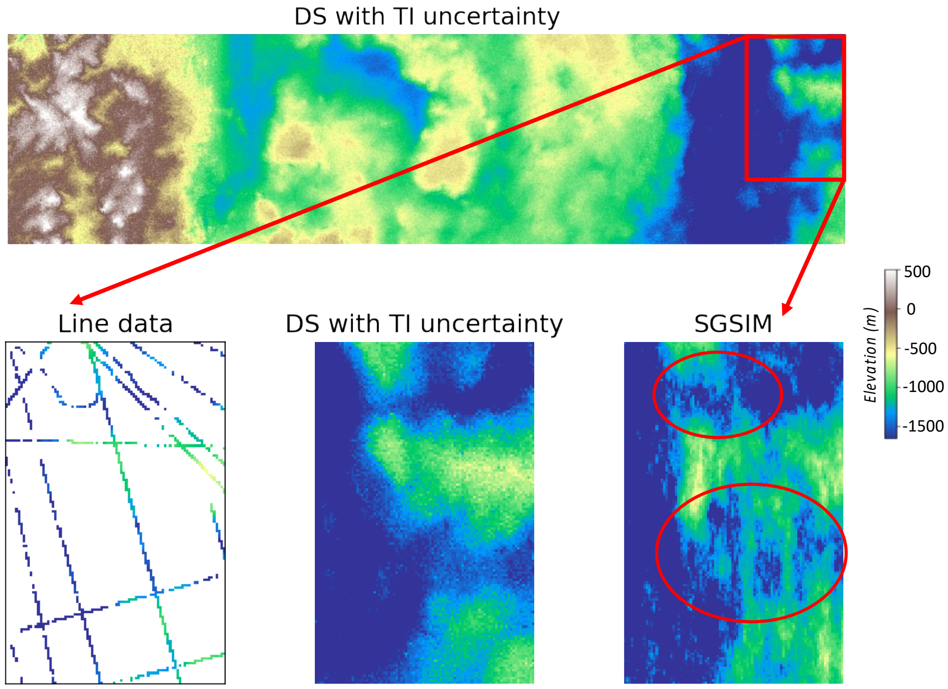
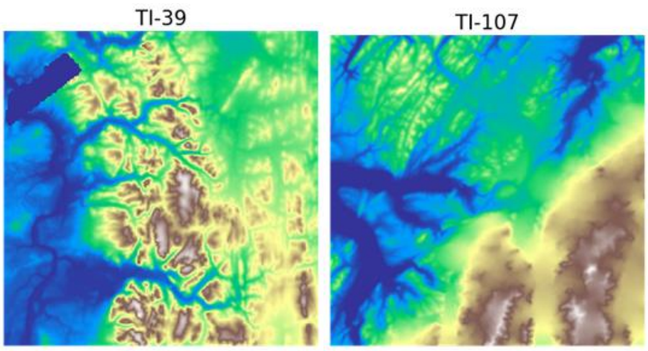


Figure 14. Comparison of SGSIM and the proposed DS with uncertain TI sampling in a local sparse lines area. Red circles highlight the areas where SGSIM failed to simulate meaningful channels.



Training images provided to our MPS simulation in the channelized area

Based on the preceding realizations, the advantages of our MPS realizations can be summarized as follows:

(1) MPS models have comparable morphology with TIs. Gaps between radar lines is suitably filled by the spatial structures in TI. By contrast, there are many fluctuations in SGSIM maps.

(2) There is no artifact around the radar lines in MPS realizations.

(3) Our MPS program has an ability to create a set of realizations with the objective to express spatial uncertainty.

(4) The computational efficiency is significantly improved by our method. Given a large number of TIs, the proposed method reduces the running time from 21 hours to less than 1 hour.

Page 20 Line 381:

Sentence: In Figure 15, we also compare the empirical variograms from the modeled topographical maps using the four different approaches.

Comment: Since this field is spatially heterogeneous, how is this computation limited?

Response:

Thanks for this advice. Here we use the global empirical variogram mainly for a simple comparison. The empirical variogram is very clear to show how the different gap-filling methods retain the spatial correlations.

Page 20 Line 384:

Sentence: Overall, it shows the TI sampling approach performs the best in terms of improving the modeling speed, simulation quality, and capturing the spatial uncertainty.

Comment: Need to do more to distinguish superiority over SGSIM.

Response:

We agree with reviewer. Therefore, we added Figure 14 in the revision to specifically compare our TI sampling approach with SGSIM.

Point-to-point responses to RC2

The MPS-BedMappingV1 paper focuses on the subglacial topography of the Amundsen Sea Embayment (ASE), which is one of the most important parts of Antarctica to resolve so that ice sheet models can better estimate ice loss and hence sea level rise projections. Using a spatial statistical method, the authors have matched suitable digital elevation models (DEMs) of bed topography from deglaciated areas (in the Arctic and Antarctic off-shore regions) with sparse groundtruth ice-penetrating radar bed elevation data over Pine Island Glacier and Thwaites Glacier. These matched DEMs are training images (TIs) which are used as the basis for a gap-filling algorithm which can be used to fill in the bed topography of areas without data coverage in between radar flightline. The product of this exercise is multiple realizations of subglacial topography over the ASE region, which the authors then use to quantify subglacial hydrological flow uncertainty.

The paper itself was technically challenging, but the authors have done a great job at walking through the procedure step by step, with generous use of figures to illustrate their methodology. One of the core contributions and fascinating pieces of this work is the probabilistic framework of finding training images (from a large pool of deglaciated bed DEMs) that match best with sparse radar data observations over a glaciated area. The amount of work behind sourcing the relevant datasets and building the model framework is no small feat, and I applaud the authors for taking the time to implement this very important piece of work.

In general, the paper is very readable, though I found that some of the terminology could be made more precise (e.g. high-resolution could be quantified as 500 m spatial resolution). Using subglacial hydrology as the motivation for this subglacial topography modelling work seemed a little weak in my opinion, but I think the authors can do better job to persuade us on what other benefits a high-resolution bed topography model entail. The authors could definitely be a lot more generous with their citations to mention other related interpolation or gap-filling techniques (e.g. Graham et al 2017), and list some more newer and key publications in the cryospheric science field. Following this will be more specific comments and technical suggestions on ways to improve the paper.

Responses: We thank the reviewer for thoughtful and thorough review of our manuscript. We have addressed the line-by-line comments below and added more citations recommended by the reviewer. We agree that there are potentially high-impact scientific directions beyond hydrological modeling. However, this is beyond the scope of this study. Instead, we have added additional discussion on the scientific implications of this work. Changes are highlighted as Red in the revised manuscript.

Please see below our “point-to-point” responses.

Specific comments and technical corrections

General point on figure presentation

- In general, the quality of the figures in this paper is great and gets the point across, but being a perfectionist, I have some nitpicks on missing units and coordinates labels. I encourage you to check out Rougier et al., 2014 for some tips on making better figures, both for this paper, and in your future publications.

Response: Thanks for the suggestion! We agree it is important to make proper scientific figures. We have carefully read the paper written by Rougier et al. We have revised/re-made most of the figures, by adding missing units and colorbars, and revising captions. We also added coordinates labels to the figures as suggested.

- Note that the current colour map used in this study may distort the data and is not accessible to people with colour-vision deficiencies. I encourage you to use one of the freely available Scientific colour maps provided at https://www.fabiocrameri.ch/colourmaps or see Crameri et al., 2020 for more details. You can use the package at https://github.com/callumrollo/cmcrameri which integrates with your matplotlib plots.

Response: We appreciate this advice and agree that it is more proper to use universal colormaps. We therefore have re-made all the Figures to color-blind friendly in the revision. We thank the reviewer for kindly sharing color-blind friendly colormaps.

Section 0: Abstract

- pg1, L1: Suggest changing "West Antarctica" to "Amundsen Sea Embayment" in the title, since the study doesn't cover some parts of West Antarctica like the Siple Coast or Weddell Sea.

Response: While our focus is primarily on the Amundsen Sea Embayment, our study area also encompasses the Bentley Subglacial Trench and upstream areas of the Ross ice streams. Therefore, we believe that “West Antarctica” is appropriate.

- pg1, L13-14: "However, mapping of subglacial topography is subject to high uncertainty". Please quantify the order of the uncertainty, is it 10 metres? 100 metres?

Response: We have clarified that this is 100s of meters.

- pg1, L14-15: "This [high uncertainty] is mainly because the bed topography is measured by airborne ice-penetrating radar along flight lines with large gaps up to tens of kilometers". Note that ice penetrating radar measures bed elevation accurately, I think what you mean to say is that the gaps (without radar coverage) is where the uncertainty lies. Please reword sentence.

Response: This has been clarified, though we note that radar measurements can sometimes have large uncertainties (e.g. Gogineni 2014).

- pg1, L19: "We collect 166 high-resolution topographic training images". Please quantify high-resolution, i.e. write high-resolution (500 m).

Response: This has been added.

- pg1, L23: "then provide candidate" -> "then provides candidate".

Response: This has been amended.

- pg1, L24: "traditional MPS" -> "traditional multi-point geostatistics (MPS)".

Response: We have added the definition of the MPS acronym earlier in this section.

- pg1, L25: "demonstrates significant improvement". Please quantify improvement over previous baseline, e.g. "significant improvement of 10 metres"

Response: The objective of stochastic simulation is not to improve local accuracy, so we cannot provide a specific value for elevation improvement. Rather, the objective of geostatistical simulation is to improve the geologic realism of the topography. We clearly demonstrate that our method improves the morphological fidelity of interpolations (e.g. Figure 12), and is able to do so more quickly than traditional MPS approaches. As such, it is sufficient to say that we "demonstrates significant improvement in the topographic modeling quality and efficiency of the simulation algorithm."

- pg1, L27: "We use the multiple realizations to investigate the impact of basal topography uncertainty on subglacial hydrological flow patterns". This sentence should be expanded to mention what the results of the investigation are exactly. From reading Section 4.2, besides seeing that multiple subglacial hydrological realizations were made, I could not quantify how the method here improves upon an existing water flow model (e.g. from the subglacial water flux map of Le Brocq et al., 2013).

Response: The objective of this section is not to improve hydrological models. Rather, we demonstrate that our method can be used to quantify uncertainty in subglacial hydrological flow paths with respect to topographic uncertainty. For example, in Figure 19, the flow path locations differ in each realization, and the average across the realizations is different from the flow paths calculated using BedMachine. Our objective is to show that topography contributes to significant uncertainty in hydrological modeling, and that hydrological models made using a single DEM are at risk of over-interpretation. As such, we have left this sentence unmodified.

Section 1: Introduction

- pg1, L29: "The topography beneath the Greenland and Antarctic ice sheets is essential for nearly every ice sheet investigation". Need to cite more examples, the following sentences (L30-L32) only include citations for Antarctic papers, and are not representative of the cryospheric literature. I have listed a few suggested citations in the next bullet points, but you are welcome to add others too.

Response: Citations have been added.

- pg1, L30: "modeling subglacial hydrology (MacKie et al., 2021)". Please cite more, e.g. Siegert et al., 2016; De Fleurian et al., 2018; Sommers et al., 2018; etc

Response: Citations have been added.

- pg1, L30: "interpreting geologic conditions (Holschuh et al., 2020)". This sentence seems incorrect, how can bed topography be used to interpret geologic conditions? The cited article by Holschuh et al., 2020 mentions "a likely role for preexisting geology in glacial bedform shape", i.e. that geology affects topography, not the other way round. Please reword the sentence, and add more Greenland/Antarctic subglacial geology papers e.g. Anandakrishnan et al., 1998; Aitken et al., 2014; Bell et al., 1998; Lowry et al., 2020; etc

Response: Topographic measurements can be very helpful for making general (and often qualitative) geologic interpretations (e.g., King et al., 2009, Bingham et al., 2009). The Holschuh et al. study is particularly informative because it provides the only 2D imaging of subglacial topography. We have added citations to studies making geologic interpretations from topographic measurements, including some of the work that builds on the Holschuh swath radar surveys.

- pg1, L31: "estimating ice volume and sea level rise contributions (Fretwell et al., 2013)". Please also cite BedMachine papers by Morlighem et al., 2017; Morlighem et al., 2019.

Response: This has been added.

- pg1, L31: "ice sheet modeling for sea level rise projections (Le clec'het al., 2019; Seroussi et al., 2017)". Please cite more, e.g. Schlegel et al., 2018.

Response: This has been added.

- pg2, L35: "radar along flight lines separated by up to tens of km (Fretwell et al., 2013; Herzfeld et al., 1993)". Please cite more of the underlying radar data sources, e.g. Robin et al., 1970; Shi et al., 2010; etc. Also, you should mention ice penetrating radar surveys with closer sub-kilometre spacing e.g. over iSTAR data over Pine Island Glacier by Bingham et al., 2017 and Rutford Ice Stream data by King et al., 2016.

