

1 **Response to Reviews for: “GERALDINE (Google earth Engine**
2 **supRaglAciaL Debris INput dEtector) – A new Tool for Identifying and**
3 **Monitoring Supraglacial Landslide Inputs” (MS No: esurf-2020-40)**

4 Dear Dr. Conway,

5 We thank you and the reviewers for your time, and constructive and helpful comments on our
6 manuscript. We have addressed each one of the comments below, and, have added suggested
7 manuscript additions formatted as “new text” where appropriate.

8 Reviewer comments are *grey highlighted*, and, use their numbering where it was used in the
9 review.

10 William Smith, on behalf of all authors.

11 **Gioachino Roberti (Reviewer 1)**

12 *I enjoyed reading the paper "GERALDINE (Google earth Engine supRaglAciaL Debris Input dEtector)*
13 *- A new Tool for Identifying and Monitoring Supraglacial Landslide Inputs" and I recommend it for*
14 *publication. The paper presents a new tool to exploit Landsat images in Google Earth Engine to map*
15 *debris onto glaciers, therefore providing a semi-automatic tool to identify rock avalanches emplaced*
16 *on to glaciers, and to track supraglacial debris movement. This tool can complement seismic analysis,*
17 *and, if extensively applied, help developing F-M curves of rock avalanches onto glaciers in the past 37*
18 *years.*

19 We thank the reviewer for their time and comments. We have addressed each one of these below.

20 *The following comments can help to further improve the paper. I think a better overview of satellite*
21 *spatial resolution and detectable landslide size is needed in introduction.*

22 We agree and will add a new paragraph on line 87 as follows:

23 “Since the launch of Landsat 1 in July 1972, optical satellites have imaged the earth surface at increasing
24 temporal and spatial frequency. Six successful Landsat missions have followed Landsat 1, making it
25 the longest continuous optical imagery data series, revolutionising global land monitoring (Wulder et
26 al., 2019). Analysis ready Landsat data is available for Landsat 4 (1982-1993), Landsat 5 (1984-2012),
27 Landsat 7 (1999-present) and Landsat 8 (2013-present), providing 38 years of data at a 30 m spatial
28 resolution and a 16-day temporal resolution. These data are categorised into three tiers: (1) Tier 1 data
29 that is radiometrically and geometrically corrected (< 12 m root mean square error); (2) Tier 2 data
30 which is of lower geodetic accuracy (> 12 m root mean square error); and (3) Real Time imagery, which
31 is available immediately after capture but uses preliminary geolocation data and thermal bands require
32 additional processing, before being moved to its final imagery tier (1 or 2) within 26 days for Landsat
33 7, and 16 days for Landsat 8. Traditionally, it has been difficult to exploit extensive optical imagery
34 collections such as Landsat, without vast amounts of computing resources. However, in the last decade,
35 cloud computing has become increasingly accessible. This allows a user to manipulate and process data
36 on remote servers, removing the need for a high-performance personal computer. Google Earth Engine
37 (GEE) is a cloud platform created specifically to aid the analysis of planetary-scale geospatial datasets
38 such as Landsat and is freely available for research and education purposes (Gorelick et al., 2017).

39 Here, we utilise Google Earth Engine (GEE), and the Landsat data archive of 37 years of optical
40 imagery, to present the Google earth Engine supRaglAciaL Debris INput dEtector (GERALDINE). An

41 open-source tool to automatically delimit new supraglacial landslide deposits over wide areas and
42 timescales...”

43 We shall remove any information from the methods section that is explained in this paragraph.

44 Wulder, M.A. et al.: Current status of Landsat program, science and applications, *Remote Sensing of*
45 *Environment*, 225, 127-147, doi: 10.1016/j.rse.2019.02.015, 2019.

46 *In the discussion section you could add some paragraphs:*

47 *1) A paragraph about development F-M curves, as it is a topic mentioned in introduction and conclusion*
48 *but not directly addressed in the discussion*

49 We do not present any results in this primarily methodological paper that revise magnitude frequency
50 relationships. We consider this outside of the present scope of the paper, which is to present the tool,
51 its capability, and to validate it against known events so it can be applied to specific areas in the future.
52 However, it is an important point for the introduction and discussion as this is how the tool will be
53 applied in future (by us, and, by others) once the tool is accepted and published. In this study we are
54 signposting the user to use GERALDINE with confidence, even in areas with existing landslide
55 inventories where GERALDINE can/should be tested to confirm or modify F-M curves. We will revise
56 the introduction to include more detailed information of recently published inventories of glacial RAs,
57 which provide a basis for F-M testing. On line 71 we will reword and add:

58 “Manual imagery analysis to identify supraglacial landslide deposits and RAs has principally been
59 applied in Alaska. This technique enabled detection of 123 supraglacial landslide deposits in the
60 Chugach Mountains (Uhlmann et al. 2013), 24 RAs in Glacier Bay National Park (Coe et al. 2018), and
61 more recently, 220 RAs in the St Elias Mountains (Bessette-Kirton and Coe, 2020). These studies
62 acknowledge that their inventories are incomplete/underestimates due to analysis of summer only
63 imagery and an inability to detect events that are rapidly advected into the ice. These are critical
64 drawbacks preventing accurate magnitude frequency relationships from being derived but analysis of
65 more imagery over larger areas is unfeasible due to time and computational requirements. Studies of
66 this kind are also typically in response to a trigger event e.g. earthquake or a cluster of large RA events
67 (e.g. Coe et al. (2018) in Glacier Bay National Park), spatially biasing inventories into areas with known
68 activity. They therefore provide a snapshot in time, with no continuous record. Methods are needed
69 which are accessible, quick and easy to apply and require no specialist knowledge, to re-evaluate
70 magnitude frequencies in glacial environments. Currently, the only method capable of identifying a
71 continuous record of such events, is seismic monitoring (Ekström and Stark, 2013). Seismic detection
72 utilises the global seismic network to detect long-period surface waves, characteristic of seismogenic
73 landslides. Seismic methods have identified some of the largest supraglacially deposited RAs...”

74 Bessette-Kirton, E.K. and Coe, J.A. A 36-Year Record of Rock Avalanches in the Saint Elias Mountains
75 of Alaska, With Implications for Future Hazards, *Frontiers in Earth Science*, 8:293, doi:
76 10.3389/feart.2020.00293, 2020.

77 *2) A paragraph about "eliminating the time-intensive process of manually downloading, processing*
78 *and inspecting numerous satellite images" that is then mentioned in the conclusion. With considerations*
79 *about transferability of the method to other satellites and data storage and processing platforms*

80 We fully agree this needs quantifying given it is a major benefit and will add this brief section to line
81 233:

82 “GERALDINE used a total of 228 Landsat images for analysis; 107 to determine the 2017 debris extent
83 and 121 to determine the 2018 debris extent. Landsat tiles vary from 200 MB to 1000 MB when
84 compressed, so, if we assume an average tile is 500 MB, a user would require 114 GB of local storage,
85 a large bandwidth internet connection to download (which comes with an associated carbon cost), and,

86 a PC capable of processing these data. GEE required none of these requirements and completed analysis
87 in under two minutes, extracting information from every available cloud-free pixel, to maximise use of
88 the imagery. The new debris output map produced was 6.5 MB, and contained all relevant ‘new’ debris
89 information from 2018.”

