Earth Surf. Dynam. Discuss.,
https://doi.org/10.5194/esurf-2020-40-RC3, 2020

Earth Surface Dynamics Discussions
Open Access

Interactive comment on "GERALDINE (Google earth Engine supRaglAciaL Debris INput dEtector) – A new Tool for Identifying and Monitoring Supraglacial Landslide Inputs" *by* William D. Smith et al.

Sam Herreid (Referee)

samherreid@gmail.com

Received and published: 24 July 2020

The article "GERALDINE (Google earth Engine supRaglAciaL Debris Input dEtector) – A new tool for Identifying and Monitoring Supraglacial Landslide Inputs" By Smith at al. describes a tool that subtracts composite debris maps from two stacks of Landsat images, one from a period of interest and the other from the preceding year, to isolate new debris additions. A user can then interpret this output to locate supraglacial rock avalanches or landslides. I found the paper mostly easy to read and I think the research objective is timely and useful. I also appreciated the user guide provided in

Printer-friendly version

Discussion paper

the supporting information. However, I think the authors stopped their tool development prematurely leaving some fundamental elements unaddressed. My main points of concern are briefly summarized here with more detailed comments inline below along with minor comments.

Looking at the two map figures of the article, it is clear that, even within the GEE stack methodology, which is in principal sound, debris cover is not confidently mapped. There is unphysical debris in the accumulation zone and many instances of “new debris additions” that are not new debris additions. These areas accumulate into tool output false positives that are neglected by the authors who rather only report true positive success, leading to statements like L283-284: “GERALDINE outputs [had a] 100% successful identification”. By neglecting to calculate a metric like precision or the false positive rate, the study is lacking a meaningful assessment of performance. I think it is reasonable to state, as the authors do, that some of these debris map errors stem from errors in the RGI, but these then need to be either mitigated or quantified in the error assessment of your tool.

From my view, the main incentive for a tool that considers every image acquired in a stack, is to detect rock avalanches that are deposited onto a glacier’s accumulation zone and automatically assign a best constrained deposition date. The automated detection of rock avalanches deposited onto bare glacier ice in ablation zones is also useful, but there is less chance of missing one since there will be a surface expression in every snow/cloud free image after deposition until it is too heavily reworked or evacuated from the glacier. Further, in ablation zones there is the case, that will likely only grow in frequency, where a rock avalanche is deposited onto existing debris cover, or earlier deposited rock avalanche debris, which is an entirely undetectable event using this method. By summing debris cover over one year or longer, the method presented here will likely catch a deposit onto the accumulation zone, but by not finding the difference between each sequential image the approach loses any ability to assign a deposition date. I understand the incentive to aggregate debris, but from the comment above, I think the quality of the resulting debris maps are still low relative to other automated debris maps in the literature.

I think a rock avalanche deposit onto bare glacier ice is a strong signal that can be detected automatically. For example, the area of a rock avalanche feature will almost always be much larger than any other location of debris additions from other sources (if mapped accurately and dt is short, e.g. 1 year). The authors leave this step to the user which I think significantly reduces the applications of this tool. I can accept that this version does not need to perfectly resolve all of elements to mapping rock avalanches onto clean glacier ice, but I think providing an automated selection of rock avalanches from new debris additions is only a minor addition that will increase both the tool application as well as ability to quantify true positives, false positives and false negatives. I also think that looking at the differences between every image after a rock avalanche is detected to constrain the date of deposition is a reasonable and achievable result at this stage of tool development.

If this method is to be a starting point for a globally applicable tool (L22), I am concerned that the authors cite limitations of GEE that cause the region of interest to be limited to <5000 km2. Do the authors anticipate that this method could be written in a more computationally efficient way such that this limit will be dramatically increased? Highly useful functionality of a tool like this one will be when all of Earth's glaciers can be assessed in near real time, but if there are intrinsic limitations within GEE is this a feasible future for this tool?

