Responses to the Comments from the Reviewer 1

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. 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.

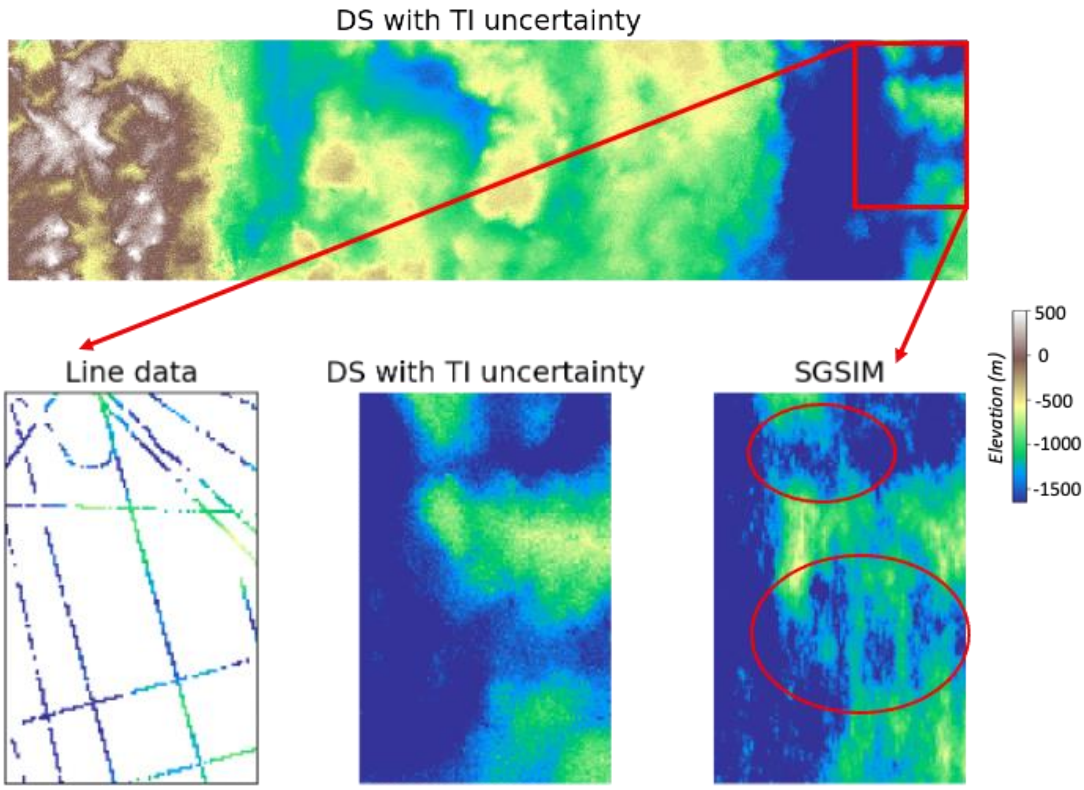
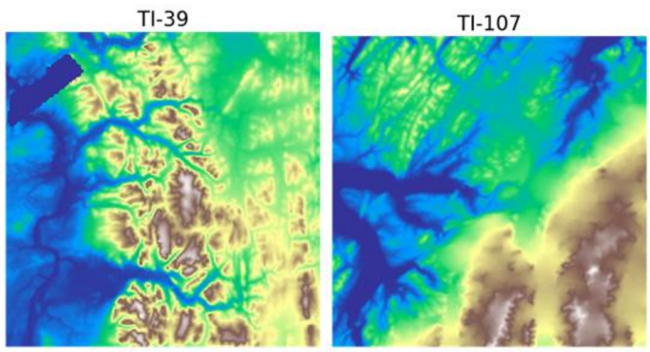


Figure 14. Comparison of SGSIM and proposed DS with uncertain TI selection in the local sparse lines area. Red circles highlight the areas where SGSIM failed to simulate meaningful channels.



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."