90 *3) and (eventually) how a similar approach may be used in other context (landslides in forested areas*
91 *etc.)*

92 We understand why the reviewer asks this question, but the methodology used as presented here is not
93 suitable for use in other environments and has been deliberately tuned/developed to detect in snow and
94 ice landscapes where landslide deposits have little residence time. The ability to threshold a band ratio
95 technique into two distinct categories (ice/snow and debris), is what GERALDINE exploits. In contrast,
96 there are multiple substrates which must be categorised in other environments, and therefore a different
97 method is required. Scheip and Wegman exploit percentage change and NDVI in their tool (developed
98 at the same time as ours) which is tuned to vegetated landscapes. We will make reference to:

99 Scheip, C.M. and Wegmann, K.W. HazMapper: A global open-source natural hazard mapping
100 application in Google Earth Engine, Nat. Hazards Earth Syst. Sci. Discuss., doi: 10.5194/nhess-2020-
101 108, in review, 2020.

102 **In line comments:**

103 *Line 14: Quantify? What’s the size of the smallest detectable landslide?*

104 A specific size is not possible because it depends on the strength of the seismic signal. We shall clarify
105 and reword to:

106 “Although large landslides can be detected and located using their seismic signature, smaller landslides
107 ($M \leq 5.0$) frequently go undetected because their seismic signature is less than the noise floor,
108 particularly supraglacially deposited landslides which feature a “quiet” runout over snow.”

109 *Line 18: You can detect only the large landslides? This sounds in contrast with your earlier statement*

110 We shall amend to “from which large debris inputs such as supraglacial landslide deposits ($> 0.05 \text{ km}^2$)
111 can be rapidly identified”

112 *Line 21: Ok cool. So large landslides that may not have been identified seismically. I don’t think you*
113 *need to try to "sell it" as alternative or better method than seismic identification, they can be used*
114 *together. A tool like the one you have developed is cool by itself, seismic identification or not.*

115 We think that highlighting that GERALDINE can detect landslides, which are both seismically and
116 non-seismically detectable is important in the abstract as it is not a well-known point. However, we do
117 explain in the manuscript, that the available tools should be used in conjunction – this is now happening
118 with the authors of the original seismic landslide detection paper.

119 *Line 24: Very cool! But you should expand the 37 year F-M topic in the discussions*

120 See above comments (Line 47-77 of our response) on F-M topic. We think this is the end-point use of
121 GERALDINE, establish a past supraglacial inventory globally, then run this as near to live as Landsat
122 allows. However, we wish to have GERALDINE published and accepted, and in use by others, as we
123 continue to try and develop/collate F-M for regions in collaborative ways.

124 *Line 38: Can you method distinguish these 3 types of debris cover?*

125 GERALDINE cannot distinguish between debris types. We think it is important to highlight all ways
126 of debris transport into supraglacial environments, so the user has an idea of all debris sources/pathways
127 and can evaluate fluxes.

128 *Line 40: How do you distinguish re-emerged debris vs supraglacially emplaced debris?*

129 It is not possible to use GERALDINE to identify the different debris transport pathways from remotely
130 sensed data alone and would require further analysis in the field.

131 *Line 43: Ok, you should be clearer about the landslide size in the abstract too*

132 We are going to remove the landslide size from this section of the work as GERALDINE can detect
133 events smaller than this. We will reword to:

134 “Here we focus on supraglacial landslide deposits ($>0.05 \text{ km}^2$), commonly associated with RAs, defined
135 as landslides: (a) of high magnitude ($> 10^6 \text{ m}^3$); (b) perceived low frequency; (c) long runout; and (d)
136 where there is disparity between high present-day rates of slope processes above ice (Allen et al., 2011;
137 Coe et al., 2018) and expected rates based on theories of lagged paraglacial slope responses (Ballantyne,
138 2002; Ballantyne et al., 2014a).”

139 *Line 80: How small is "smaller landslides"*

140 In the context of seismic detection, this is difficult to define because it depends on the seismic signal,
141 which can be determined by a plethora of things such as: source area, volume, drop height and horizontal
142 runout distance. We think it is better to state the magnitude threshold from which they are difficult to
143 determine but we will make this clearer by rewording Line 80 to:

144 “This also results in an inability to detect landslides that are relatively low in volume, due to their weak
145 seismic fingerprint ($M \leq 5.0$)...”

146 *Line 82: It sounds like you are detecting large landslides that have no seismic signature rather than*
147 *"small landslides"... maybe you can reword a bit to put emphasis on the combination of size, frictional*
148 *melting etc.*

149 This sentence is simply to describe the limitations of the seismic method for detecting landslides onto
150 glaciers. We shall reword Line 82 to emphasise that these are two substantial inter-related drawbacks
151 of the method:

152 “This also results in an inability to detect landslides that are relatively low in volume, due to their weak
153 seismic fingerprint ($M \leq 5.0$) and causes underestimation of landslide properties (e.g. event size and
154 duration) because their runouts are seismically “quiet”, likely due to frictional melting of glacier ice
155 (Ekström and Stark, 2013).”

156 *Line 89: It may be worth expanding this paragraph/add new paragraph and give an overview of GEE*
157 *and Landsat satellites. I see you discuss landsat satellites in method and validation sections but an*
158 *overview of the satellites (different tiers, spatial resolution, accuracy, years of operation and revisiting*
159 *time etc...) in the intro will help the reader.*

160 As discussed on line 20-45 of our responses, we will add a new paragraph discussing Landsat and GEE.

161 *Line 97: I think the resolution should be mention earlier in the paper too*

162 Agreed. See above additional paragraph (line 20-45 of our response) in response to line 89.

163 *Line 197: Would this still be useful to assess some of the other supraglacial debris types presented in*
164 *the introduction? Expand on this (see my comment of figure 5).*

165 As mentioned above it is not possible to distinguish between debris types, but on line 277 we will
166 discuss different types of debris and add:

167 “We note other areas are flagged as ‘new debris’ in 2013 and 2014. These are typically where glacier
168 downwasting has occurred exposing more of the valley walls, or where there has been temporal

169 evolution of the debris cover e.g. glacier flowline instabilities. These flow instabilities can cause double-
170 counting of debris when larger time windows are specified (see Herreid and Truffer, 2015). Both
171 processes subsequently cause false classification as ‘new debris’. However, neither glacier
172 downwasting nor evolution of the debris cover display supraglacial landslide characteristics, so it is
173 highly unlikely that a user would mistake them for one.”

174 Herreid, S. and Truffer, M. Automated detection of unstable glacier flow and a spectrum of speedup
175 behaviour in the Alaska Range, *Journal of Geophysical Research: Earth Surface*, 121(1), 64-81, doi:
176 10.1002/2015JF003502, 2016.

177 *Line 217: maybe you could also briefly discuss how this method could be applied (maybe not in GEE*
178 *but in some other environment) to other satellites*

179 We shall add this to the end of the paragraph (Line 219):

180 “We also envisage development with other higher resolution and higher repeat satellites e.g. the Sentinel
181 2 and Planet Lab constellations. However, we found that current cloud mask algorithms for these data
182 are not sufficient for accurate global glacial debris delineation.”