Finally, there is a factor present in the quantity "new debris additions" that is not quantified or discussed. Unstable glacier flow will produce debris structures that deviate from flow lines parallel to a glacier's valley wall (e.g. the surge loops on Susitna Glacier in your Figure 4) and a difference map of debris cover over some dt will show a false gain and false loss of debris cover that is really just debris structure translation. Where glacier flow instabilities are present, a simple difference of debris cover maps cannot be strictly new debris additions. Herreid and Truffer, 2016 provides a discussion on this topic.

Herreid, Sam, and Martin Truffer. "Automated detection of unstable glacier flow and a spectrum of speedup behavior in the Alaska Range." Journal of Geophysical Research: Earth Surface 121.1 (2016): 64-81.

L1: Perhaps stylistic but I think “A new tool for identifying and monitoring supraglacial landslide inputs” is a better title, without the less straightforward and somewhat redundant acronym.

L9: Why not use “rock avalanche” throughout? I believe rock avalanche is more precise and consistent with the literature for what you are looking at. If the authors prefer the more general term landslide, then early in the introduction make clear what is and is not a landslide vs rock avalanche for this study and keep the language consistent. It’s strange to read landslide in the title and have rock avalanche be the first sentence of the abstract.

L9-12: There is a missing step here, rock avalanches can happen far from glacier ice. Detection of RAs for the study of RAs alone, or to answer frequency questions with respect to climate or ice factors, should consider all RAs independent of their runout happening to be on a glacier. This is either a very big sampling bias or you should pose a glacier specific problem.

L14: It reads like you are focusing on filling this small to medium gap but on L43 you say you focus on the inputs of high magnitude, > 10ˆ6 mˆ3, RAs. Please clarify/fix and keep consistent throughout. L215 considers a 0.062 km2 event.

L22: From the abstract alone you don’t mention measuring area or volume or event timing, so I don’t quite see the jump to a global product. Further, on L118 you advise ROIs <5000 km2. Do you anticipate a less computationally costly version of your method or are there HPC options in GEE? Finally, it is a little strange to have a first step towards a revision, a revision implies several steps have already been taken.

L26: With the known errors in the RGI, it's better to avoid presenting the number of glaciers to the accuracy of a single glacier. Consider ">200,000".

L27: Consider a revised global estimate of debris cover from Herreid and Pellicciotti, accepted by Nature Geoscience, which should be available by August 2020 at this DOI: 10.1038/s41561-020-0615-0

L34: Either add "e.g." to the citations or also add a citation to Kirkbride and Deline, 2013 whose Table 1 gives a more complete list of citations for expanding debris cover.

Kirkbride, Martin P., and Philip Deline. "The formation of supraglacial debris covers by primary dispersal from transverse englacial debris bands." Earth Surface Processes and Landforms 38.15 (2013): 1779-1792.

L35: What is the difference between sub- and en- glacial sediments in this context? I don't think sub-glacial sediments can melt out.

L35: Anderson, 2000 addresses general dispersion of medial moraines which you don't explicitly mention here. Does "debris store" mean extraglacial debris? This is not clear.

Anderson, Robert S. "A model of ablation-dominated medial moraines and the generation of debris-mantled glacier snouts." Journal of Glaciology 46.154 (2000): 459-469.

L36: It might be worth distinguishing here high volume low frequency mass movements from low volume high frequency.

L43: How are you able to focus on landslide of a particular volume? Throughout you do not calculate or consider volumes. And do you mean high volume? Magnitude of what?

L44: "where there is disparity between current high rates of activity above ice" this is unclear.

L46: lag ice-free conditions in terms of what?

L47 What does “relatively low in the landscape” mean?

L58: I’m not sure if there are remote sensing methods yet to see englacial debris. Maybe you mean geophysical methods, e.g. GPR.

L59: “[add: potentially] considerable modification”

L60: “Deposited”? “Emplaced” is odd.

L72: Landslides vs RA confusion here.

L87: Open access or open source?

L90: Define what you mean by “wide” in parentheses

L109: RGI errors are further quantified in Herreid and Pellicciotti, accepted by Nature Geoscience, available around August 2020 at DOI: 10.1038/s41561-020-0615-0

L116: add: “[and all images in the] year preceding. . .”