Response: We have added citations.

- pg2, L36: "using methods such as kriging (Fretwell et al., 2013)". BEDMAP2 was not created using kriging, they used the ArcGIS Topogrid algorithm (see pg 381 of Fretwell et al., 2013). Please find an alternate paper citation that uses kriging, or change kriging to Topogrid.

Response: Correct. We have modified this sentence and added a citation to kriging.

- pg2, L37: "model inversions (Huss and Farinotti, 2012; Morlighem et al., 2017, 2020)". Recommend citing ITMIX paper by Farinotti et al., 2017 too. Also note that ice sheet model inversion techniques are not deterministic, I suggest you separate this sentence from the previous L36 "generally interpolated deterministically using methods such as kriging" part.

Response: While there are some stochastic elements to sampling the parameter space when solving inverse problems, this is done in order to converge on a unique solution. Ice sheet model inversions are underdetermined, so there is no unique solution. However, regularization parameters are always used to force a unique solution (at the cost of excessive smoothing). As such, we consider this to be deterministic.

- pg2, L37: You may also want to mention other spatial statistical papers by Graham et al., 2017 and Goff et al., 2014 somewhere here for completeness.

Response: Goff is cited. We have added Graham.

- pg2, L49-50: "contain a globally wide range" -> "contain a range". Make this sentence more concise.

Response: This has been changed.

- pg2, L59: "non-stationary bed topography is directly measured using high-resolution remote sensing data such as satellite images". Bed topography is not measured directly by optical satellite images (e.g. Landsat/Sentinel-2), they can be measure indirectly using photogrammetry. Direct measurements would require using satellite altimeters such as Cryosat-2 or ICESat-2. The ArcticDEM dataset by Porter et al., 2018 you cited uses stereo-photogrammetry on WorldView imagery. Please reword this sentence, and also mention which version of ArcticDEM was used in this study (is it v3.0?).

Response: This has been changed to “The non-stationary bed topography is measured using high-resolution remote sensing data such as satellite imagery…”. We use version 3. We have added this clarification to the training data description in Section 2.

- pg2, L60-61: "reveal glaciated morphologies resembling the topography beneath the ice sheets" -> "reveal glaciated morphologies resembling the topography beneath the contemporary ice sheets". Need to be clear that you are using deglactiated paleo-ice-sheets that resemble the current present-day ice sheet.

Response: This has been clarified.

- pg2, L61-62: "They therefore bear significant information on the subglacial topography". Clarify exactly what is meant by 'information', something to do with the texture or structure of the topography? Or something else?

Response: This has been changed to “morphological information”

- pg2, L63: "satellite imagery of deglaciated topography has not been explored to stochastically simulate subglacial topography". Again, optical satellite imagery does not contain direct topography information without extra processing. Be clear that you mean satellite imagery derived DEMs.

Response: This has been changed to “deglaciated topography has not been used to stochastically simulate…”

- pg3, L66: "high-resolution training images" -> "high-resolution (500 m) training images".

Response: This statement is referenced by several different studies which use different resolutions. So we do not add this.

- pg3, L76-77: "Then MPS then uses" -> "MPS then uses".

Response: This has been changed.

- pg3, L83: "More importantly," -> "Moreover,". Be concise.

Response: This has been changed.

- pg3, L88-89: "there are no available training images because" -> "there is limited training images because".

Response: This has been changed to “there is limited training imagery available…”

- pg3, L89-90: "Satellite altimetry observations from deglaciated areas in the Arctic offer a potential source of training imagery". Note that satellite altimetry products (e.g. from

Cryosat-2/ICESat-2) are typically point-based measurements, not raster grid images. Suggest rewording this sentence to "Satellite-based observations of deglaciated areas...".

Response: This has been modified.

- pg3, L90: "training imagery. However, training images retrieved from the Arctic would be" -> "training imagery, but these training images would be". Be concise.

Response: This has been changed.

- pg3, L92: "significant computational burden". Could you quantify what significant means?

Response: MPS algorithms search through training images to perform a simulation. So the computational demand scales with the training image volume. And ArcticDEM is very large. It is especially important to avoid large computational requirements when doing stochastic simulation because we typically generate many realizations.

- pg3, L96: "We will address" -> "We address". Be concise.

Response: This has been changed.

- pg3, L98-99: "We first collect a large amount of topographic images to serve as the training images. These images are taken from the deglaciated areas in the Arctic and Antarctica" -> "We first collect topographic images from deglaciated areas in the Arctic and Antarctic regions to serve as training images (Fig 2)". Be concise and link to Fig 2.

Response: This has been changed.

- pg4, L103: "We will demonstrate our method using the entire" -> "We demonstrate our method over the entire".

Response: This has been changed.

- pg4, L106-107: "We use the topographic simulations to" -> "We then use these simulated topographic images to"

Response: This has been changed.

- pg4, L107: "in order to investigate the impact of topographic uncertainty on hydrologic" -> "in order to investigate how topographic uncertainty impacts hydrologic"

Response: This has been changed.

Section 2: Radar data set & training images

- pg4, L110-118: Recommend to provide a table of all the data sources used since the list is long. At a minimum, the table columns should include the placename of the data (e.g. Thwaites Glacier, Laurentide Ice Sheet, etc), the data citation, and the paper citation.

Response: We provided detailed explanations and reference of the training images in the provided Zenodo repo. We don't think it is necessary to provide another table. Such a table does not provide additional important information, but will require a lot of time.

- pg4, L110: "seafloor bathymetry measurements". Fig 1 map doesn't show bathymetry, i.e. areas outside the grounding zone.

- pg4, L111: "subaerial topography" -> "ice surface topography". Also, could you show where the REMA DEM is being used over the ASE region? There are not many deglaciated places in the ASE on land.

Response to the above two points: we are using subaerial topography. There is some exposed topography in the ASE, which is included in Figure 1, though there is so little exposed rock that it is difficult to see.

- pg4, L113-114: "gridded at a 500-meter resolution" -> "gridded at a 500-meter spatial resolution". Specify the gridding algorithm used, and how data gaps (i.e. NaN values) are treated.

This has been modified.

- pg4, L114-115: "(Holschuh et al., 2020) (provide" -> "(Holschuh et al., 2020) provide"

This has been modified.

- pg4, L115-116: "we augment the available training data with deglaciated subaerial topography from the ArcticDEM (Porter et al., 2018)" -> "we increase the available training data with deglaciated subaerial topography from ArcticDEM data (Porter et al., 2018)". Augmentation can have a slightly different connotation in machine learning, such as synthetic data generation, suggest to just use 'increase'.

This has been modified.

- pg4, L119: "may have experienced additional depositional processes" -> "may have experienced additional erosional or depositional processes".

This has been changed.

- pg4, L119-120: "any topographic alterations are likely minimal at a 500 m resolution". How is it likely minimal? Need to provide a citation or have some numbers to back up this claim.

Response: In Pine Island Bay, it’s taken tens of thousands of years to build up several meters of sediment (e.g. Kirshner et al., 2012, QSR). A few meters of vertical change are not going to alter the spatial statistics of a training image with a horizontal resolution of 500 m.

- pg4, L120: "geologic settings" -> "geological settings".

This has been changed.

- pg4, L121: "100x100 km^2" -> "100 km x 100 km".

This has been changed.

- pg4, L121-122: "training image data repository is publicly accessible from Zenodo repository (https://zenodo.org/record/5083715#.YQT2JI5Kiiw, DOI 10.5281/zenodo.5083715)" -> "training image data is publicly accessible on the Zenodo repository (https://doi.org/10.5281/zenodo.5083715)".

Response: This has been changed following the reviewer’s suggestion.

- pg5, L123: Figure 1 colorbar. Suggest using Scientific Colour Map instead of "terrain" because values at 1500 m are white, which is the same as the NaN values. You may want to select a different colour for NaN values and/or indicate with a small box that NaN values are of a certain colour.

Response: Thanks for pointing this out. We have change the colormap to according the Crameri et al., 2020 suggested by the reviewer.

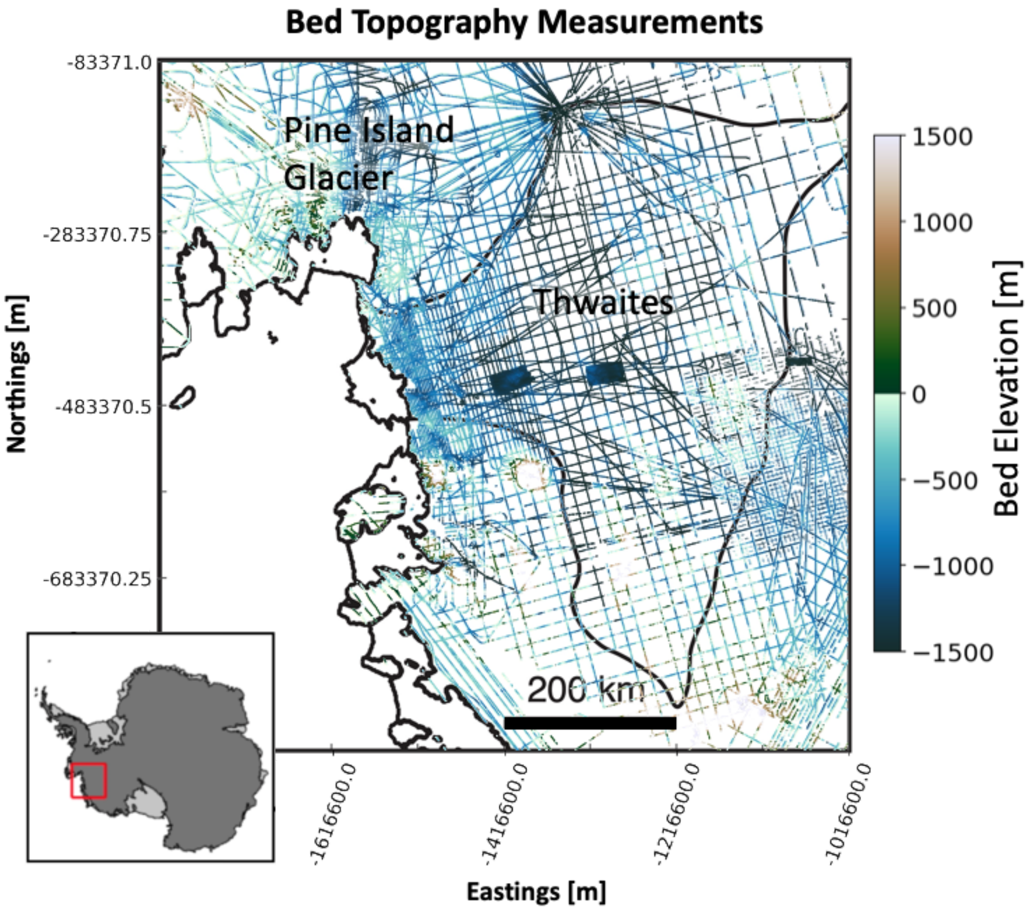


Figure 1. Radar line surveys of the Thwaites and Pine Island glaciers in the Amundsen Sea Embayment of West Antarctica. The red box implies the location of the study area in Antarctica. Black lines indicate boundaries for Thwaites Glacier, ice shelves, and the grounding line (the point where the ice detaches from the bed and achieves flotation). The topography patches in the center of Thwaites Glacier were measured using swath radar (Holschuh et al., 2020).

- pg5, L123: Figure 1 axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

Response: Coordinates have been added

- pg5, L127: Need to indicate that the inset map's red box shows the study region.

This has been changed.

- pg5, L128: "The topography patches" -> "The two topography patches".

Response: This has been changed.

- pg6, L131: Figure 2a axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

Response: Again, sorry, we do not have the exact coordinate locations of the training images data in our received dataset. Otherwise we would add the coordinates. After all, the coordinates do not have influences on applying our proposed MPS method

- pg6, L131: Figure 2a map. What do the red boxes with the checkerboard patterns mean? Mention in the caption.

Response: This has been changed.

- pg6, L131: Figure 2b axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

Response: Please see response to “- pg6, L131:Figure 2a axes”

- pg6, L131: Figure 2b map. Why are there some black lines on the main map? Black does not seem to be in the colorbar.

Response: Some areas appear darker because it is a shaded relief map.

- pg6, L132: Figure 2c axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

Response: See response to “- pg6, L131:Figure 2a axes”

- pg6, L132: Figure 2c map. T1-165 and T1-166 look like DEMs from Holschuh et al., 2020, are they rotated 90 degrees? Where is North? Ideally there would be x/y or lon/lat coordinates to show the location and orientation of these training images.