183 *Line 276: In the intro you mentioned the possibility of the development of frequency/magnitude curves*
184 *for landslides onto glaciers, but there is not discussion of that point here. Maybe you can add a short*
185 *paragraph (with example?) to explore that potential.*

186 We hope the above comments (Line 47-77) on M-F topic resolve this.

187 *Line 277 Something else that feels like may be missing in the discussion is the overview of the value of*
188 *"eliminating the time-intensive process of manually downloading, processing and inspecting numerous*
189 *satellite images" that is then mentioned in the conclusion*

190 See above (Line 77-89 of our response) for section which will be added to discuss this to Line 233.

191 *Line 287: I agree, but you should discuss this in the discussion. See my comment about frequency-*
192 *magnitude*

193 See above comments (Line 47-77) on F-M topic.

194 *Figure 4: Can you can mark the collapse scar of these landslides? will help the reader*

195 Agreed. We will add a source scar.

196 *Figure 5: Same, where is the landslide coming from? Can you discuss the origin of the other debris*
197 *addition in the glacier on the top right of the picture? in relation to my comment at Line 197.*

198 We will highlight the source scar in the image. See above comments (line 162-176 of our response)
199 regarding Line 197.

200 **Michelle Koutnik (Reviewer 2)**

201 *In this study the authors develop a powerful new tool to identify supraglacial landslides. They present*
202 *the tool, as well as demonstrate how it can work and the value of it by identifying two previously*
203 *unknown landslide debris events. This is an exciting development, and valuable to capturing these*
204 *events where evidence of them is often lost quickly on the landscape, and yet they are important debris*
205 *sources. I enjoyed reading the paper and I really enjoyed thinking about what may be possible using*
206 *the tool. I have some questions and suggestions for the authors, but all of these points are minor.*

207 *Overall, great work on this*

208 We thank the reviewer for their time and comments. We have addressed each one of these below.

209 *1) It could be worthwhile to put the size of landslide deposits that you can identify in more context with*
210 *the size of glaciers that you can reliably search over and/or something about the size distribution of*
211 *glaciers around the world. You don't have to answer this but it made me curious: over what proportion*
212 *of the total number of glaciers would be possible to detect a rock avalanche of the size that you search,*
213 *assuming that an event occurred? Is this the same as the base number of glaciers with debris mentioned,*
214 *which was 4.4% of 215,547 glaciers worldwide? And, the abstract mentions >2km² area but around*
215 *line 43 the mention is volume. It would be helpful to relate these together and also indicate how volumes*
216 *are estimated. With respect to the events it may also help to explain why these are referred to as 'high*
217 *magnitude' - is this your designation?*

218 We thank the reviewer for raising an interesting point. We will reword line 26 to:

219 “There are currently >200,000 glaciers worldwide covering >700,000 km², of which 8.2% are less than
220 1 km² (Herreid and Pellicciotti, 2020), excluding the Greenland and Antarctic ice sheets (RGI
221 Consortium, 2017).”

222 The abstract states that for the area of the deposits we show examples of, our actual minimum deposit
223 size that we have confidence in is 0.05 km². This is mentioned on Line 216. We will omit the volume
224 from line 43 and utilise “supraglacial landslide deposit” as an umbrella term for all events as volume
225 requires an (not well agreed on) empirical relationship to be applied. We will reword this section to:

226 “Here we focus on supraglacial landslide deposits (>0.05 km²), commonly associated with RAs, defined
227 as landslides: (a) of high magnitude (> 10⁶ m³); (b) perceived low frequency; (c) long runout; and (d)
228 where there is disparity between high present-day rates of slope processes above ice (Allen et al., 2011;
229 Coe et al., 2018) and expected rates based on theories of lagged paraglacial slope responses (Ballantyne,
230 2002; Ballantyne et al., 2014a).”

231 Herreid, S. and Pellicciotti, F. The state of rock debris covering Earth's glaciers, Nature Geoscience,
232 doi: 10.1038/s41561-020-0615-0, 2020.

233 *2) Another question is if this tool could detect smaller-scale events. Is it that any smaller events are not*
234 *considered rock avalanches and/or that they cannot be detected? (I thought that “rock avalanches”*
235 *were defined being >1Mm³, but I could be wrong about that) What about rock avalanche events on*
236 *already heavily debris-covered glacier surfaces - would those be detectable?*

237 As above (Point 1, line 201), we are going to be much clearer about this in the text and reword sections
238 of the manuscript referring to all large debris inputs detected (>0.05 km²) as “supraglacial landslide
239 deposits”. This is because we do not know the processes which resulted in slope failure, and, as you
240 rightly point out changing between different terms is confusing. We believe that one umbrella term such
241 as “supraglacial landslide deposit” will address this. However, we do validate GERALDINE against
242 RA deposits as this was the primary design for the tool and there are good inventories to test against, as
243 these examples have been investigated, confirming their failure/deposition process.

244 We shall add a sentence at L228 addressing multiple failures:

245 “GERALDINE can also not detect landslide debris deposition onto an existing debris cover. Therefore,
246 if a supraglacial landslide consists of multiple failures, a GERALDINE output map would only detect
247 one event, with the deposit extent being the combined total of all failures. It would be highly beneficial
248 to combine GERALDINE with seismic detection to help delineate the amount of failures that occur.”

249 *3) It could also be worthwhile putting the need / value of this tool in context with the total number of*
250 *rock avalanches of this scale that have been found to date. Was the validation set of 48 known events*
251 *chosen to span as many regions as possible, or are these all of the events that have been catalogued to*
252 *date? More context on the likelihood to find additional, unidentified events would be helpful. Another*
253 *way to expand on that could be to illustrate just how labor intensive it would be to search the Landsat*
254 *archive manually. What is the range of repeat times of Landsat? This would also help put in context the*
255 *two new events that you did identify.*

256 We provide reasons why those 48 validation RAs were suitable for this study on Line 173 and will add
257 a map of their locations in the supplementary information. The cited sources feature the largest datasets
258 of supraglacial RAs, which does induce some spatial bias but they do span a range of conditions and
259 glacial landscapes. We will add more recent context (RA frequencies in Alaska) in the paragraph
260 beginning on Line 68 (see lines 47-77 of Gioachino Roberti’s review).

261 As addressed in Gioachino Roberti’s review (reviewer 1), we will add a section in the introduction (see
262 lines 20-45 of our response to Gioachino’s Roberti’s review) detailing the Landsat data archive
263 (resolution, repeat times and data tiers) and in section 3.2 we will give an example of the local
264 requirements needed to undertake the Hayes range identification of both RAs (see lines 77-89 of our
265 response to Gioachino Roberti’s review).

266 Finding additional, unidentified deposits is one of the driving purposes of the tool, both in areas already
267 studied, and in those with no inventory. This work is underway, but, relies on GERALDINE being
268 accepted by peer review as being able to reliably identify supraglacial landslides, based on known
269 validation events. The revision of inventories and new inventories is the logical following paper.