L118: What do you mean by “specify annual date ranges”? Are you saying the tool can only work for one time window between two specified years? This seems like a pretty critical limitation to the functionality to the tool. Are you sure GEE is the correct platform if its memory capacity is such a bottleneck? Maybe JuypterLab is a better cloud-based platform? Or your code could select a single optimal image of a one year stack and then make your calculations on single images? Also if you clip the RGI first, then all of your calculations will be less computationally costly.

L122: This section is not very clear, but if I understand correctly, the tool will collect two stacks, one from the year before a defined date range and one for the full defined date range, and then perform a single subtraction to find a single map of new debris. There is an issue of accumulating “new debris additions” if the stack of images aggregate debris from, say, 10 years, there will be much more new debris additions that are not sourced from RAs. You also lose the ability to automatically detect a deposition date which is, in my view, the main incentive to use GEE and consider stacks of images rather than single optimal images. I think maybe you should change the wording of a “user-specified date range”, and rather say “a user specified year where the tool will give you a map you can look for RAs deposited since the preceding year.” But 1. I don’t understand why finding the RAs can’t also be automated, this should be a very clear signal if deposited on clean ice (you will entirely miss RAs that are deposited onto existing debris cover); and 2. As a user I can think of two uses for a tool like this: (a) getting the location of all RAs that have been deposited onto a glacier and are still present at the surface and a deposition date if deposited since Landsat 4; and (b) near-real-time detection. I think your tool could be successful for the latter, although to be practical it should be able to analyze all of Earth’s glaciers at once or at least all glaciers in, say, Alaska (Bearing Glacier in SE Alaska alone is larger than the recommended <5000 km2 ROI), but I think there is still a lot of improvement needed for the former. The difference map needs to be computed annually to keep other debris addition signals small and also facilitate a deposition date.

131: I can appreciate that the method used to assess cloud mask performance considers clouds in an entire stack, thus incorporating a variety of cloud types in a simple run of your code. However, I would like to see more direct evidence that clouds themselves are accurately mapped. From my experience cloud mapping algorithms are unreliable in glacierized areas. Could you show a side by side image of a raw satellite image and an overlay of the output of the cloud mask with scores >20%, perhaps one where it worked well and a second where it was at its worst. I’m concerned that you’re only mapping 60% of RA area. How were the studies that make up your validation dataset able to map 100% of the RA area and you cannot? Surely with the stack methodology the aggregate over many images should, together, capture 100% of RA area unless it’s a particularly snowy or cloudy year. Does this suggest you have a 40% error rate in detecting RAs?

L132: What about cast shadows from topography? Herreid and Pellicciotti, accepted by Nature Geoscience (available August 2020 at DOI: 10.1038/s41561-020-0615-0)

Printer-friendly version

Discussion paper

found it necessary to remove area in shadow in order to accurately map debris cover. The band ratio method is able to negotiate some shading, but when a surface becomes too dark there is still the possibility for false positive debris classification (e.g. Herreid and Pellicciotti, 2020 removed 760 km2 of shaded glacier area in Alaska and Western Canada).

L134: I don't really see a justification for the step of mapping supraglacial lakes or ponds. These features generally develop in heavily debris-covered portions of glaciers where your tool will fail to detect a RA by not having the prior bare ice context. Further, if these features are 22 pixels on average, as you cite in the SI, then the above discussed 40% omission error dwarfs the stream/pond signal. If you elect to keep this component please provide an example in the SI that shows how mapping streams and ponds leads to a higher rate of RA detection.

L150: One of your inequality signs should include "or equal to"

L158: There is a missing discussion on double counting translated debris features that deviate from a flowline parallel the glacier valley walls. Also summed non-RA debris additions if the user defined time period is not sufficiently short. Herreid and Truffer, 2016 established a very similar methodology to the one presented here in order to detect glacier flow instabilities. In this study RA are identified but considered an error in the context of the flow instability research question. For your work, RAs are signal and the features identified by Herreid and Truffer, 2016 are errors. These should be discussed.