Response: See response to “- pg6, L131:Figure 2a axes”

- pg6, L132: Figure 2 caption. "(b) Antarctica". Are training images gathered from the whole of Antarctica, or just the Amundsen Sea Embayment region as shown in Fig 2b?

Response: Mainly the Amundsen Sea Embayment region. We have revised the paper.

Section 3: methodology

Section 3.1: Multiple-point geostatistics

Section 3.1.1: Overview

- pg7, L142-143: "Both MPS and covariance-based methods have the ability to interpolate data exactly". Please clarify what "exactly" means, e.g. is the vertical precision 1 metre? 5 metres?

Response: This means that grid cells with data are unchanged. We added more explanations.

- pg7, L149: "section 0" -> "section 3.1.2".

This has been changed.

Section 3.1.2: Direct Sampling (DS)

- pg7, L164: "a unknow" -> "an unknown".

This has been changed.

- pg7, L165: "to find the similar structure in TI. The similarity within" -> "to find a similar structure in the TI pool, Similarity within".

This has been changed.

- pg8, L166: "defined by a certain distance metric". What are examples of distance metrics that can be used? Suggest change sentence to "defined by a distance metric (e.g. Euclidean distance, Hausdorff distance, etc)".

Response: The suggested sentence is beneficial for improving expression. We change this sentence into:

"The similarity within DS is defined by a distance metric (e.g. Hamming distance for categorical variable and Euclidean distance for continuous variable)."

- pg8, L170: Figure 3b. Put a dashed box (width: 4 pixels, height: 3 pixels) around the [37], [?][86][80] cells, otherwise they seem disconnected. It took me a while to understand what was happening.

- pg8, L170: Figure 3 colorbar. Add units for the colorbar.

Response: Thank you. In Figure 3, conceptual training image and simulation grid are provided as an example. The new figure is shown below:

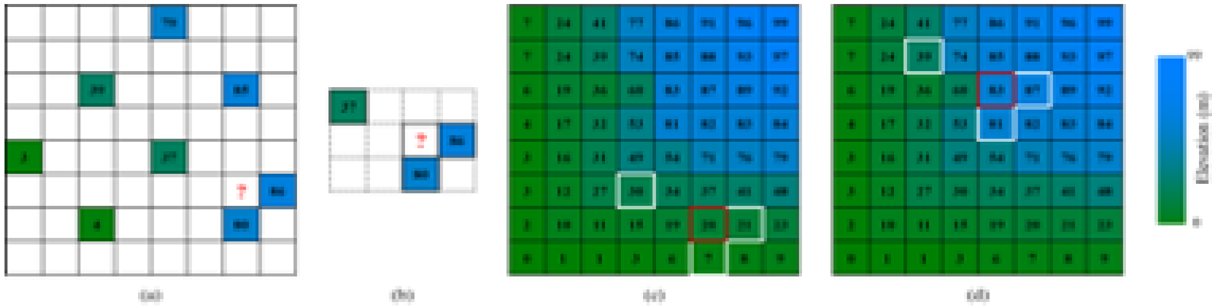


Figure 3. Conceptual example of the DS point simulation. (a) Radar lines on the simulation grid; (b) Three known points (value: 37, 80, 86) constitute a conditioning data pattern; (c) A mismatch pattern in TI; (d) A similar pattern in TI.

- pg8, L174: "Based on the explanation above, there are mainly three important parameters in DS" -> "Based on the explanation above, there are three main parameters in DS to tweak".

Response: This has been changed.

- pg8, L178: "The TI pattern with mismatch distance below t will be accepted and pasted to the simulation grid". What happens when 2 or more TI patterns are found below the threshold? Is the one with the lower distance metric value (i.e. more similar) picked? Or is it done on a first comes first serve basis? Please add a sentence to clarify in text.

Response: It searches the TI until the first acceptable pattern is found. So it stops after finding a match. This is a widely-used, well-documented algorithm and we have included the appropriate explanations for further information.

- pg8, L182: "avoid verbatim copy, an recommended value of" -> "avoiding verbatim copies, a recommended value of"

This has been changed.

Section 3.1.3: A metric space for training images

- pg9, L192-193: "we rescale each TI to range between 0 and 1 by min-max normalization (Han et al., 2012)". Are the min and max values used for normalization obtained per image (i.e. local) or derived from the entire training dataset (i.e. global). Depending on the implementation, I have some concerns on how suitable this normalization is. Consider a training image with high mountains and low valleys, and a training image which is relatively flat and lacks topographic features. A local min-max normalization would stretch both images into the same 0-1 range, i.e. little bumps on the flat terrain training image would become mountains. In effect, this could lead the subsequent model to use training images from flat areas to inform the gap filling of mountainous areas.

Response: The reviewer is insightful to raise this comment.

The simulations are carried out in a transformed reference. It is unlikely that a transformed flat area would be used to simulate a mountainous region because there are significant morphological differences between the two beyond just elevation scale.

- pg9, L194: "from each TI with a" -> "from each TI using a".

Response: This has been changed.

- pg9, L195-196: "into a finite number of groups". How many groups? Please state the number here.

Response: We have added a description of how the number of groups is determined.

- pg9, L197-198: "We referred to the commonly used distance threshold in DS approach to set it as 0.1 (Meerschman et al., 2013) of the maximum pattern distances of the TI". Sentence not clear, maybe reword to "We set the distance threshold for DS to 0.1, as per Meerschman et al., 2013"? Also, the Meerschman et al., 2013 paper uses 0.05 as a default (though they have a range of recommended values), could you elaborate why 0.1 used here instead, and not another value like 0.2?

Response: This has been clarified. Thanks for this comment.

According to experiments performed by Meerschman et al. (2013), it is advised to use $t \leq 0.1$ and $n \geq 30$ and avoid $t \geq 0.2$ and $n \leq 15$ for a continuous training image. As mentioned before, the distance threshold t is applied to define the similar pattern with a given pattern. A large value of t increases the number of similar structures in TI. Thus, it is possible that the program pastes an incompatible structure in the simulation grid. Based on empirical experiments, $t=0.2$ is not proper in the continuous variable scenario.

- pg9, L200: "The distance used in the clustering is the normalized Euclidean distance". The agglomerative hierarchical clustering scheme seems to typically works on points, but your training images are raster grids (with a z value), so how is the Euclidean distance computed? Are you treating each pixel as a point with x,y,z values and running the clustering on those, or are you computing the pixel-wise distance (i.e. elevation difference) between the grids and taking the mean? Please clarify.

Response: We compute the pixel-wise distance. We added more clarification:

“The distance used in the clustering is the normalized Euclidean distance between the pixel-wise evaluations”

- pg9, L202: Figure 4 colorbar. Add units for colorbar (e.g. Normalized Elevation). Also, does the min-max normalization turn NaN values to 0 (because I what looks like data gaps as dark blue, which is 0 on the colorbar), or are NaN values still treated as NaN?

Response: We added the colorbar to Figure 4. The image in Figure 4 is processed by the min-max normalization. Let z denote the value of elevation in the original image. zmin and zmax are the minimum and maximum elevations, respectively. The min-max normalization is defined as follows:

z'=(z-zmin)/(zmax-zmin)

where z' is the value in the normalized image.

Therefore, there is no unit in the normalized image.

In our method, the unknown points (NaN values) do not participate in the normalization step. Therefore, Nan values are still treated as NaN. With the aim of avoiding confusion, we add an explanation in our paper:

"Like other MPS programs such as SNESIM (Strebelle, 2002), we extract a set of spatial patterns from TI with a fixed template. It is worth noting that the unknown points in TI are not involved."

- pg9, L204: "training images are now represented expressed" -> "training images can now be expressed".

Response: This has been changed.

- pg9, L207: "used to define the" -> "used to quantify the".

Response: This has been changed.

- pg9: L211: "x_A is any representative pattern in A". The term "any representative pattern" can sound a bit ambiguous, suggest rewording to clarify that "x_A is any one of the possible representative patterns of A".

Response: This has been changed in the revision.

- pg9, L211-212: "d(.) is the Euclidean distance between any two representative patterns". Again, see my point above on pg9, L200. Is the Euclidean distance computed by vectorizing the pixels to points first, or pixel-wise on the images?

Response: See response to L200 comment.

- pg9, L212-213: "the modified Hausdorff distance". My understanding of Hausdorff distance is that it is a method originally developed for x, y point cloud measurements, but your training images are raster grids. Depending on whether you are treating each pixel in your training images as x, y, z points, or computing the Euclidean distance pixel-wise (i.e. distance is calculated for pixel-values in the same image coordinates), this can result in very different results, so you will need to clarify and possibly justify your approach.

Response: The modified Hausdorff distance is an effective tool to measure the similarity between two datasets. In our method, we apply the Hausdorff distance to quantify the similarity between two training images. Let's take the Figure 4 as an example. Since the size of template is 23×23, there are 529 points in one pattern. In other words, each pattern represents a point in a 529-dimensional space. In addition, our program identifies 16 patterns from TI_A. Therefore, the pattern set A has 16 points in the high-dimensional space. By comparison, 21 patterns are detected from TI_B.

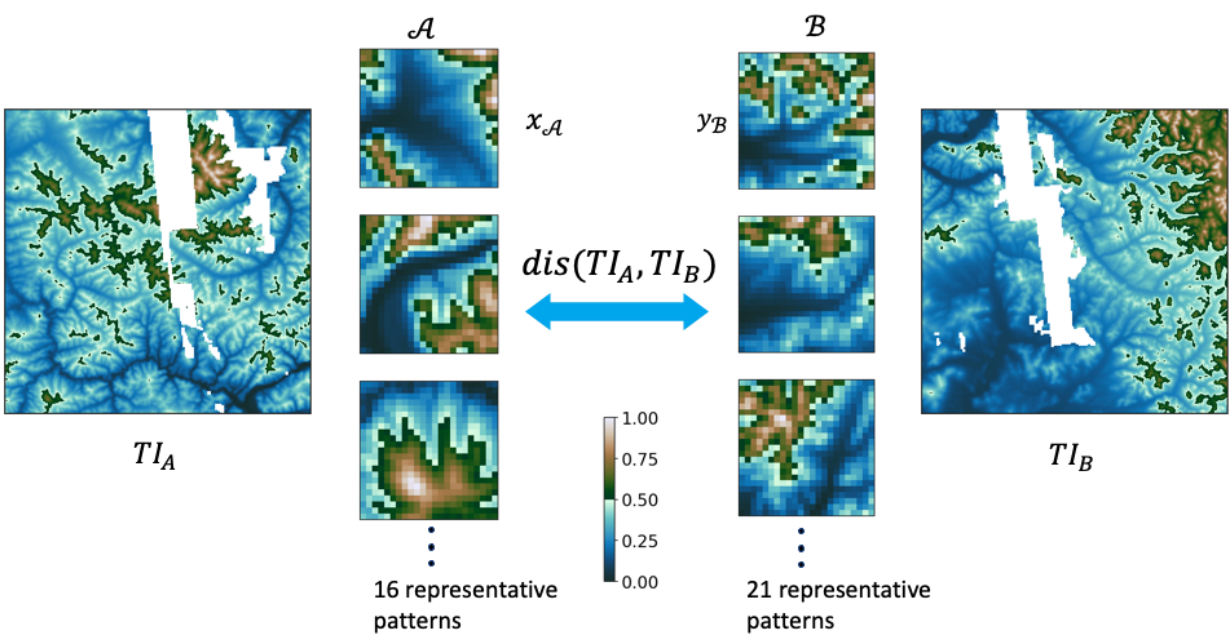


Figure 4. Calculating the distance between any two training images (TI_A and TI_B) using modified Hausdorff distance. There are three key steps: (1) Extracts training patterns with a fixed template. (2) The representatives are selected by a hierarchical clustering method. In this example, the computer found 16 important patterns from TI_A and 21 patterns are from TI_B. The number of representatives is dependent on the complexity of morphology. (3) Calculates the modified Hausdorff distance between two pattern sets. The output distance becomes an indicator of similarity between two TIs.

Next, we employ the modified Hausdorff distance to quantify the similarity between two datasets using Eq 1.

- pg10, L219: Figure 5 colorbar. Add units for the colorbar.

Response: The units have been added in Figure 5 in the revision.