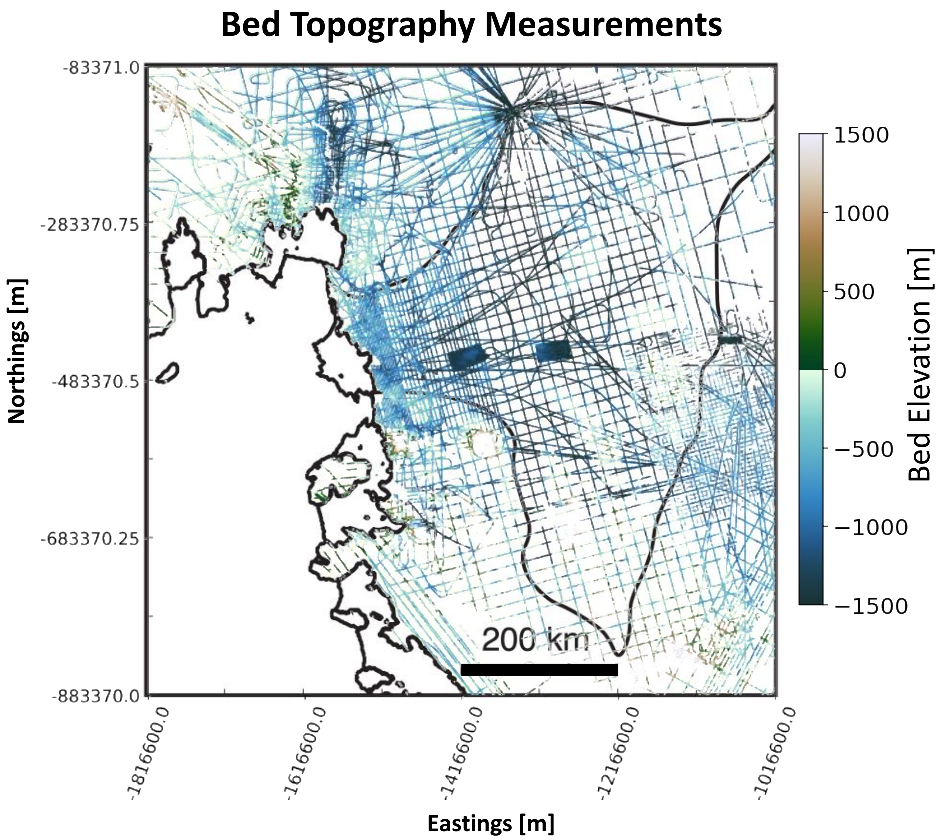
Page 5 Line 126:

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 $P_{n=3}^{1} = (86,80,37)$ with three known points is checked. Then, the computer launches a random searching procedure. As shown in Figure 3(c), a pattern $P_{n=3}^{2} = (21,7,30)$ is created. The distance metric is as follows:

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

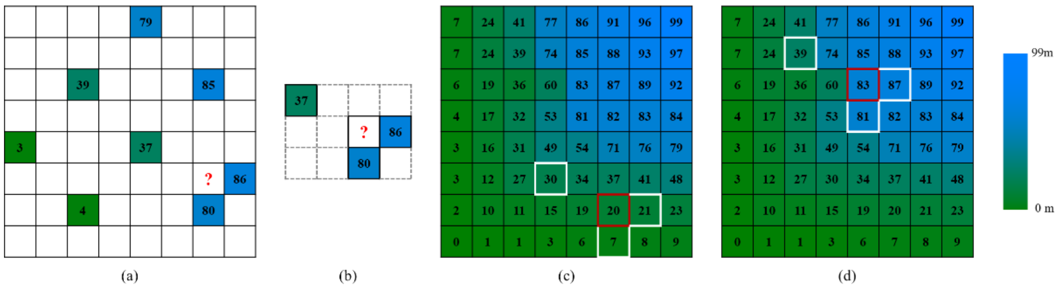


Figure 3. A conceptual example of the DS point simulation. (a) Radar lines on the simulation grid; (b) Three known points (value: 37, 80, 86) constitute a conditioning data pattern; (c) A mismatch pattern in TI; (d) A similar pattern in TI.

Because the distance is much larger than the threshold, the program has to test another point in TI. As Figure 3(d), a pattern $P_{n=3}^{3} = (87,81,39)$ is found. The distance between $P_{n=3}^{1}$ and $P_{n=3}^{3}$ is 0.01. Consequently, $P_{n=3}^{3}$ 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.