270 *Related to this point: I may have missed it, but how computationally and user-labor intensive is applying*
271 *this tool. It sounds well beyond the scope of what a team like yours could do, but how far from possible*
272 *would it be to search all glaciers where events may have occurred for the past 37 years? Is the challenge*
273 *on the GEE computation side or on the validation side? When the latest RGI outlines come out is this*
274 *something that could be done?*

275 As addressed in Gioachino Roberti’s review (reviewer 1) (see lines 77-89 of our responses), in section
276 3.2 we will explain the amount of images GERALDINE processes for the Hayes region in 2018 and the
277 computational storage/processing/time that it saves. The challenge would certainly be validation of
278 events over very large spatial extents (necessary to ensure accuracy), but it would be possible if
279 numerous people worked on different regions and would be an interesting avenue for future work.

280 *It may be your goal to let the curiosity of the users take over here, but are there some outcomes you*
281 *think this makes possible in the short term and would advocate for (or may be doing yourselves?). Not*
282 *having a sense of how intensive the process is, I was left to wonder the scope of study that may be*
283 *reasonable to undertake - maybe in the conclusions you could indicate something about studies that*
284 *seem worthwhile? For example, is this best applied to target regions of a certain size and/ or over target*
285 *timeframes of a certain duration?*

286 We will highlight and cite a recently published paper, which calls for systematic, long-term observations
287 of RAs and the regions suggested for analysis (Bellwether sites). On line 86 we will add:

288 “These links, coupled with the availability of high spatial and temporal resolution optical satellite
289 imagery, have demonstrated the need for systematic observations of landslides in mountainous
290 cryospheric environments (Coe, 2020). Five ‘bellwether’ sites have been suggested for these purposes:
291 the Northern Patagonia Ice Field, Western European Alps, Eastern Karakorum in the Himalayas,
292 Southern Alps of New Zealand and the Fairweather Range in Alaska (Coe, 2020).

293 We think the focus should be on six main areas, which we will highlight by adding to Line 287:

294 “We suggest users should apply GERALDINE at standardised time intervals in recently identified
295 ‘bellwether sites’ (Coe, 2020) in glaciated high mountain areas undergoing rapid change i.e. Greenland,
296 Alaska, Patagonia, the European Alps, New Zealand Alps and the Himalaya, to investigate annual rates
297 of these large debris inputs.”

298 Coe, J.A. Bellwether sites for evaluating changes in landslide frequency and magnitude in cryospheric
299 mountainous terrain: a call for systematic, long-term observations to decipher the impact of climate
300 change, Landslides, doi:10.1007/s10346-020-01462-y, 2020.

301 *In the supplement it was mentioned that you compared to a Planet image. Is there anything that can be
302 said about the future of applying this tool to other image sets? I understand that Planet images are not
303 openly available, but is Landsat the only archive that makes sense to use? Is there anything to say about
304 coordinating Landsat-based results with other image sets, or does that just need to be taken on a
305 glacier-by-glacier basis? This would be important to at least mention, but doesn’t take away from the
306 achievement of getting this to work for Landsat data.*

307 This was also raised by Gioachino Roberti’s review (reviewer 1) and is a good point. We will add a
308 paragraph to the introduction explaining that the time-span of Landsat data makes it most suitable for
309 this purpose, and that Sentinel/Planet imagery incorporation is a future goal but that cloud masks for
310 these datasets are currently too inaccurate (see line 177-182 of our response).

311 *4) This is a subtle point, but it seemed like one that was done deliberately in the text so wanted to raise
312 my reaction. The title (and acronym for the tool) uses “identifying” to describe what is done by the tool.
313 And, the tool is referred to as a “detector”. However, typically the text refers to what the tool provides
314 as “highlighting” new events. It is only after user evaluation that they are “identified”. If this was
315 deliberate then I would check for complete consistency and maybe say that directly somewhere. I
316 suggest that identify (or detect) is a reasonable term for what the tool does, and then the user confirms
317 or validates that finding - but, any word choices you prefer will work as long as explained clearly and
318 used consistently. (Pay particular attention to this in the conclusion where the language seems to be
319 mixed.)*

320 We thank the reviewer for raising this important point and shall check for consistency of these terms.
321 We agree that the tool is a detector to aid identification of large debris inputs, but we also believe that
322 saying tool outputs ‘highlight’ these events is also in-keeping with the language.

323 *Another subtle point on language is if all “supraglacial landslide inputs” are the same as “rock
324 avalanches”? And, assuming that debris inputs are also the same thing? I would be check over to be
325 clear and consistent.*

326 As mentioned above and by reviewer 1 (Roberti), we will check for consistency and refer to all deposits
327 as “supraglacial landslide deposits (>0.05 km²)”. We believe that many of the supraglacial landslides
328 are likely to be emplaced through a rock avalanche process, but, this is difficult to verify, and, for this
329 tool supraglacial landslide removes any process related issues.

330 **In line comments:**

331 *It may be worth mentioning in the main text that updates to RGI can be readily accommodated. I have*
332 *seen at least one announcement that RGI v7.0 has a release target by the end of 2020. This is indicated*
333 *in the supplement but not stated directly (but maybe it is obvious).*

334 We will update line 112 to read:

335 “Any updated version of the RGI will be incorporated when available. Additionally, the RGI can be
336 replaced by the user with shapefiles of the Greenland and Antarctic ice sheets, if analysis is required in
337 these regions, or higher resolution (user defined) glacier outlines, if the RGI is deemed insufficient.”

338 *Line 35: Consider referring to point (i) as glaciological and climatological controls?*

339 Agreed. We shall change to: “(i) glaciological and climatological controls such as thrusting and meltout
340 of sub- and en-glacial sediment onto the surface (e.g. Kirkbride & Deline, 2013; Mackay et al., 2014;
341 Wirbel et al., 2018)”

342 *Line 55: “rapidly transported away from source areas” - in addition to rapid sequestration, which is I*
343 *think the point focused on in the sentence following the one where this is mentioned, is there a citation*
344 *about how runout extent of the event is different when deposited primarily on ice?*

345 We shall reword to “In supraglacial settings, landslides, where topography allows, travel much further
346 than their non-glacial counterparts due to the reduced friction of the ice surface (Sosio et al., 2012).
347 Rapid transportation away from source areas also occurs because of glacier flow. This removes the
348 simplest diagnostic evidence of a subaerial mass movement process – a linked bedrock source area and
349 debris deposit. Without the associated deposit bedrock source areas are easily mistakenly characterised
350 as glacial cirques (Turnbull and Davies, 2006)”

351 Sosio, R. et al. Modelling rock avalanche propagation onto glaciers, Quaternary Science Reviews, 47,
352 23-40, doi: 10.1016/j.quascirev.2012.05.010, 2012.

353 Turnbull, J.M. and Davies, T.R.H. A mass movement origin for cirques, Earth Surf. Proc. and
354 Landforms, 31(9), 1129-1148, doi: 10.1002/esp.1324, 2006.