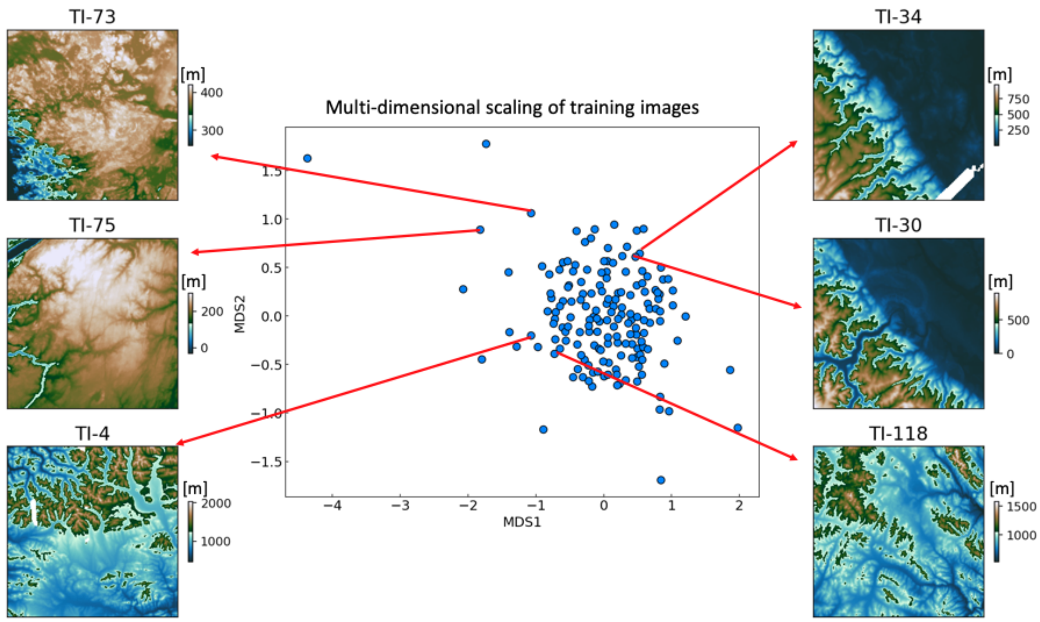


Figure 5. Visualization of the metric space using multi-dimensional scaling (MDS) into a two-dimensional cartesian space. Each dot on the plot represents a TI. It is clear that TIs with comparable morphology are close in this metric space.

- pg10, L220: Figure 5 caption. "TI" -> "training image (TI)".

Response: We are grateful for this suggestion. However, the abbreviation TI has already been explained in the preceding paragraph. It is not necessary to redefine it in the caption of Figure 5.

Section 3.3: Formulation of the problem through probability aggregation

- pg11, L245. Equation 3. Need to state what is i and j.

Response: The definition of i and j are added.

- pg12, L258: "additional weight term" -> "additional weight term (w_{ij}) as follows"

Response: This has been changed.

Section 3.4: Probability of training images given radar line data.

Section 3.4.1 Most probable set of training images

- pg12, L268: "challenging because the d_{Ai} are very high-dimensional" -> "challenging because the flight radar line data d_{Ai} is high-dimensional".

Response: This has been changed.

- pg12, L269: "low-dimensional" -> "low-dimensional problem"

Response: This has been changed.

- pg12, L277: Equation 8. Does this function yield n? Maybe do "n = argmin..."

Response: This equation does not yield n. it yields (TI)ˆ, which is the indexes of the most probable TIs.

- pg12, L279: "size n of TI in the Appendix" -> "size n of TI via Particle Swarm Optimization (PSO) in the Appendix".

Response: This has been changed in the revision.

- pg13, L283: Equation 9. This does not look like a Hausdorff distance formula like in Equation 1, more like a directed Hausdorff with no max argument. Please check that this formula is correct.

Response: Thanks for this attentive check. In Equation 1, we calculate the similarity between two TIs. By comparison, the consistency between TI and radar data is computed in Equation 9. In order to inspire a high-quality simulation, our program focuses on TI capability to provide a similar structure to radar data. Therefore, the max argument is removed in Equation 9.

- pg13, L285: "We use flexible sized templates". Please state the range (min and max) of template sizes used, e.g. 10x10, 20x20, etc.

Response: We added template size in Figure 7. The size of template is case-dependent. In this step, we apply a flexible template to gather conditioning data. Visiting a point in the simulation grid, our program collects 40 closest points and creates a conditioning pattern. Thus, the radius of template is not a fixed number.

- pg13, L286: "randomly chosen between the maximum radius up to 15 pixels to include 40 measurement points". This sentence carries ambiguity. Firstly, radius implies a circle, is that what was used, or was a rectangle used as in Figure 3? Secondly, does 'include 40 measurement points' mean that only 40 measurements points are needed and no more (i.e. =40), or that it is a minimum (i.e. >=40).

Response: This is a nice suggestion. In combination with the previous comment, we change the sentence as:

"The template size varies in order to include 40 neighboring line data points, but with maximum radius up to 15 pixels."

- pg13, L288: Figure 7 colorbar. Add units for the colorbar.

Response: Thanks for this recommendation. Like Figure 4, the TI and radar data have been processed by the normalization to remove unit. We revised the colorbar.

- pg13, L289: Figure 7 caption. "training image and radar line data" -> "training image (TI) and radar line data (d)".

Response: This has been changed in the revision.

- pg13, L291: "We use a particle swarm optimization (PSO) to minimize the distance function dis(.)" -> "We use particle swarm optimization (PSO) to minimize the distance dis(.) (see Appendix)".

Response: This has been changed in the revision.

- pg13, L292: "PSO has its specific advantages in requiring less parameterizations". Are the parameters you used in the particle swarm optimizer for this Amundsen Sea Embayment study applicable to other parts of Antarctica, or would people need to re-run things to find new optimal parameters? I.e. how stable are the parameters across different study areas?

Response: Parameterization is an important component in the PSO optimization. In order to ensure the fast convergence, we directly employ the recommended parameters tested by Rezaee Jordehi and Jasni et al (2013). These parameters are not only used by the Amundsen Sea Embayment and Antarctica topographic simulation, but adopted by a range of engineering applications. Therefore, it is not necessary to find other parameters in every new case.

- pg14, L298: Figure 8b colobar. Add units for the colorbar.

Response: Please see response for "- pg13, L288: Figure 7 colorbar."

- pg14, L298: Figure 8b axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

Response: See response for - pg6, L131

- pg14, L298: Figure 8b map. Is A1=T1-99, A2=TI-88 and so on? Not sure why the TI numbers are on the bottom of the map.

Response: This has been changed in the revision as following:

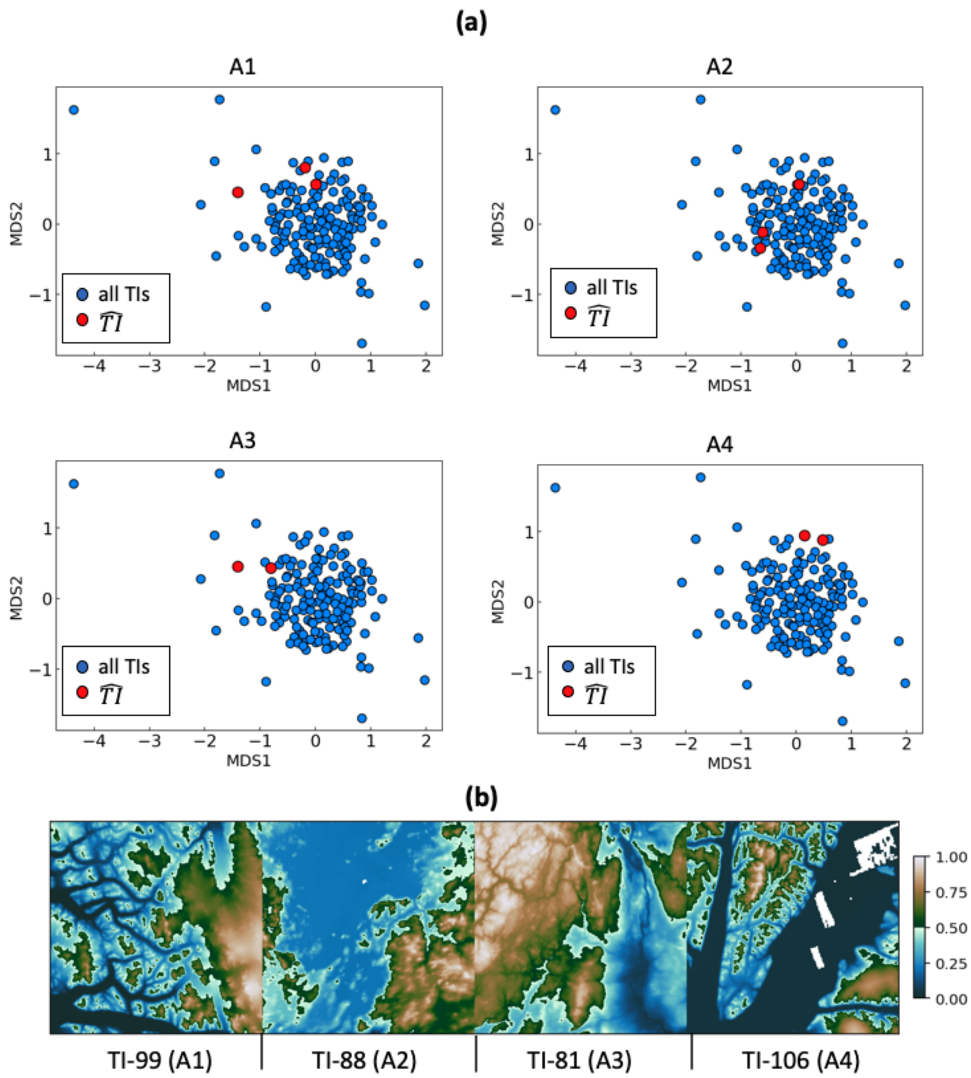


Figure 8. (a) The estimated set of most probable training images (TI)ˆfor each area displayed on MDS plots. The red dots highlight the estimated (TI)ˆ. (b) Examples of (TI)ˆ displayed in the topographic modeling space.

- pg14, L299. Figure 8 caption. "for each area displayed on MDS plots" -> "for each subarea (A1, A2, A3, A4) displayed on multi-dimensional scaling (MDS) plots".

Response: This has been changed

- pg14, L299-300. Figure 8 caption. "The red dots highlight the estimated TI" -> "The blue dots represent a TI, and the red dots highlight the estimated TI". Also, why do some subareas have two red dots while some has three?

Response: As mentioned above, the optimal number of TIs in each area (A1, A2, A3, A4) is determined by a profile log-likelihood function. Therefore, our program allocates three TIs to A1 and A2. By contrast, A3 and A4 separately get two TIs. We revised the caption of Figure 8.

Section 3.4.2: Kernel density estimation of P(TI(A_i)|d_{A_i})

- pg15, L303: "We assume that the TIs" -> "We assume that TIs".

Response: This sentence is improved below:

"We assume that TIs near the (TI)ˆon MDS plot (Figure 5) tend to have similarly high probability of being assigned to the radar data subarea"

- pg15, L304: "(Figure 4)" -> "(Figure 5)".

Response: This has been changed

- pg15, L305: "we can observe the spatial patterns of nearby TIs look similar" -> "we can observe that spatial patterns of nearby TIs look similar". Please define how it looks 'similar'. Is it a qualitative similarity, i.e. the topographic structure looks similar, or something else?

Response: The similarity in this circumstance indicate the difference between morphological and geometrical features. Therefore, we improve this sentence into:

"This is because nearby TIs in the MDS metric space (see Figure 5) show similar morphological patterns."

- pg15, L306: "predict the probability to each TI". Probability of what to each TI? Please state.

Response: In our method, we use the kernel density estimation to predict the probability that our program selects a TI to model a simulation area.

With the aim of better explain our method, we add more description in this sentence.

"We therefore consider a Gaussian kernel density estimation (KDE) to predict the probability of TI being assigned to a subarea A_i."

- pg15, L311: "dis(TI, TI_k) is distance between a TI and" -> "dis(TI, TI_k) is the modified Hausdorff distance between a TI and".

Response: This has been revised.

- pg15, L314: Figure 9 colorbar. Add units for the colorbar.

Response: The colorbar in the Figure 9 reveal the prior P(TI(A_1)|d_(A_1))of each TI with respect to the radar measurements in the subarea A1. Therefore, these values do not have a unit. We added more explanations in the figure.

- pg15, L316: "(a) Estimated TI for A1" -> "(a) Estimated TI for subarea A1".

Response: This has been revised.

- pg15, L316: "plotted in MDS space" -> "plotted in multi-dimensional scaling (MDS) space".

Response: Since the definition of MDS has already introduced in Section 3.1.3, we improve the caption of Figure 9 as follows:

“Figure 9. (a) Estimated (TI)ˆfor subarea A1 in MDS space.”

- pg15, L316: "The red dots are" -> "The blue dots are training images (TI) and red dots are".

Response: This has been revised.

Section 3.5 Aggregation by weighting log-ratios

- pg16, L323: "two areas" -> "two subareas".

Response: This has been revised.

- pg16, L327: "between area" -> "between subarea"

Response: This has been revised.