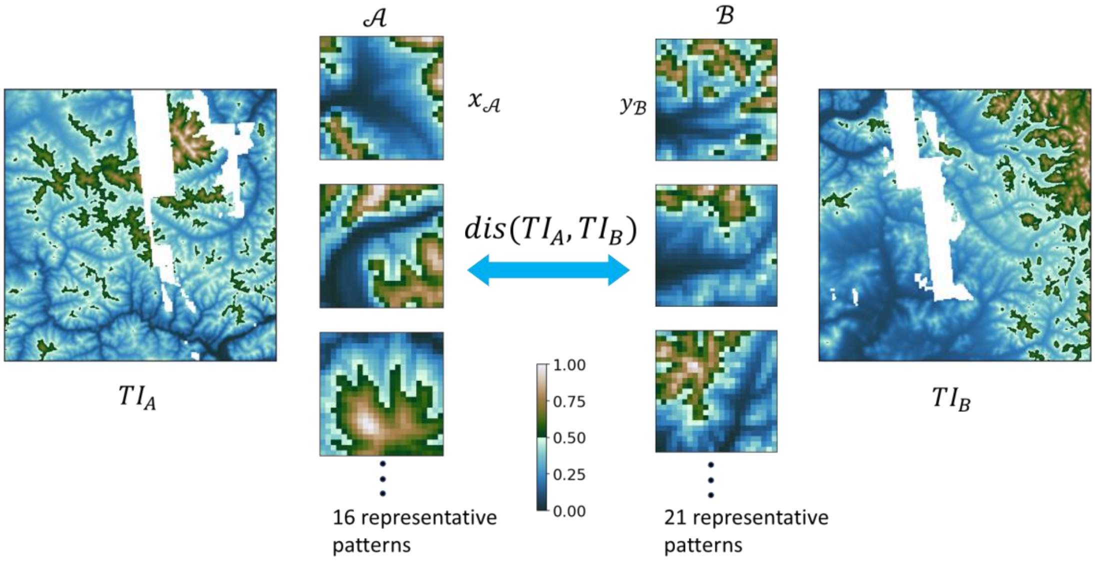


Figure 4. Calculating the distance between any two training images (TI_A and TI_B) using modified Hausdorff distance. There are three key steps: (1) Extracts training patterns with a fixed template. (2) The representatives are selected by a hierarchical clustering method. In this example, the computer 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:

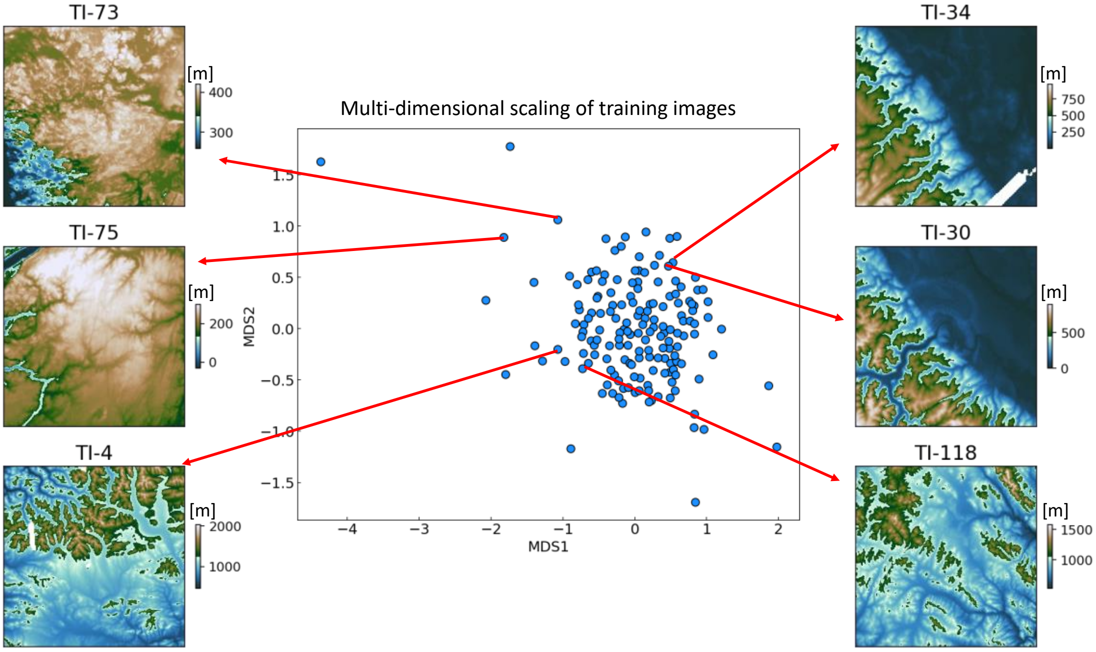


Figure 5. Visualization of the metric space using multi-dimensional scaling (MDS) into a two-dimensional cartesian space. Each dot on the plot represents a TI. It shows TIs with similar morphology are close in this metric space.