355 *Line 60: Why use the term “censoring” here?*

356 We shall change to ‘visibility’.

357 *Line 108: I would change this from “present day” since the RGI v6.0 was published in 2014 and likely*
358 *stops with digitized outlines before then*

359 We will change this to “1943 and 2014”.

360 *Line 153: I’m not sure I understand the point that “GERALDINE is in effect standardised with this*
361 *global supraglacial cover map” - it would be help to expand on this point*

362 We are going to remove this now that an updated supraglacial debris cover map is available (Herreid
363 and Pellicciotti, 2020). We will reword to “We justify our 0.4 threshold based on Scherler et al. (2018)
364 who deemed it optimum for the creation of a global supraglacial debris cover map using Landsat data.”

365 Herreid, S. and Pellicciotti, F. The state of rock debris covering Earths glaciers, Nature Geoscience,
366 doi: 10.1038/s41561-020-0615-0, 2020.

367 *Line 185: I had to read this sentence a few times. Maybe stating this in terms of candidate events*
368 *(instead of outputs) or being clear that identification step is the one that the user executes and that*
369 *GERALDINE only presents candidates? (See point above, as I’m advocating for a particular language*
370 *choice, just that it is a bit more clear and consistent)*

371 We shall reword to:

372 “Of the 48 validation RAs, the user was able to correctly identify 44 of these events from GERALDINE
373 output maps, a true positive detection accuracy of 92 %. False negatives all pre-date 1991 (Figure 3),
374 giving 100% successful user identification post-1991. These false negatives can be explained by a
375 failure of Landsat satellites from imaging the RA deposit. This was due to reduced (and insufficient in
376 this case) tier 1 Landsat image availability pre-Landsat 7 within the GEE data catalogue, inhibiting
377 GERALDINE from highlighting the RA as new debris.”

378 *Line 211: Is introducing the acronym SLC necessary? It is only used once (I think). In general there*
379 *are a lot of acronyms (see comment below on Figure 1)*

380 We will remove this acronym.

381 *Line 255: Am I understanding this right that GERALDINE could not detect multiple landslide deposits*
382 *in about the same spot but at different times? This may never (or only rarely) occur, but I wasn't sure*
383 *if that was the point this sentence was trying to make. Or, something else about how the “user will have*
384 *already determined the date of these earlier supraglacial landslides”*

385 We shall add a sentence at L228 addressing multiple failures:

386 “GERALDINE can also not detect landslide debris deposition onto an existing debris cover. Therefore,
387 if a landslide consists of multiple failures, a GERALDINE output map would only detect one event,
388 with the deposit extent being the combined total of all failures. It would be highly beneficial to combine
389 GERALDINE with seismic detection to help delineate the amount of failures that occur.”

390 Clarification of Line 255 will read:

391 “If GERALDINE is run annually for multiple years, the user will be able to determine the emplacement
392 date for these earlier supraglacial landslide deposits.”

393 *Section 3.3 - it seems like it would be worth mentioning that you can do this in the abstract. That would*
394 *also help expand on the “monitoring” side of the tool’s name up front*

395 We shall reword Line 17:

396 “GERALDINE outputs maps of new supraglacial debris additions within user-defined areas and time
397 ranges, providing a user with a reference map, from which large debris inputs such as supraglacial
398 landslide deposits can be rapidly identified and monitored. We validate the effectiveness of
399 GERALDINE outputs using published rock-avalanche inventories. We then demonstrate its potential
400 in Alaska by identifying two previously unknown, large (>2 km²) supraglacial landslide deposits and
401 track the evolution of an existing supraglacial landslide deposit.”

402 *Line 285: What are the current methods that GERALDINE outperforms? Manual inspection of*
403 *individual images? - Very very minor, but I also found that the original (and widely cited) paper by*
404 *Ostrem has his last name typically spelled with a slashed O but in the original paper it is given with an*
405 *umlaut. From my reading of this it may have been an older alphabet choice and that these are the same*
406 *(<https://en.wikipedia.org/wiki/%C3%96>;*
407 *https://en.wikipedia.org/wiki/Danish_and_Norwegian_alphabet). I just wanted to point out that the*
408 *community overwhelmingly cites this paper with the author’s name using a slashed O. And, subsequent*
409 *work by Gunnar Ostrem uses the slashed O.*

410 We shall reword to:

411 “We showcase how GERALDINE does not suffer from the traditional disadvantages of current manual
412 and seismic detection methods that can cause supraglacial landslides to go undetected, by identifying
413 two new supraglacial landslides in 2018, in the Hayes Range of Alaska, one of which could not be
414 detected using existing methods.”

415 Thank you for pointing this out. We shall change the “O” in Ostrem to a slashed O.

416 *Figure 1: I would consider giving all these acronyms in the caption. Also, the second to last step isn't*
417 *quite clear - what is “both” here?*

418 All acronyms will be given in the caption. We shall change the second to last step to read “Subtract t^1
419 debris map from t^2 debris map to highlight new debris” and add both t^1 and t^2 to the previous steps
420 where applicable.

421 *Figure 2: This figure made me look back to the text to make sure if I understood that the maximum*
422 *debris extent would merge the evolution of the event. I think that is true, regardless of the search*
423 *timeframe (and somewhat dependent on the Landsat image separation). This would mean that to track*
424 *the debris transport you would first find that an event occurred and then go back and look through all*
425 *images to characterize how it evolved - this is all a user step, right? I'm thinking of your Lituya*
426 *Mountain example: if you instead ran GERALDINE for the timeframe of 2012-2014 you would get one*
427 *maximum extent and then you would have to notice that the event occurred in 2012 and was still visible*
428 *in 2013 and 2014 frames. This is still great since it is relatively little work to analyze around a particular*
429 *event compared to finding the event in the first place. Right? I think some more context on how many*
430 *events may exist and how laborious it is search individual frames may help put this in context. And, you*
431 *could say a bit more about this workflow in Section 3.3, since what is said around line 260 isn't quite*
432 *clear how that connects to what is shown in Figure 2 (if at all).*

433 Yes, this would be a user step, and is why we suggest annual or sub-annual time frames. As mentioned
434 by reviewer 1 (Roberti reviewer response lines 77-89) on Line 233 we will add a section detailing the
435 time and computational savings GERALDINE makes vs manual inspection. We will reword Line 258
436 onwards to make it clear that this movement would be over the user-specified time period:

437 “A secondary use of GERALDINE is tracking existing supraglacial landslide deposits. These deposits
438 are transported down-glacier by ice flow, although often the initial emplacement geometry is
439 characteristically deformed and spread due to differential ablation and ice motion (Reznichenko et al.,
440 2011; Uhlmann et al., 2013). GERALDINE can give an indication of deposit behaviour and movement
441 by highlighting ‘new’ debris, at the lateral and down-glacier end of the deposit, as it moves between
442 image captures (Figure 2). Differencing the distance of this new debris from the previous year’s deposit
443 extent can give an approximation of lateral spreading and glacier velocity over the user-specified time
444 period, the latter of which is often unknown at the temporal resolution of Landsat and complex to
445 calculate in high mountain regions (Sam et al., 2015).”