- pg16, L328: "x_l and x_l' are the radar data locations". Clarify what this means. Are they point (x,y) or pixel (image) coordinates?

Response: These two variables are the coordinates of radar points. We refined this sentence in the following:

“x_l and x_l' are radar data point locations in A_i and A_j.”

- pg16, L328: "Epanechnikov kernel". Need to provide a citation, or better, state the formula.

Response: The related reference has been added .

- pg16, L329-330: "the weights w_{ij} are simply" -> "the weights w_{ij} are"

Response: This has been revised.

- pg16, L332: Equation 13. "where i, j = [1,2,3,4]". Does this mean i or j can be 1,2,3 or 4? I.e. 4x4 = 16 possibilities? I feel like there is a better way to state this mathematically. Alternatively, you could take this out of Equation 13 and describe i and j in-text at L334.

Response: i and j are indices in the Equation 13. Considering the value of an index can be 1, 2, 3 and 4, the possible values of w_ij are 16. However, the distance disim(A_i,A_j) is a symmetrical metric. In other words, disim(A_i,A_j)=disim(A_j,A_i). Moreover, the distance disim(A_i,A_j)=0 when i=j. Therefore, there are 6 possible values of weights w_ij.

- pg17, L338: Figure 10. (Optional) Would be nice to plot the A1-A4 images on the top right corner of each panel, or at least refer to the subarea images in the caption, i.e. link to Fig 8b.

- pg17, L338: Figure 10 axes. Please spell out PDF in full.

Response: We improved Figure10 as suggested:

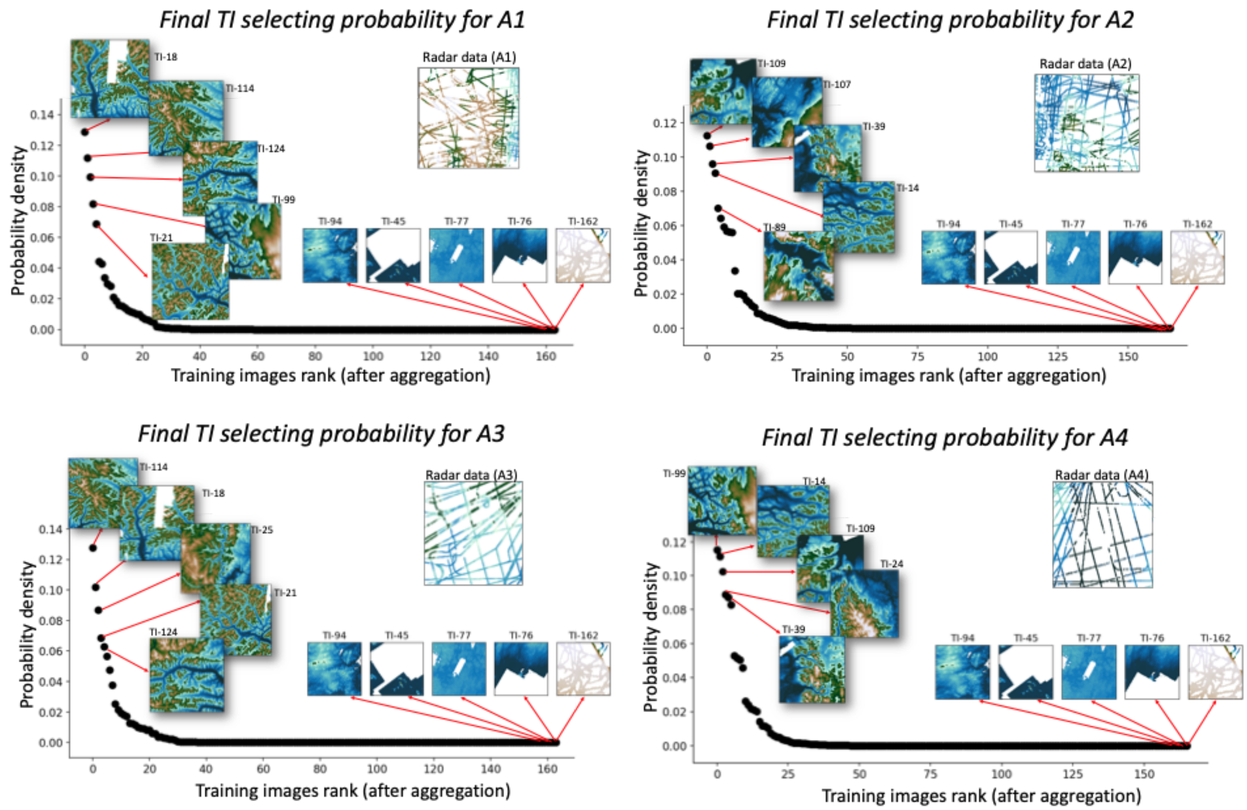


Figure 10. Probability distribution of final aggregated TI probability in each radar line subarea.

- pg17, L339: Figure 10 caption. "aggregated TI" -> "aggregated training image (TI)"

Response: We did not change this word because the definition of TI has been widely used in the preceding section.

Section 3.6 Direct sampling with TI sampling

- pg17, L344: "tend to have higher elevations". How many metres higher? 2000m? 3000m?

Response: Figure 11 provides a conceptual model to display the training image selected by our method. However, the values in TIs is normalized. Therefore, we cannot accurately state the elevation.

- pg17, L344: "larger scale low-elevation valleys". How many metres deep, and how many metres wide?

Response: See above response.

- pg17, L345: "that can be related to warm water routing". This is contentious, the valleys could just be due to ice stream action.

Response: This has been removed.

- pg17, L346: "At the end, multiple realizations of topographical models are generated with" -> "At the end of the simulation, multiple realizations of topography are generated, each with"

Response: This has been changed.

- pg18, L351: Figure 11 caption. "Examples of sampled TIs from the posterior distributions" -> "Examples of sampled training images (TIs) from the posterior distribution corresponding to each subarea (A1, A2, A3, A4) for each realization". You might need to rephrase this sentence a bit to explain the figure better.

Response: This has been rephrased.

- pg19, L353. Figure 12 axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

Response: This is a demo example, so we have added a scale bar.

- pg19, L353: Figure 12 colorbar. Add units for the colorbar.

- pg19, L355: Figure 12 caption. "DS" -> "direct sampling (DS)"

- pg19, L355-356: Figure 12 caption. "Model realization number corresponds to the TI realization number in Figure 11". I think you mean "Model realization number (#n) corresponds to the model realization number in Figure 11"?

Response: For the above 3 comments: we have re-made a complete new Figure 12. Please see below

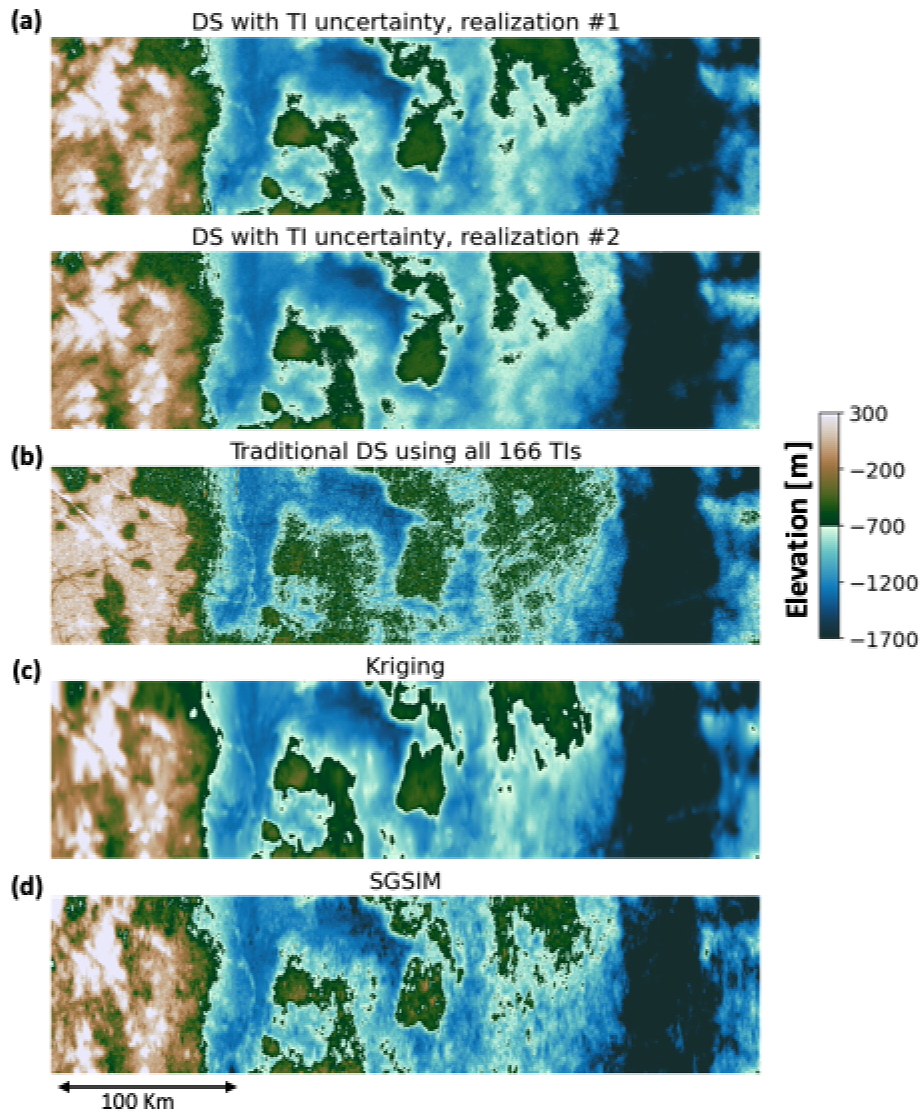


Figure 12. (a) Two topographical model realizations from using our proposed DS with uncertain TI selection to fill the radar lines gaps. Model realization number corresponds to the TI realization number in Figure 11. (b) line gaps filling by traditional DS using all the 166 TIs (without TI sampling). (c) and (d) line gaps filling using kriging and SGSIM.

Section 3.7 Comparison with traditional MPS modeling and two-point geostatistical modeling

- pg19, L360-372: This subsection seems to be interpreting some results, and you may want to consider moving it into a 'Discussion' section for the final publication.

Response: We think this section fits better in the current Methodology section, because it evaluates the performances of the proposed method.

- pg19, L361: "scanning all the 166 TIs" -> "scanning all 166 TIs".

This has been changed.

- pg19, L361-362: "Figure 13a shows one realization of the simulated result. It is obvious that the conventional approach results in a much noisier topographical model. There are significant" -> "Figure 13a shows one realization of the conventional MPS simulated result. The conventional approach appears to show a much noisier topographical model, and there are significant"

Response: This is a good correction. We have revised the sentences.

- pg19, L362-363: "we take a cross-section A-A' on the Pine Island glacier and plot the comparison" -> "we take a cross-section A-A' across the trunk of Pine Island Glacier and plot the elveation comparison".

This has been changed.

- pg19, L364-365: "We can observe that the DS without TI sampling creates a significant amount unrealistic" -> "We observe that the DS without TI sampling methods generated more unrealistic"

Response: Thank you for this comment. We modify this sentence as follows:

"We can observe that the DS without TI sampling creates more unrealistic elevation peaks and troughs."

- pg19, L365: "Pine Island glacial" -> "Pine Island Glacier".

This has been changed.

- pg19, L366: "where mostly radar data" -> "where dense radar data".

This has been changed.

- pg19, L367-368: "when using all the 166 TIs without proper sampling, the DS finds too large a set of patterns likely many incompatible with the sparse data" -> "when using all 166 TIs without proper sampling, DS without TI sampling finds too large a set of patterns that is likely incompatible with sparse data".

This has been changed.

- pg19, L369: "thereby improving the result" -> "thereby improving the result (see Fig 12)".

This has been changed.

- pg19, L369-370: "avoiding the channel artifacts is critically important for modeling subglacial hydrological flow (see section 0)" -> "avoiding topographic artifacts in subglacial channels is important for modeling subglacial hydrological flow (see section 4.2)".

Response: This has been changed.

- pg19, L371: "to simulate one realization. When using our TI sampling approach, it took less than 1 hour" -> "to simulate one realization, compared to 1 hour for our TI sampling approach".

Response: This has been changed: ""In terms of running time, the conventional DS with 166 TIs took nearly 21 hours to simulate one realization. By contrast, our TI sampling approach took less than 1 hour.""

- pg19, L372: "are run on a PC with an Intel i9-11900 of 2.5GHz processor and 32GB of RAM" -> "were ran using an Intel i9-11900 (2.5GHz) processor with 32GB of RAM".

Response: This has been changed.