Herreid, Sam, and Martin Truffer. "Automated detection of unstable glacier flow and a spectrum of speedup behavior in the Alaska Range." Journal of Geophysical Research: Earth Surface 121.1 (2016): 64-81.

L162: What do you mean by "Debris biased"?

L168: Do you mean an omission/commission validation? If not, please provide an additional sentence on why a bipartite approach was used.

L172: RA already defined.

L175: 48 suitable events were found out of how many that you considered? It is helpful for the reader to know if these are rare occurrences or the majority. I assume these inventories only consider supraglacial RAs?

L175: please add a map figure showing all of the regions you applied your tool

L189: I think if your code mapped RAs from the best available image for each event, rather than a composite, you could be very close to 100%.

L189: A relevant factor that you do not mention is a RA that crosses existing debris cover. This is likely the predominant factor of why you will not be able to map RAs to 100%.

L196: The accuracy of the satellite image remains the same, the overall significance of a single pixel of a small glacier increases.

L197: Looking at the noise in bare ice regions of Figure 4 I struggle to see what you mean by “true negative detection rate is also extremely high”

L198: I don’t agree with this justification for user verification. If you subtract two optimal satellite images before and after a RA deposition onto a non-debris-covered portion of a glacier, the signal is exceptionally prominent, and I see no reason why an algorithm cannot easily identify this automatically. I think somewhere in your GEE stack processing, the debris mapping and the cloud removal, a very clear signal becomes muddy. I think some small changes to your workflow can provide a much clearer, and likely more computationally efficient output.

L199: The problem with saying “to a user familiar with glacial and landslide processes, the [tool output is] clear” is that a user familiar with glacial and landslide processes will be able to spot large landslides onto bare ice from a raw image. The spatial domain of the tool is low <5000km2 and the tool cannot iterate over many years to pinpoint a deposition date. I think there is a lot of potential in a tool like this but in its current state I have a hard time seeing a scientific application.

L202: Please add a section to methods describing how your derived areal extent. Presumably there was a manual step involved in this.

L215: How much user interpretation was involved with isolating the 71% true-positive RA area? False positive and false negative areas must also be considered to make a statement about detection confidence.

L228: But if topographic shading is classified as debris, it will influence new debris detection.

L247: Your method has a high potential to detect all events [add: that are deposited onto initially bare glacier ice]. E.g. a hypothetical second event at the same scarp on the glacier east of Maclaren Glacier that deposited a slightly smaller volume of rock would be entirely missed by your method.

L250-256: I find this to be significant conditionality and required prior knowledge for an automated tool. Your method doesn't automatically run for multiple years sequentially, so how would someone new to the area know where to start? Reading your Figure 4 alone suggests the BRG RAs were deposited between 2017 and 2018, this is misleading. The mapped Lituya RA in Fig. 5 also appears patchy, should the logic of L254 be followed and this area be dismissed as erroneous?

L257: While translated features are present in your output (also translated features from flow instabilities, see Herreid and Truffer, 2016) and are scientifically useful, these are errors with respect to your intended tool function. If you can automatically differentiate feature translation from feature deposition then this can be a nice side component to your study, otherwise I think you need to treat this as error.

Herreid, Sam, and Martin Truffer. "Automated detection of unstable glacier flow and a

Printer-friendly version

Discussion paper

spectrum of speedup behavior in the Alaska Range." Journal of Geophysical Research: Earth Surface 121.1 (2016): 64-81.

L275: How does reduced ablation over one year around the ELA, where ablation rates are generally low, increase surface velocities?

L281: RAs on bare glacier ice in ablation zones are easy to identify from one recent image and your method also requires manual inspection. Here I think you should highlight your tool's ability to potentially catch events in the accumulation zone that have only a very short residence time.

L284: This is the first mention of 100% successful identification which should first appear in the results section, but I also think it is incorrect. By considering only true positive area, a map that is entirely "new debris additions" will also have a 100% successful identification rate but is clearly meaningless. You need to score your success against false positive and false negative area.

Interactive comment on Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2020-40, 2020.