446 *Supplement: - Section 4.0 first paragraph should be “complementary” instead of “complimentary”*

447 We will change to “complementary”.

448 **Sam Herreid (Reviewer 3)**

449 *The article “GERALDINE (Google earth Engine supRaglAcial Debris Input dEtector) – A new tool*
450 *for Identifying and Monitoring Supraglacial Landslide Inputs” By Smith at al. describes a tool that*
451 *subtracts composite debris maps from two stacks of Landsat images, one from a period of interest and*
452 *the other from the preceding year, to isolate new debris additions. A user can then interpret this output*
453 *to locate supraglacial rock avalanches or landslides. I found the paper mostly easy to read and I think*
454 *the research objective is timely and useful. I also appreciated the user guide provided in the supporting*
455 *information. However, I think the authors stopped their tool development prematurely leaving some*
456 *fundamental elements unaddressed. My main points of concern are briefly summarized here with more*
457 *detailed comments inline below along with minor comments.*

458 We thank the reviewer for their time and comments. We have addressed each one of these below.

459 As an overarching response, the main purpose of GERALDINE is the rapid analysis of hundreds of
460 Landsat images over large areas to aid in the detection of supraglacial landslide deposits, without the
461 need for any computational/digital storage capacity or basic programming skills. This is a clear research
462 gap. We have purposely designed the tool in this basic way using GEE, to keep outputs rapid, easy to
463 use for those collating inventories, and viewable in a web browser, making it as accessible as possible.
464 The validation shows the tool is fit for purpose, with few benefits, but many disadvantages of adding
465 complexity at this stage.

466 We believe a number of the comments come back to the purpose of our work, to identify supraglacial
467 landslides rapidly, in the cloud, and with non-remote sensing expert users, and, the reviewer expertise
468 and recent publications on debris cover extents. We are not aiming (or wanting) to produce precise maps
469 of all debris cover with minimum noise, we are aiming to detect slope process inputs that are usually
470 time consuming to identify, or, not identified at all.

471 *Looking at the two map figures of the article, it is clear that, even within the GEE stack methodology,*
472 *which is in principal sound, debris cover is not confidently mapped. There is unphysical debris in the*
473 *accumulation zone and many instances of “new debris additions” that are not new debris additions.*
474 *These areas accumulate into tool output false positives that are neglected by the authors who rather*
475 *only report true positive success, leading to statements like L283-284: “GERALDINE outputs [had a]*
476 *100% successful identification”. By neglecting to calculate a metric like precision or the false positive*
477 *rate, the study is lacking a meaningful assessment of performance. I think it is reasonable to state, as*
478 *the authors do, that some of these debris map errors stem from errors in the RGI, but these then need*
479 *to be either mitigated or quantified in the error assessment of your tool.*

480 We have not designed GERALDINE to map all supraglacial debris cover in the most accurate and
481 confident way, that would require a large amount of performance enhancements and accuracy
482 assessments, as the reviewer rightly points out, and is beyond the aims of this research. Using the image
483 stacking method, GERALDINE finds the maximum debris extent. This approach would be unsuitable
484 for accurate debris cover mapping, as any temporally inconsistent/misclassified debris pixel is amplified
485 into the final debris mosaic, evidently creating some debris false positives. To map global debris cover
486 in an accurate way (which is not the aim of our work), an average approach would need to be used,
487 which has been done elsewhere (e.g. Scherler et al. 2018). Using the maximum debris extent does,
488 however, allow supraglacial landslides to be detected effectively, and is particularly useful for those
489 with a short surficial residence time e.g. landslides in accumulation zones. It is therefore wrong to think
490 of GERALDINE as a tool to accurately map all supraglacial debris cover. Instead it should be used as
491 a tool to highlight new possible supraglacial landslide deposits, which is not often done. We agree that
492 it would be optimal to do a validation in which we could quantify all true/false positives/negatives, with
493 an error matrix and associated statistics. However, due to the way the tool gets a maximum debris extent
494 using the image stacking method (if just one pixel in the image stack is debris, that pixel in the final

495 mosaic will be debris), there is no dataset we can use to perform such a validation. All existing datasets
496 rely on an average or singular image to calculate debris coverage, which is completely unsuitable for
497 validating GERALDINE outputs against. We have confidence in outputs though because the underlying
498 image classification methods (cloud removal and band ratio algorithms) work, as they have been used
499 and peer-reviewed elsewhere. We have therefore undertaken a validation in this way to provide some
500 measure of supraglacial landslide detection accuracy and believe it is suitable for these purposes. With
501 regards to allowing a user to detect supraglacial landslides, our 100% successful user identification of
502 validation RA deposits post-1994 is valid, as an expert in slope processes was able to successfully
503 identify 100 % of them from GERALDINE output maps. To begin a discussion on the tool being used
504 beyond its purpose, and it failing to do that, is not of benefit here.

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

518 Although it is relatively easy to spot landslide deposits in glacier ablation zones by viewing individual
519 images, GERALDINE eliminates the need to manually analyse the entirety of 22+ Landsat images,
520 making it 22x less time consuming for a user (any one Landsat sensors capture 22 images per year, at a
521 frequency of every 16 days. Except for 1993-1999, there are always two Landsat sensors imaging the
522 earth surface, making it likely that there are 44+ images every year). GERALDINE outputs are also
523 characterised by high contrast between new and old debris, making it much easier to identify
524 supraglacial landslide deposits and narrow the window of event occurrence. As reviewer 3 points out,
525 the main benefit of GERALDINE is the ability to spot supraglacial landslides deposited in accumulation
526 zones, which have thus far not been quantified well. With regards to debris deposition on existing debris
527 cover, this is a limitation of the tool. However, we argue that identifying debris onto clean ice (i.e.
528 expanding debris cover) is of greater importance than debris onto existing debris, which is likely to
529 have a much lower impact on reducing glacier melt, and only affects the glacier through mass input
530 (i.e. accumulation). It would of course impact frequency-magnitude estimates over these portions of a
531 glacier. We agree this issue may grow if debris cover expands over most glacial areas (as the reviewer
532 has just published on). The ability to assign a deposition date is an element we have tried to incorporate
533 to GERALDINE but by using a mosaic derived from an image stack, the metadata of pixel date/time
534 used is lost, so it is not possible using our current approach. A user would typically always want to view
535 the 'raw' time-stamped imagery that created the image stack after a positive supraglacial landslide
536 deposit ID, to investigate its characteristics. Determining an event date and time adds little workload to
537 this procedure. As mentioned previously, we know our debris maps are of lower quality than others in
538 the literature, but, again, that is not the aim of our tool. Rather than trying to accurately map the entire
539 debris cover, we aim to quickly highlight where potential supraglacial landslide events have occurred.
540 Using GERALDINE this task can be done easily by a user (especially those without expert remote
541 sensing abilities, but, with slope process expertise) without considerable resources, as we eliminate the
542 need for any specialist computing/storage/programming requirements.