- pg20, L374: "compare to the two-point geostatistical modeling with kriging and Sequential Gaussian Simulation (SGSIM)" -> "compare the two-point geostatistical modeling results (Fig 12) with kriging (Fig 13b) and Sequential Gaussian Simulation (SGSIM; Fig 13c)".

Response: This has been changed.

- pg20, L375: "Figure 13b and Figure 13c plot topographic modeling results from kriging and SGSIM" -> "".

Response: This sentence is refined as follows:

"We further compare our approach with the two-point geostatistical modeling methods, kriging and Sequential Gaussian Simulation (SGSIM) (Figure 12c and 12d)"

- pg20, L375-376: "We can observe that kriging produces the most smoothed topographical model" -> "We observe that kriging produces the smoothest topography".

This has been changed.

- pg20, L376: "are very clear" -> "are clear".

This has been changed.

- pg20, L377-378: "After all, kriging is a deterministic modeling approach. Thus, it cannot capture location scale elevation variations and quantify the spatial uncertainty" -> "After all, kriging is a deterministic modeling approach which cannot capture variations in local scale elevation, nor can it quantify the spatial uncertainty".

We have rephrased this statement.

- pg20, L378: "quantify the spatial uncertainty". Note that kriging can output variance, which could be a measure of uncertainty.

This is a nice suggestion. This sentence is changed into:

“It cannot generate multiple topographical models to quantify the spatial uncertainty.”

- pg20, L378: "Our SGSIM" -> "The SGSIM".

This has been changed.

- pg20, L379: "limited the neighborhood" -> "limiting the model to the neighborhood".

Response: This has been changed.

- pg20, L379-380: "The limitation of SGSIM, an approach based on spatial covariances, lies on its limitations in capturing" -> "However, SGSIM, an approach based on spatial covariances, is limited in its ability to capture".

Response: This has been revised by adding a new Figure 14, also to address the other reviewers’ comments.

- pg20, L382: "using the four different approaches". Technically the radar is the groundtruth, and you are comparing three different approaches. Please reword the sentence.

Response: This sentence is corrected as:

“Furthermore, we compare the empirical variograms from radar data as well as the simulated topographical maps using three different approaches.”

- pg20, L382: "the DS using sampled TIs has" -> "the DS using sampled TIs method has"

- pg20, L384: "the DS without TI sampling has" -> "the DS without TI sampling method has"

Response: Response to the above two comments: our sentence means the same thing, so we didn’t change it.

- pg21, L386: Figure 13 map. It would be nice if you can combine Fig 12 and Fig 13 so that the different methods (MPS, Kriging, SGSIM) could be directly compared in one plot. Even better if you can plot the groundtruth data lines (e.g. from Fig 15a) too.

Response: We have re-created a new Figure 13 to address the reviewer's comment.

- pg21, L386. Figure 13 axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

Response: See response to "- pg19, L353. Figure 12 axes. "

- pg21, L386: Figure 13 colorbar. Add units for the colorbar.

Response: Colorbar has been revised.

- pg21, L387: Figure 13 caption. Please mention that the black ovals on Fig 13c refers to areas of sparse data.

Response: This has been changed.

- pg22, L388: Figure 14 top map. Please include a colorbar, and refrain from using jet as a colormap, choose a Scientific Colour Map instead.

- pg22, L388: Figure 14 top map. Can you put the dashed box from Fig 14 bottom (the transect plot) in the top map too?

- pg21, L388. Figure 14 top axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

- pg22, L388: Figure 14 bottom plot. Don't use red and green for the lines as they are not good for people with Deuteranopia. Also, suggest changing the black colour for the kriging line to a different colour to avoid confusion with the radar data points. Try https://colorbrewer2.org for suggestions on what to use.

- pg22, L389: Figure 14 caption. Mention that the dashed box shows main channel of Pine Island Glacier.

Response to all the above comments on Figure 14: Thanks for your suggestions. We have re-made the Figure 14 following the reviewer's suggestion, including changing the colormap to color-blind friendly and revising captions. We added a scale bar for reference.

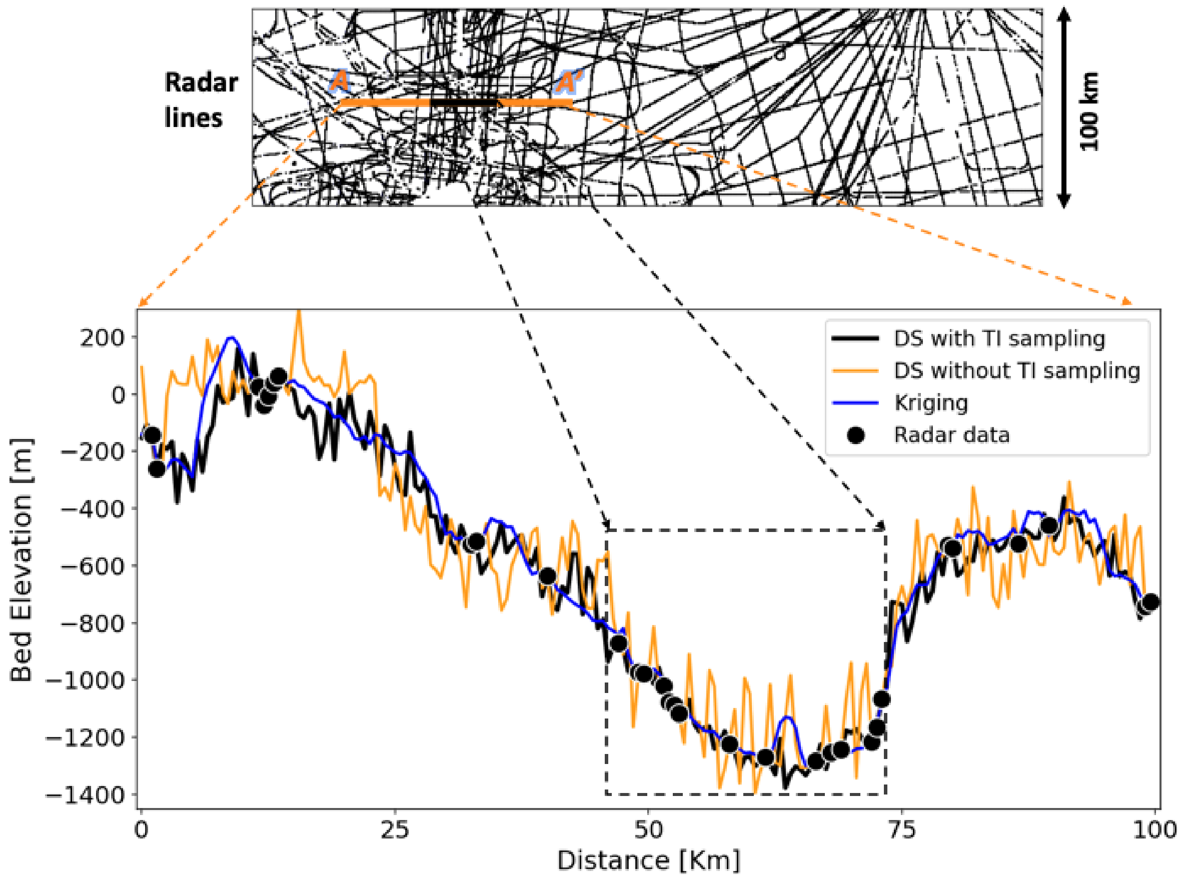


Figure 13. Cross-section view of the modeled topography maps at line A-A’. The black dashed box shows main channel of Pine Island Glacier

- pg22, L390: Figure 15. Don't use red and green for the lines.

- pg22, L390: Figure 15. For the SGSIM line (blue), which realization are you using? 1 or 2?

- pg22, L390: Figure 15. Could you describe in text, why the lines deviate away from the radar data more at 70-80km? Is it because of sparser radar data coverage?

- pg22, L390: Figure 15 axes. Why does the x-axis stop at 80km, and not continue to 100km like in Fig 14?

Response to the above comments on Figure 15:

We have re-made the Figure 15 following the reviewer’s suggestion.

We could continue to 100km and above in the variogram. But it does not reveal any further meaningfully spatial correlations to the current plot.

Section 4 Application to the entire Amundsen Sea Embayment (ASE)

Section 4.1 Training image sampling and DS simulation

- pg23, L395: "We apply the methodology to" -> "We apply the methodology in Section 3 to".

Response: This has been changed.

- pg23, L395: "high-resolution topography" -> "high-resolution (500 m) topography".

Response: This has been changed.

- pg23, L397: "diving" -> "dividing".

This has been changed.

- pg23, L398: "Equally divide the ASE area into four subareas". By four subareas, do you mean quadrants? Or 4 long strips? Are the subareas always square, or can the be rectangular?

Response: Yes. This has been changed.

- pg23, L399-401: The loop in Step 2 and 3 says "continue to divide it into four equally sized areas... until the threshold", does this mean that subarea sizes are not fixed at 100km x 100km? Does the model work with different sized subareas?

Response: Exactly, the length of subarea is not constant. Our method is capable of dealing with subareas with varying size. The areas with dense radar data are recursively divided until the number of radar points is lower than a predefined threshold.

- pg23, L400: "Repeat step 2 until amount of data" -> "Repeat step 2 until the amount of data".

Response: This has been changed.

- pg23, L401: "into totally of 56 subareas" -> "into a total of L=56 subareas".

This has been changed.

- pg23, L403: "Figure 16 shows the final ASE subareas with the corresponding radar data density". Please specify how radar data density is calculated. Is it the number of non-NaN pixels over an area?

Response: The radar data density is calculated by dividing the number of radar points by the area of the subarea of interest. The unknown points do not participate the density computation.

- pg23, L409: Figure 16 map. Need a colorbar for the radar elevation data, and I highly encourage using a Scientific Colour Map instead of jet.

Response: We have remade Figure 16 and changed the colormap

- pg23, L409. Figure 16 axes. Please show x/y (or longitude/latitude) coordinates for this map with units, instead of using image coordinates.

Response: Coordinates have been added

- pg24, L411. Figure 17 axes. Please show x/y (or longitude/latitude) coordinates for this map with units, instead of using image coordinates.

Response: Coordinates have been added

- pg24, L411: Figure 17 colorbar. Add units for the colorbar.

- pg24, L411: Figure 17 caption. "TIs assigned to the entire ASE area" -> "training images (TIs) assigned to the entire Amundsen Sea Embayment (ASE) area"

- pg25, L413. Figure 17 axes. Please show x/y (or longitude/latitude) coordinates for this map with units.

- pg25, L413. Figure 17 colorbar. Add units for the colorbar. Also, find a way to represent NaN values properly instead of as white colour.

- pg25, L411: Figure 17 caption. "ASE" -> "Amundsen Sea Embayment (ASE)".

- pg25, L413: Figure 17 caption. "circles" -> "ovals".

Response to all the above comments on Figure 17: We have revised the figure by changing the colorbar and caption. Coordinates also have been added

Section 4.2 Uncertainty in subglacial hydrological flow

- pg26, L419: "model was applied to 20 realizations to model" -> "model was applied to the 20 topography realizations in Section 4.1 to model".

This has been changed.

- pg26, L419:420: "The direction of water flow is determined by calculating hydraulic potential, phi, using the Shreve equation (Shreve, 1972)" -> "The direction of subglacial water flow can be determined by the gradient of hydraulic potential, phi, which is calculated using the following equation (Shreve, 1972)".

This has been changed.

- pg26, L424: "100 kg m-3" -> "1000 kg m-3"!!

Response: Thanks. This has been changed.

- pg26, L424: "gravitational acceleration" -> "gravitational acceleration (9.8 m s-1)". Or whatever value you used.

This has been changed.

- pg26, L425: "The hydrological model is implemented" -> "The subglacial hydrological model was implemented".

This has been changed.

- pg26, L425-426: "These functions use the hydraulic potential gradient to compute flow accumulation, or the number of pixels that flow into another pixel". Describe the algorithm briefly for someone unfamiliar to the tools, e.g. is it using a D8 routing model?

Response: We do not use D8. We use the multiple flow directions (MFD) model, which is stated in the text. We have added additional details.

- pg26, L428: "We assume spatially uniform basal melt rates". Specify the basal melt rate quantity used.

Response: The flow algorithm is based on pixel contribution, so melt rates are not explicitly used. The uniform melt rate assumption means that all pixels are weighted equally.

- pg26, L428: "water pressure" -> "subglacial water pressure".

This has been changed.

- pg26, L432: "grounding line" -> "grounding zone".

This has been changed.

- pg26, L433: "These tributaries are located near a system" -> "Some of these tributaries are located over a system".

This has been changed.