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."

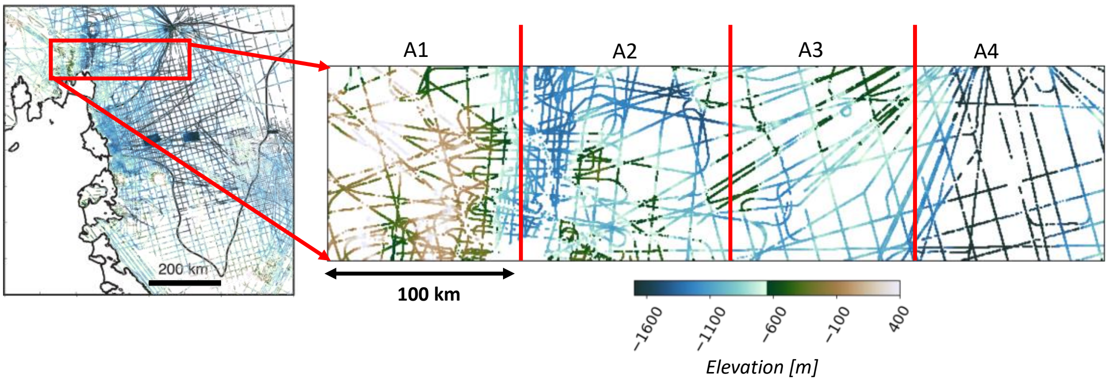


Figure 6. A subset of Pine Island Glacier is used to illustrate the methodology. Apparently, A2 and A3 share the similar morphology. Thus, our program assigns comparable TI to these two areas. In contrast, there is a considerable difference between A1 and A4.

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

Comment: define

Response:

We have defined this formular in detail:

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

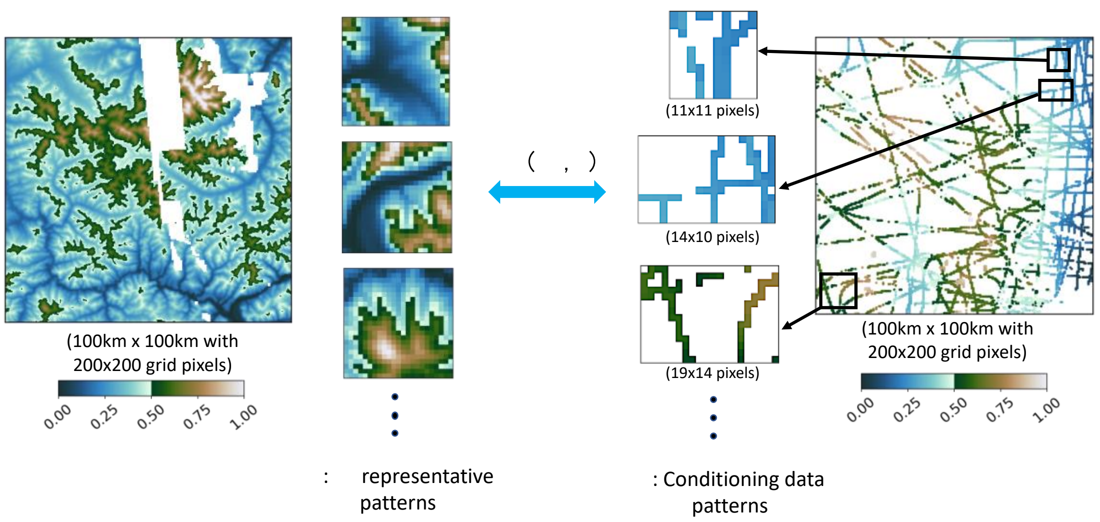


Figure 1. Illustration of measuring the distance between training image ($\boldsymbol{TI}$) and radar lines data ($\boldsymbol{d}$) in subarea A_1. We first extract a group of radar data patterns $\mathcal{D}$ from the simulation grid with flexible sized templates. Then the Hausdorff distances between 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 $dis(I_{TI}(\widehat{TI}), d_{A_i})$.

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."