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

553 In an ideal world we would provide some form of automatic detection of supraglacial landslide deposits,
554 but their size and shape vary so considerably on glaciers, particularly in steep, meandering glacier
555 terrain, with frequent cloud cover, that this was decided as unfeasible for the initial tool creation. It
556 would result in many missed deposits – especially for those where snow and ice entrainment make the
557 morphology complex. Any threshold on size and shape would almost certainly lead to some supraglacial
558 landslide deposits being overlooked. We do envisage this to be a future part of tool development, but
559 this initial version is already vastly superior in speed and processing requirements than other methods,
560 so we want it to be available to the community as quickly as possible in its current form. We also feel
561 that only basing the tool on existing validation inventories trains bias into any auto-detection, we wish
562 GERALDINE to be run with the semi-automatic approach to derive key features that may allow
563 automation of detection. With regards to looking at the differences between every image, this is possible
564 and earlier versions of GERALDINE did utilise this method. However, cloud dominated images would
565 have large areas masked out, therefore differencing them with any other image provides no useful
566 comparison, due to a lack of data from which to difference with. Using an annual stack of images gives
567 a solid baseline from which differencing can be undertaken. In addition, we would still advise manual
568 validation using original Landsat imagery, so this image differencing was deemed unnecessary, adding
569 additional processing time and user interaction with the tool. It is also likely that a user will want to
570 determine supraglacial landslide source areas, which would also require viewing the original Landsat
571 imagery.

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

578 GERALDINE has been through multiple iterations to ensure the code is written in the most
579 computationally efficient way possible with the current method (to our knowledge). It can handle much
580 bigger areas than 5000 km² but processing is significantly slower, as is panning/zooming around the
581 map in the browser, due to the way GEE computes these layers on the fly. If calculating large areas, the
582 best approach is to export them and view them in a GIS. We will explain this in the methods section of
583 the manuscript by rewording from Line 116 onwards:

584 “GERALDINE gathers all Landsat images from the user-specified date range and the year preceding
585 this user-specified date range, within the user-specified region of interest (ROI), creating two image
586 collections within GEE. Users should note that smaller ROIs and annual/sub-annual date ranges
587 increase processing speed, with processing slowing considerably with >800 Landsat images (~160-
588 1500 GB of data). The software clips all images to the ROI, applies a cloud mask, then a water mask,
589 before finally delineating supraglacial debris cover from snow and ice. GERALDINE acquires the
590 maximum debris extent from both image collections, creating two maximum debris mosaics, then

591 subtracts these mosaics and clips them to the RGI v6.0 (or user defined area if not using RGI) to output
592 a map. This map highlights debris within the user-specified time period that was not present in the
593 preceding year, which we term ‘new debris additions’. This map is viewable within a web browser as a
594 layer in the map window. However, as it is calculated ‘on-the-fly’ (Gorelick et al., 2017), large areas
595 can be slow to navigate. All files can be exported in GeoJSON (Georeferenced JavaScript Object
596 Notation) format for further analysis, including to verify if detections are discrete landslide inputs. This
597 is recommended for large ROIs. An overview of the workflow is presented in Figure 1 and the detail
598 for each step described in Sections 2.1.1–2.1.4.”

599 *Finally, there is a factor present in the quantity “new debris additions” that is not quantified or*
600 *discussed. Unstable glacier flow will produce debris structures that deviate from flow lines parallel to*
601 *a glacier’s valley wall (e.g. the surge loops on Susitna Glacier in your Figure 4) and a difference map*
602 *of debris cover over some dt will show a false gain and false loss of debris cover that is really just*
603 *debris structure translation. Where glacier flow instabilities are present, a simple difference of debris*
604 *cover maps cannot be strictly new debris additions. Herreid and Truffer, 2016 provides a discussion*
605 *on this topic. Herreid, Sam, and Martin Truffer. "Automated detection of unstable glacier flow and a*
606 *spectrum of speedup behavior in the Alaska Range." Journal of Geophysical Research: Earth Surface*
607 *121.1 (2016): 64-81.*

608 We thank the reviewer for raising this interesting observation. This is a valid point if the use of
609 GERALDINE was for accurate debris cover maps, but, as mentioned previously, we are only interested
610 in supraglacial landslide deposits. If a user familiar with glacial landslides/glacial flow was to view
611 these areas classified as new debris, they would be able to determine straight away that these were not
612 supraglacial landslide deposits, due to their size and shape. However, we will add a section about this
613 on line 277 as these features can be seen in figure 5:

614 “We note other areas are flagged as ‘new debris’ in 2013 and 2014. These are typically where glacier
615 downwasting has occurred exposing more of the valley walls, or where there has been temporal
616 evolution of the debris cover i.e. glacier flowline instabilities. These flow instabilities can cause double-
617 counting of debris when larger time windows are specified (described further in Herreid and Truffer,
618 2015). Both processes subsequently cause false classification as ‘new debris’, however, neither display
619 supraglacial landslide characteristics, so it is highly unlikely a user would mistake them for one.”

620 Herreid, S. and Truffer, M. Automated detection of unstable glacier flow and a spectrum of speedup
621 behaviour in the Alaska Range, Journal of Geophysical Research: Earth Surface, 121(1), 64-81, doi:
622 10.1002/2015JF003502, 2016.

623 **In line comments:**

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

626 We thank the reviewer for proposing an alternative title, but remain with our original wording because
627 the tool name is key to identifying its purpose and is more memorable for a user – similar for example
628 to the well-known ‘Google Earth Engine Digitisation Tool (GEEDiT)’. Furthermore, it includes the
629 name of the platform – Google Earth Engine – where the tool is executed.

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

635 We agree that terminology is not consistent throughout. As per our response to reviewers 1 and 2, we
636 will be much clearer about this in the text and reword sections of the manuscript referring to all large
637 debris inputs detected ($>0.05 \text{ km}^2$) as “supraglacial landslide deposits”. This is because we do not know
638 the processes which resulted in slope failure (although for many, and within the validation data set they
639 are almost certainly of RA origin), and, as reviewer 3 rightly points out, changing between different
640 terms is confusing. We propose that one umbrella term, e.g. supraglacial landslide deposit, will address
641 this. However, we do validate GERALDINE against RA deposits, as these examples have been
642 investigated, confirming their failure/deposition process. We will reword the abstract introduction to:

643 “Supraglacial landslides are high-magnitude, long runout events, believed to be increasing in frequency
644 as a paraglacial response to ice-retreat/thinning, and arguably, due to warming temperatures/degrading
645 permafrost above current glaciers.”

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

650 With our terminology updated to “supraglacial landslides”, we believe this section is now clear at
651 proposing a glacier specific problem. The purpose of this paper is to exactly fill the sampling/spatial
652 bias you refer to. Off glaciers subaerial landslide deposits have far longer residence time in landscapes
653 and there is less likely to be under-detection, although this can vary depending on how rapid geomorphic
654 processes are.

655 *L14: It reads like you are focusing on filling this small to medium gap but on L43 you say you focus on
656 the inputs of high magnitude, $> 10^6 \text{ m}^3$, RAs. Please clarify/fix and keep consistent throughout. L215
657 considers a 0.062 km^2 event.*

658 As per above, all terminology is to be changed to “supraglacial landslides”, with the tool an aid to detect
659 “supraglacial landslide deposits ($>0.05 \text{ km}^2$)”. As per our response to reviewers 1 and 2 we will reword
660 L43 to:

661 “Here we focus on supraglacial landslide deposits ($>0.05 \text{ km}^2$), commonly associated with RAs, defined
662 as landslides: (a) of high magnitude ($> 10^6 \text{ m}^3$); (b) perceived low frequency; (c) long runout; and (d)
663 where there is disparity between high present-day rates of slope processes above ice (Allen et al., 2011;
664 Coe et al., 2018) and expected rates based on theories of lagged paraglacial slope responses (Ballantyne,
665 2002; Ballantyne et al., 2014a).”