- pg26, L434: "(Smith et al., 2017)". Please also cite more recent Thwaites Glacier active subglacial lake papers by Hoffman et al., 2020 and Malczyk et al., 2020.

This has been changed.

- pg26, L434-435: "with a drainage and refill cycle" -> "with a drain and refill cycle".

This has been changed.

- pg26, L435-436: "Lake drainage events are associated with increases in ice velocity (Stearns et al., 2008), making it important to characterize the connectivity of active lake systems". This is stated too simplistically. Lake drainage doesn't always cause ice speedup, see Section 4.2 of Smith et al., 2017. You can however, say that "Lake drainage events are sometimes associated with increases in ice velocity", though I recommend mentioning the caveat that lake drainage may not necessarily lead to ice speedup.

This has been changed.

- pg26, L436: "(Stearns et al., 2008)". Please consider reading and citing other relevant papers such as Wright et al., 2014 and Fricker et al., 2016.

This has been added.

- pg26, L436: "Our results" -> "Our topography modelling results in Section 4.1".

This has been changed.

- pg26, L436-437: "additional observational constraints" -> "additional radar data observational constraints".

Response: We have left this as is. These findings could be used to inform survey decisions from a number of instruments.

- pg26, L436-438: You will need to better quantify and justify why a better bed topography is needed, since ice surface elevation is 11x more important than bed elevation for hydropotential calculations (which can be derived from Equation 14). Otherwise it will be hard to argue your case that additional bed elevation observations are needed for deriving better hydropotential maps when ice surface elevation is the more important driving force.

Response: It is correct that the ice surface has a greater influence on hydropotential calculations than topography. However, the uncertainties in topography are orders of magnitude greater than uncertainties or temporal variations in the ice surface, so topography is the dominating source of uncertainty in hydrological models. We have added the following sentences to clarify our main scientific takeaways:

”The average of the hydrological models across different realizations is different from the hydrological model made using the mass conservation topography, particularly in the main trunk of Thwaites Glacier (Figure 19d and 19e). This demonstrates that deterministic DEMs cannot be used to sample the range of possible flow path locations, which could lead to the misinterpretation of hydrological conditions. In contrast, geostatistical simulation provides a framework for quantifying hydrological uncertainty with respect to topographic uncertainty.”

- pg27, L439: Figure 19 axes. Please show x/y (or longitude/latitude) coordinates for these maps with units.

Response: Coordinates are the same as Figure 18.

- pg27, L439: Figure 19 maps. Mention in the caption where you obtained the glacier boundaries, provide a data citation if possible.

Response: Boundary is the same as Figure 1. We don't prefer to repeat this again.

- pg27, L439: Figure 19a map. Is the Mass conservation map derived using BedMachine Antarctica data? If so, please mention in the caption and cite accordingly.

This has been done.

- pg27, L439: Figure 19e map. You have shown the mean flow accumulation which is good, but maybe show a standard deviation plot as well to see how the different realizations differ, and where they differ the most. Also, it would be nice to show a difference map between Fig19d and Fig19e to highlight the uncertainty of the subglacial water flow model between the mass conservation map and your different realizations.

Response: We appreciate the reviewer's in-depth suggestion. We are currently engaged in hydrological modeling projects that build on this work. These analyses will take a more in-depth look at the hydrological uncertainties in the ASE and the scientific implications, but are beyond the scope of this study.

Section 5 Conclusions

- pg27, L446-447: "fill large-scale geophysical data gaps and applied it to map high-resolution subglacial topography in West Antarctica" -> "fill geophysical data gaps and applied it to produce a high-resolution (500 m) subglacial topography map of the Amundsen Sea Embayment".

This has been revised.

- pg28, L448: "high-resolution topographic" -> "high-resolution (500 m) topographic".

This has been revised.

- pg28, L449-450: "from the Arctic and Antarctica. These training images represent" -> "from deglaciated regions in the Arctic and Antarctica, which represent".

This has been revised.

- pg28, L450: "We have placed them in" -> "We have placed the training images (TIs) in".

This has been revised.

- pg28, L450: "publicly available repository". Link to the repository here, or point to the Data and code availability section.

This has been revised.

- pg28, L457: "based on the distance" -> "based on the modified Hausdorff distance".

This has been changed.

- pg28, L459: "posterior TI distribution allowed us" -> "posterior TI distribution then enabled us".

This has been revised.

- pg28, L459: "direct sampling" -> "direct sampling (DS)".

This has been revised.

- pg28, L461: "Such non-stationary TI sampling framework avoided the use" -> "Such a non-stationary TI sampling framework avoids the use".

This has been changed.

- pg28, L461-462: "It significantly improved" -> "It has significantly improved".

This has been changed.

- pg28, L462: "reduced the DS running time" -> "reduced the DS running time from 21 hours to 1 hour".

This has been revised.

- pg28, L470: "across realizations" -> "across different realizations".

This has been revised.

- pg28, L470: "tributaries are near a system" -> "tributaries cross a system".

This has been changed.

- pg28, L471-472: "and could have the potential to influence ice sheet velocity. The high hydrological uncertainty in this area highlights the need for additional measurement constraints". Similar to my point made above for pg26, L436-438, you will need to argue your case better to arrive at this conclusion, because ice surface elevation is 11x more important than bed elevation for calculating hydropotential, so additional measurement constraints (assuming you mean radar bed data) will not affect hydropotential as much as better ice surface elevation data from satellite altimeters. My suggestion is that you can also mention why high-resolution (<500 m) fine-scale topography matters, see e.g. Kyrke-Smith et al., 2018 and related papers.

Response: Please see response to comment for L436-438.

Section Appendix

- pg30, L520: Figure 20a axes. The y-axis needs units for distance, is it metres? Also, the x-axis says "size n of TI", but I think you mean "swarm size n" according to the caption.

Response: The distance is a measure of similarity between a training image set and radar data in a subarea. Since input values have been normalized, there is no unit for the distance metric.

In addition, the axis in Figure 20a and 20b indicate the size of selected TIs. It is clear the similarity decreases with the number of TIs. The reason lies in that an increasing number of TIs is capable of providing various spatial structures to our program. However, numerous TIs considerably extend the searching scope and have a negative influence on the computational efficiency. Therefore, we apply a profile log likelihood function to find a balance between simulation quality and running time.

The swarm size plays an important role in the particle swarm optimization (PSO). According to the investigation conducted by Rezaee Jordehi and Jasni (2013), the number of particles should increase with the dimension of problem at hand. Based on the study performed by Piotrowski et al. (2020), the recommended value is range from 70-500. Therefore, we set the swam size m=10×n, where n is the number of candidate TI^.

Section Code and Data availability

- pg30, L524: "training image database". The training image (TI) database of 166 images was stored as a single text file. In order to be more user friendly (see FAIR guidelines by Wilkinson et al., 2016), I strongly advice that you 1) separate the TIs to be individual files, one for each image, so that people don't need to figure out how to reshape the arrays; 2) use an alternative file format like NetCDF or GeoTIFF, so that you can have geographical coordinates and other metadata associated with each TI, especially important for the sake of data provenance (where was the TI sourced from? The Arctic or Antarctic?). Also, you have not included any of your output topography or subglacial flow model realizations on Zenodo, but I expect those to be released in time for the final publication.

Response: We thank the reviewer for the suggestions. We notice most of the above comments concern the data formats in our public repository. We provided the data and code as a supplement for this submitted research "development and technical paper". We think using different data formats (here ".txt") is not a critical issue, as long as we provide descriptions and codes to access them. Besides, we do not think it's proper to include model realizations to the Zenodo training image repo. It's because these realizations are training image derived models with high uncertainty. They cannot be used again as training images.

- pg30, L525: "modelling codes" -> "modelling code".

Response: This has been changed.

- pg30, L526: "https://github.com/sdyinzhen/MPS-BedMappingV1". I highly appreciate the source code being made available. For the sake of reproducibility, could you provide a dependency specification (e.g. a pip requirements.txt or conda environment.yml) file so that others know what libraries you have used?

Response: We have added more descriptions on the Github code repository.

References

Response: References have been added to our revision.

Aitken, A. R. A., Young, D. A., Ferraccioli, F., Betts, P. G., Greenbaum, J. S., Richter, T. G., Roberts, J. L., Blankenship, D. D., & Siegert, M. J. (2014). The subglacial geology of Wilkes Land, East Antarctica. Geophysical Research Letters, 41(7), 2390–2400. https://doi.org/10.1002/2014GL059405

Anandakrishnan, S., Blankenship, D. D., Alley, R. B., & Stoffa, P. L. (1998). Influence of subglacial geology on the position of a West Antarctic ice stream from seismic observations. Nature, 394(6688), 62–65. https://doi.org/10.1038/27889

Bell, R. E., Blankenship, D. D., Finn, C. A., Morse, D. L., Scambos, T. A., Brozena, J. M., & Hodge, S. M. (1998). Influence of subglacial geology on the onset of a West Antarctic ice stream from aerogeophysical observations. Nature, 394(6688), 58–62. https://doi.org/10.1038/27883

Bingham, R. G., Vaughan, D. G., King, E. C., Davies, D., Cornford, S. L., Smith, A. M., Arthern, R. J., Brisbourne, A. M., De Rydt, J., Graham, A. G. C., Spagnolo, M., Marsh, O. J.,

& Shean, D. E. (2017). Diverse landscapes beneath Pine Island Glacier influence ice flow. Nature Communications, 8(1), 1618. https://doi.org/10.1038/s41467-017-01597-y

Crameri, F., Shephard, G. E., & Heron, P. J. (2020). The misuse of colour in science communication. Nature Communications, 11(1), 5444. https://doi.org/10.1038/s41467-020-19160-7

De Fleurian, B., Werder, M. A., Beyer, S., Brinkerhoff, D. J., Delaney, I., Dow, C. F., Downs, J., Gagliardini, O., Hoffman, M. J., Hooke, R. L., Seguinot, J., & Sommers, A. N. (2018). SHMIP The subglacial hydrology model intercomparison Project. Journal of Glaciology, 64(248), 897–916. https://doi.org/10.1017/jog.2018.78

Farinotti, D., Brinkerhoff, D. J., Clarke, G. K. C., Fürst, J. J., Frey, H., Gantayat, P., Gillet-Chaulet, F., Girard, C., Huss, M., Leclercq, P. W., Linsbauer, A., Machguth, H., Martin, C., Maussion, F., Morlighem, M., Mosbeux, C., Pandit, A., Portmann, A., Rabatel, A., … Andreassen, L. M. (2017). How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment. The Cryosphere, 11(2), 949–970. https://doi.org/10.5194/tc-11-949-2017

Fricker, H. A., Siegfried, M. R., Carter, S. P., & Scambos, T. A. (2016). A decade of progress in observing and modelling Antarctic subglacial water systems. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2059), 20140294. https://doi.org/10.1098/rsta.2014.0294

Goff, J. A., Powell, E. M., Young, D. A., & Blankenship, D. D. (2014). Conditional simulation of Thwaites Glacier (Antarctica) bed topography for flow models: Incorporating inhomogeneous statistics and channelized morphology. Journal of Glaciology, 60(222), 635–646. https://doi.org/10.3189/2014JoG13J200

Graham, F. S., Roberts, J. L., Galton-Fenzi, B. K., Young, D., Blankenship, D., & Siegert, M. J. (2017). A high-resolution synthetic bed elevation grid of the Antarctic continent. Earth System Science Data, 9(1), 267–279. https://doi.org/10.5194/essd-9-267-2017

Hoffman, A. O., Christianson, K., Shapero, D., Smith, B. E., & Joughin, I. (2020). Brief Communication: Heterogenous thinning and subglacial lake activity on Thwaites Glacier, West Antarctica [Preprint]. Glaciers/Subglacial Processes. https://doi.org/10.5194/tc-2020-80

King, E. C., Pritchard, H. D., & Smith, A. M. (2016). Subglacial landforms beneath Rutford Ice Stream, Antarctica: Detailed bed topography from ice-penetrating radar. Earth System Science Data, 8(1), 151–158. https://doi.org/10.5194/essd-8-151-2016

Kyrke-Smith, T. M., Gudmundsson, G. H., & Farrell, P. E. (2018). Relevance of Detail in Basal Topography for Basal Slipperiness Inversions: A Case Study on Pine Island Glacier, Antarctica. Frontiers in Earth Science, 6, 33. https://doi.org/10.3389/feart.2018.00033

Le Brocq, A. M., Ross, N., Griggs, J. A., Bingham, R. G., Corr, H. F. J., Ferraccioli, F., Jenkins, A., Jordan, T. A., Payne, A. J., Rippin, D. M., & Siegert, M. J. (2013). Evidence from ice shelves for channelized meltwater flow beneath the Antarctic Ice Sheet. Nature Geoscience, 6(11), 945–948. https://doi.org/10.1038/ngeo1977