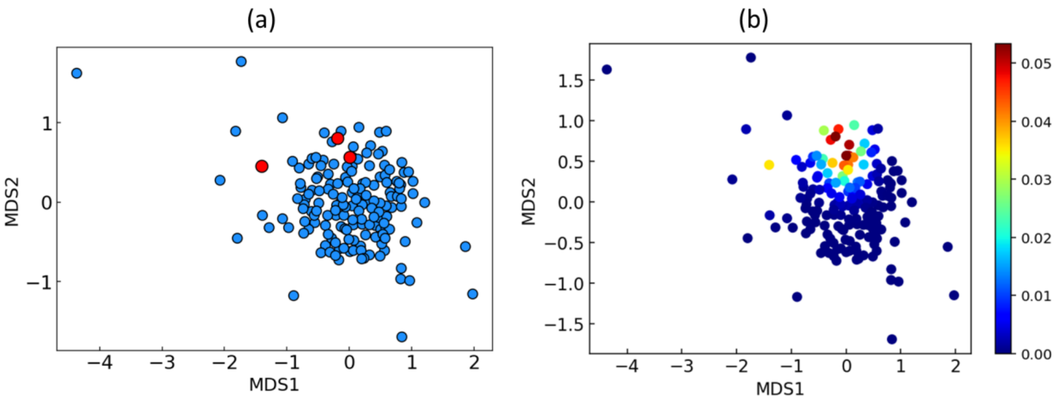


Figure 9. Probability computation based on the selected TIs. (a) Estimated $\widehat{TI}$ for subarea A1 in MDS space. The red dots are $\widehat{TI}$ while blue points represent other TIs. (b) Prior probability $P(TI(A_1)|d_{A_1})$ of each TI. Our kernel density estimation gives a high possibility to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