Lowry, D. P., Golledge, N. R., Bertler, N. A. N., Jones, R. S., McKay, R., & Stutz, J. (2020). Geologic controls on ice sheet sensitivity to deglacial climate forcing in the Ross Embayment, Antarctica. Quaternary Science Advances, 1, 100002. https://doi.org/10.1016/j.qsa.2020.100002

Malczyk, G., Gourmelen, N., Goldberg, D., Wuite, J., & Nagler, T. (2020). Repeat Subglacial Lake Drainage and Filling beneath Thwaites Glacier. Geophysical Research Letters. https://doi.org/10.1029/2020GL089658

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., Catania, G., Chauché, N., Dowdeswell, J. A., Dorschel, B., Fenty, I., Hogan, K., Howat, I., Hubbard, A., Jakobsson, M., Jordan, T. M., Kjeldsen, K. K., Millan, R., Mayer, L., … Zinglersen, K. B. (2017). BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation. Geophysical Research Letters, 44(21), 11,051-11,061. https://doi.org/10.1002/2017GL074954

Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., Eisen, O., Ferraccioli, F., Forsberg, R., Fretwell, P., Goel, V., Greenbaum, J. S., Gudmundsson, G. H., Guo, J., Helm, V., Hofstede, C., Howat, I., Humbert, A., Jokat, W., … Young, D. A. (2019). Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet. Nature Geoscience, 13(2), 132–137. https://doi.org/10.1038/s41561-019-0510-8

Robin, G. D. Q., Swithinbank, C. W. M., & Smith, B. M. E. (1970). Radio echo exploration of the Antarctic ice sheet. In A. J. Gow, C. Keeler, C. C. Langway, & W. F. Weeks (Eds.), International Symposium on Antarctic Glaciological Exploration (ISAGE) (Vol. 86, pp. 97–115). International Association of Scientific Hydrology. https://iahs.info/uploads/dms/086017.pdf

Rougier, N. P., Droettboom, M., & Bourne, P. E. (2014). Ten Simple Rules for Better Figures. PLoS Computational Biology, 10(9), e1003833. https://doi.org/10.1371/journal.pcbi.1003833

Schlegel, N.-J., Seroussi, H., Schodlok, M. P., Larour, E. Y., Boening, C., Limonadi, D., Watkins, M. M., Morlighem, M., & van den Broeke, M. R. (2018). Exploration of Antarctic Ice Sheet 100-year contribution to sea level rise and associated model uncertainties using the

ISSM framework. The Cryosphere, 12(11), 3511–3534. https://doi.org/10.5194/tc-12-3511-2018

Shi, L., Allen, C. T., Ledford, J. R., Rodriguez-Morales, F., Blake, W. A., Panzer, B. G., Prokopiack, S. C., Leuschen, C. J., & Gogineni, S. (2010). Multichannel Coherent Radar Depth Sounder for NASA Operation Ice Bridge. 2010 IEEE International Geoscience and Remote Sensing Symposium, 1729–1732. https://doi.org/10.1109/IGARSS.2010.5649518

Siegert, M. J., Ross, N., & Le Brocq, A. M. (2016). Recent advances in understanding Antarctic subglacial lakes and hydrology. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2059), 20140306. https://doi.org/10.1098/rsta.2014.0306

Sommers, A., Rajaram, H., & Morlighem, M. (2018). SHAKTI: Subglacial Hydrology and Kinetic, Transient Interactions v1.0. Geoscientific Model Development, 11(7), 2955–2974. https://doi.org/10.5194/gmd-11-2955-2018

Wright, A. P., Young, D. A., Bamber, J. L., Dowdeswell, J. A., Payne, A. J., Blankenship, D. D., & Siegert, M. J. (2014). Subglacial hydrological connectivity within the Byrd Glacier catchment, East Antarctica. Journal of Glaciology, 60(220), 345–352. https://doi.org/10.3189/2014JoG13J014

Point-to-point responses to RC3

This manuscript addresses an important question: the interpolation of sparse data in a non-stationary context. It develops a framework that is seen as generic and demonstrates it in the specific context of interpolating subglacial bedrock topography. The novelty is the use of a large number of training images coming from a variety of deglaciated areas, and to devise a scheme to select a subset of them that is specific to each simulated region.

My overall assessment is that the manuscript is scientifically sound and will have an impact in the glaciology community, and also possibly more generally in other applications of MPS to large-scale problems, including space-time simulation. The tests in section 4 clearly show the value of the approach. Importantly, this study improves our understanding of the usage of MPS in the case of very large models that are fed by lots of data and training images. It demonstrates the need to select a specific subset of training images for each region rather than using the entire training image database for all areas. This strategy provides gains in terms of quality and computation, which is very valuable. Also, it could fuel the discussion on using machine learning approaches to address such problems (e.g. GANs) which would require a new training for each sub-area, making them inefficient.

This said, I do have remarks on some technical aspects of the proposed approach that I outline below. Despite these comments, I find the method convincing and I believe the manuscript deserves publication in GMD.

Response: We highly appreciate the reviewer's positive evaluation on our paper. The remarks and suggestions are also very constructive to improve the paper. They therefore have been incorporated into the newly revised manuscript. Changes are highlighted in red in the revised manuscript. Please see below our point-to-point responses.

- While the approach is novel, some of the existing literature was missed. For example, on l.63-54, it is mentioned that this is the first application of MPS to subglacial topography. A recent paper doing just this appeared in The Cryosphere: https://tc.copernicus.org/preprints/tc-2021-161/

  Response: We have added this new paper to our reference list and revised the manuscript accordingly.

- The proposed method involves a number of modeling steps and choices. This is fine when justified, but here I did not see a clear justification for many of the choices made. I detail some instances of this below:
    - I do not see a clear motivation for using a modified Hausfdorff distance in eq. 1. Why using a distance between patterns rather than some statistical similarity between patterns, e.g. in terms of integral scales?

      Response: This is a very good question. The reason is that we need a simple method to measure distance between TIs' representative patterns. Using statistical similarity approaches would require two steps 1) specifying statistics 2) specifying distance.The Hausdorff distance captures both in one because it measures difference between shapes, which is relevant to this application. But yes, the approach suggested by the reviewer will also work.

- There exists a literature on methods to select one or more TIs based on conditioning data (e.g. Pérez, C., G. Mariethoz, and J. M. Ortiz (2014), Verifying the high-order consistency of training images with data for multiple-point geostatistics, Computers and Geosciences, 70, 190-205, 10.1016/j.cageo.2014.06.001, or Abdollahifard, M. J., M. Baharvand, and G. Mariéthoz (2019), Efficient training image selection for multiple-point geostatistics via analysis of contours, Comp. & Geosci., 128, 41-50, https://doi.org/10.1016/j.cageo.2019.04.004). This is the same problem that is addressed here with the probability aggregation approach. Such methods are not considered or mentioned. It is fine that the authors propose a new methodology, but the reason for not using existing approaches should be given.

  Response: We thank the reviewer for providing the above TI selection methods. We agree that they share similarity with the TI selection problem in our paper. However, the unique contribution of our paper's approach is that we quantify the posterior probability, then sample from it. We have added these papers to our references to Section 3.7 and explained our reasons in the revision.

- The general approach proposed is rather sophisticated (probability aggregation - PSO optimization – kernel density estimation), which complexifies the implementation. From a user point of view, these steps should be justified. Basically, it is pretty clear how things are done but the why is not always explicit.

  Responses: The methodology framework may be complex, but the execution is simple from a user's perspective. Also, from users' perspective, we provided the implemented code as supplementary material. The codes only require the user to provide the line data, define local areas, and a few DS parameters to run it. They don't need to reimplement the theoretical details.

- 303-304: this is a strong assumption because these distances are considered in a high-dimensional space. Can it be justified?

Response: Yes, it can be justified. The MDS plot of Figure 5 justifies this assumption. It shows the nearby TIs in MDS space are very similar to each other and tend to have the same geological features of fjords, mountain ranges, etc. They will thus have a similar probability of being assigned to the same radar data subarea. We added a line to explain this.

- The procedure for the choice of the TIs is rather complex. This is justified by the dependence between neighboring TIs which is modeled by probability aggregation. However, since there are local data everywhere in the domain, simply selecting a set of TIs statistically similar to the data might also perform well. Has this been tested? For instance, looking at figure 10, I am wondering whether a similar ranking could be obtained with a simpler approach, for instance considering similarity in terms of histogram and variogram similarity.

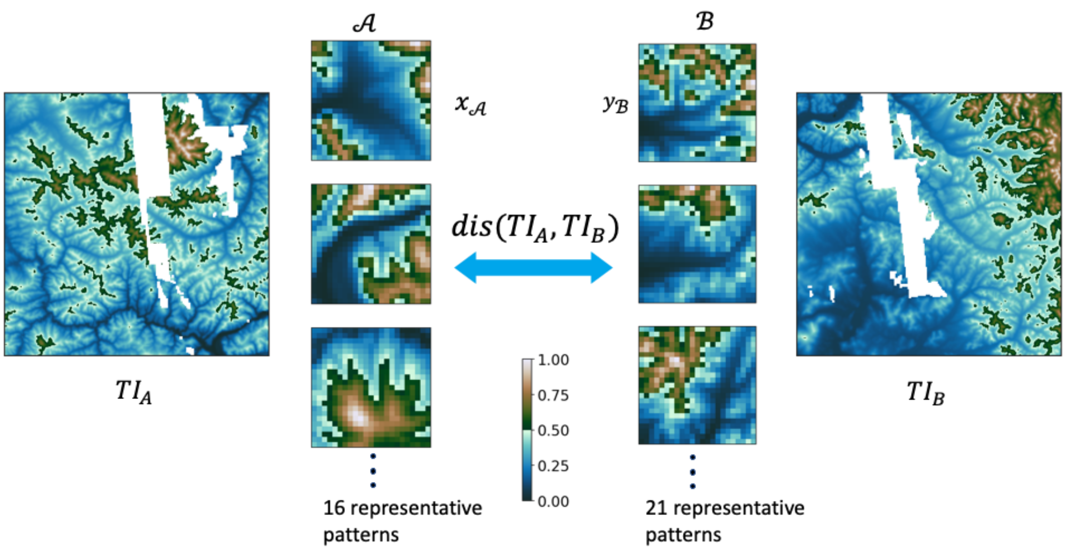
  Response: That has not been tested. Our approach is general, it will work in cases where areas with dense data border areas with sparse data, where the TI selection of the sparse data area may lead to conflict with the dense data area. Secondly, the approach also creates multiple realization of possible pairs of TI in neighboring zones, thereby possibly create a fuller sampling of the possible TI selection in any two neighboring areas

- It is likely that the training images need to be optimally oriented prior to simulation. Or possibly, one could use a rotation-invariant distance. Has this been tested?

  Responses: Yes, we tested to rotate the TIs with different angles. It turned out to make the DS computationally demanding when including more rotated TIs, but it produces similar topography results to non-rotating.

- Section 3.1.3 and the legend of figure 4 are quite unclear to the uninitiated reader and should be improved.

  Responses: we have re-plotted the legends of Figure 4 to make it clearer. Below is the new figure:



- 227: This statement about boundaries seems incorrect. DS is affected by the boundaries of the TI, as the edges of the TI cannot be sampled, especially for large data events.

  Responses: Here we didn't mean the TI boundaries, but cross the specified radar line regions. Thanks for pointing this statement out. We have rephrased this sentence to avoid confusion.

Minor comments/typos:

l.22: sentence grammatically incorrect

Corrected

l.23: provides

Corrected

l.24: flight lines data

Corrected

l.86-87: here the work by G. Pirot could be mentioned as it does exactly what is mentioned in this sentence (https://doi.org/10.1016/j.geomorph.2014.01.022, and also doi:10.1002/2015WR017078).

The two references are added.

l.96-97: Awkward sentence, rephrase

Sentence has been rephrased.

l.115: remove extra parentheses

Corrected

l.149: section 0 is a mistake

Corrected

l.182: verbatim copy should be explained

We have rephrased this line.

l.204: remove word expressed

Corrected

l.286.287: this reference seems out of place. This is the TI selection, not DS simulation

Corrected

l.287: erroneous reference: there is no section 3.13

Corrected

l.292: parameterization … convergence

Corrected

l.293: make PSO

Corrected

l.320: “is to reflect” -> should be rephrased

Rephrased to “is to quantify”

l.370: section 0 doesn’t exist

Corrected

l.397: rewrite sentence (diving ->dividing)

Corrected

Figure 18: problem of color scale where white areas appear in the simulation domain

We have changed the white area to black.

l.424: density of water is 1000 kg/m3

Corrected

Check the citations: several are duplicated.

Duplicated citations are removed