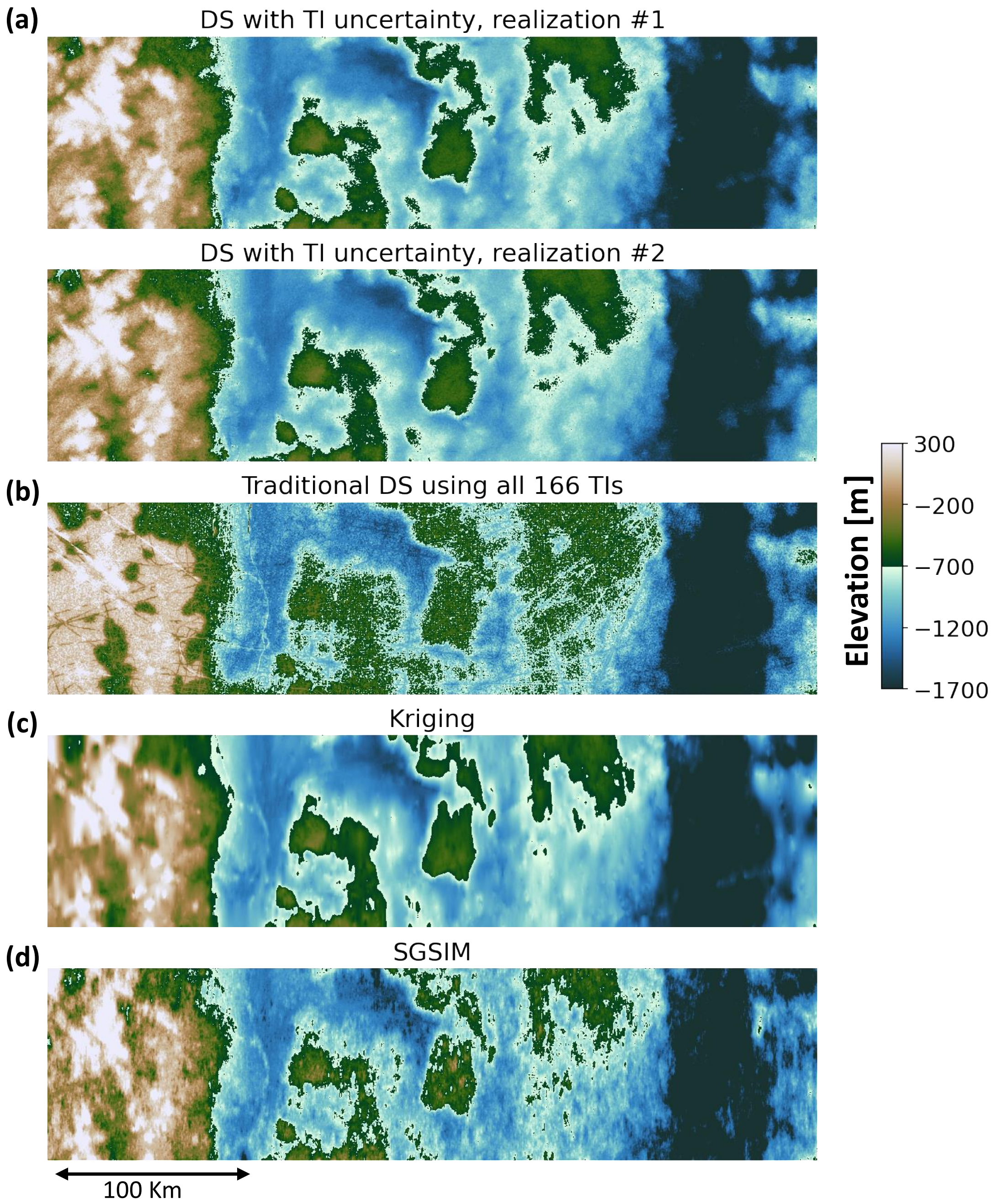


Figure 12. (a) Two realizations of DS simulated topographical models by filling the radar line gaps. Model realization number corresponds to the TI realization number in Error! Reference source not found.**. (b) line gaps filling by traditional DS using all the 166 TIs (without TI sampling). (c) and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

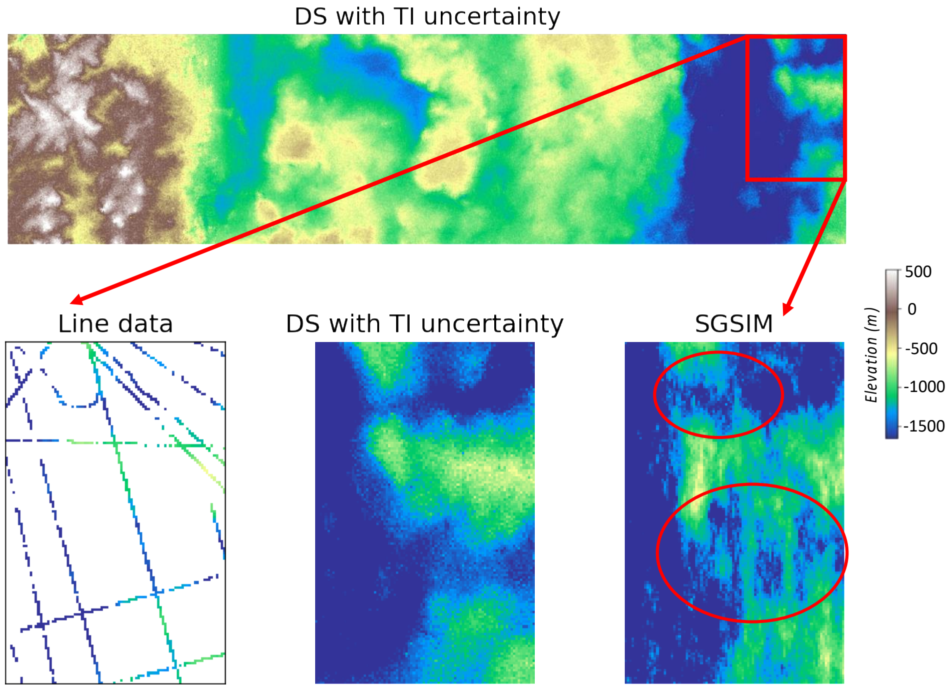
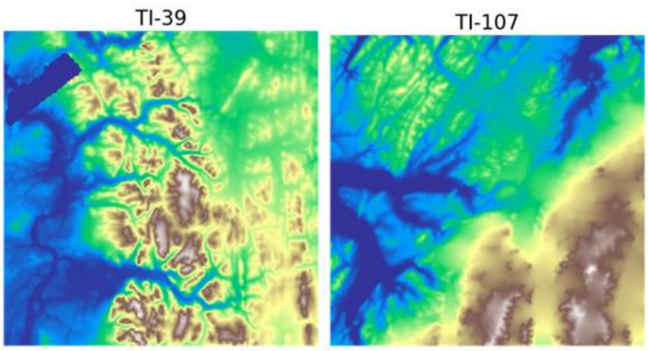


Figure 14. Comparison of SGSIM and the proposed DS with uncertain TI sampling in a local sparse lines area. Red circles highlight the areas where SGSIM failed to simulate meaningful channels.



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.