666 *L22: From the abstract alone you don't mention measuring area or volume or event timing, so I don't
667 quite see the jump to a global product. Further, on L118 you advise ROIs $<5000 \text{ km}^2$. Do you anticipate
668 a less computationally costly version of your method or are there HPC options in GEE? Finally, it is a
669 little strange to have a first step towards a revision, a revision implies several steps have already been
670 taken.*

671 The sole purpose of GERALDINE is to identify supraglacial landslide deposits – mentioning area,
672 volume or event timing in the abstract implies greater tool capabilities than it has, as these are manual
673 steps. We are clear that the tool only produces maximum debris cover maps. As mentioned in above
674 comments, it is possible to run the tool with areas $>5000 \text{ km}^2$. We shall substitute ‘revised’ to
675 ‘complete’. Volume requires area-volume scaling relationships that are uncertain, and timing within
676 Landsat repeats is best done with a focussed search through seismic data (as we are doing in
677 collaboration).

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

680 Agreed. We shall amend this.

681 *L27: Consider a revised global estimate of debris cover from Herreid and Pellicciotti, accepted by*
682 *Nature Geoscience, which should be available by August 2020 at this DOI: 10.1038/s41561-020-0615-*
683 *0*

684 We shall change this to read “Recent estimates suggest supraglacial debris only covers 7.3% of the area
685 of this glacier (Herreid and Pellicciotti, 2020), up from 4.4% estimated by Scherler et al. (2018). For
686 many glaciers...”

687 *L34: Either add “e.g.” to the citations or also add a citation to Kirkbride and Deline, 2013 whose Table*
688 *1 gives a more complete list of citations for expanding debris cover. Kirkbride, Martin P., and Philip*
689 *Deline. “The formation of supraglacial debris covers by primary dispersal from transverse englacial*
690 *debris bands.” Earth Surface Processes and Landforms 38.15 (2013): 1779-1792.*

691 We shall add a reference to Kirkbride and Deline (2013).

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

694 We cite Mackay et al. (2014), who provide evidence from Antarctica that subglacial sediment can melt
695 out. In their case much of this debris were rockfalls that entered in the accumulation area and reached
696 the basal zone. We’d also direct the reviewer to the literature on Blue Ice Moraine where subglacial
697 debris stores are brought to the surface by compressive flow and melt out as distinct debris bands.

698 *L35: Anderson, 2000 addresses general dispersion of medial moraines which you don’t explicitly*
699 *mention here. Does “debris store” mean extraglacial debris? This is not clear. Anderson, Robert S. “A*
700 *model of ablation-dominated medial moraines and the generation of debris-mantled glacier snouts.”*
701 *Journal of Glaciology 46.154 (2000): 459-469. L36: It might be worth distinguishing here high volume*
702 *low frequency mass movements from low volume high frequency.*

703 We shall add “(ii) dispersion of medial moraines (Anderson, 2000)” and subsequently shift ii to iii and
704 iii to iv. We shall reword to “(iv) remobilisation of ice proximal, extraglacial debris stores, particularly
705 lateral moraines (Van Woerkom et al., 2019).”

706 On line 43 we will discuss magnitude frequencies:

707 “Magnitude-frequency relationships suggest these low frequency, high magnitude events have a
708 disproportionate effect on sediment delivery (Malamud et al., 2004; Korup and Clague, 2009). One of
709 these large events mobilises enough debris to dominate overall volumetric production and delivery
710 rates, exceeding that of the much higher frequency but lower magnitude events.”

711 Korup, O. and Clague, J.J. Natural hazards, extreme events and mountain topography, 28(11-12), 977-
712 990, doi:10.1016/j.quascirev.2009.02.021, 2009.

713 Malamud, B.D. et al. Landslide inventories and their statistical properties, 29(6), 687-711, doi:
714 10.1002/esp.1064, 2004.

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

717 This is a fair point, and as mentioned above in previous comments (Line 630-645 of reviewer responses)
718 we shall remove the 10^6 m^3 volume from this section and define what we are interested in as
719 “supraglacial landslide deposits ($>0.05 \text{ km}^2$)”. Volumes require scaling laws. Part of this goes back to
720 the RA process identification, which by definition involves over 10^6 m^3 volumes.

721 *L44: “where there is disparity between current high rates of activity above ice” this is unclear.*

722 Recent research cited evidences high rates of RA activity in glacial environments, but it is expected that
723 this response is typically delayed until deglaciation (see Ballantyne references). We believe this is clear
724 if the full sentence is quoted.

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

726 Theory of delayed slope response to deglaciation. We shall change to “lagged paraglacial slope
727 responses since deglaciation (Ballantyne, 2002; Ballantyne et al., 2014)”.

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

729 We shall amend this to “relatively low elevations in the landscape”.

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

732 Operation Icebridge data can image englacial debris, and these data are sensed remotely – this is a
733 language/discipline point- many publications use for example, ‘GPR Remote Sensing in Archaeology’
734 (Springer) where a geophysical technique is used to remotely sense a target. We shall reword to “non-
735 ice-penetrating remote sensing and ground-based techniques”.

736 *L59: “[add: potentially] considerable modification”*

737 We shall apply this change.

738 *L60: “Deposited”? “Emplaced” is odd. L72: Landslides vs RA confusion here.*

739 We shall change to ‘deposited’. See earlier comments (Line 630-645 of reviewer responses) r.e.
740 Landslide/RA confusion and continuity.

741 *L87: Open access or open source?*

742 After consideration we think it is wrong to describe GERALDINE as either open access or open source.
743 Google Earth Engine requires a sign-up, so it is not 100 % open access and you cannot access the
744 underlying code of certain functions, so it is not truly open source. We shall change the text to reflect
745 this by rewording mentions of open access to “free-to-use” and remove any mention of open source.

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

747 We shall substitute ‘wide’ for ‘large’. It is however difficult to quantify because it depends on the extent
748 of glaciers in the region, the amount of Landsat images to be processed and whether a user wants to
749 view it in a web browser or export it and view it in a GIS. For example, the study area could be 10^6 km²
750 and have 10^3 km² of glacier coverage and run fine, but a study area of 10^4 km² with 10^4 km² of glacier
751 ice could cause processing issues. GERALDINE can struggle to display the results of larger areas with
752 >800 images within browsers, as they are calculated on the fly, as explained on Line 584-598 of our
753 response to reviewer 3 comments. However, if this layer was exported and viewed in a GIS, there would
754 be no issue. As mentioned on line 551-577 we will add this information to the method section.

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

757 We shall cite this study, in addition to Scherler et al. (2018).

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

759 We shall adopt this change.

760 *L118: What do you mean by “specify annual date ranges”? Are you saying the tool can only work for
761 one time window between two specified years? This seems like a pretty critical limitation to the*

762 *functionality to the tool. Are you sure GEE is the correct platform if its memory capacity is such a*
763 *bottleneck? Maybe JupyterLab is a better cloud-based platform? Or your code could select a single*
764 *optimal image of a one year stack and then make your calculations on single images? Also if you clip*
765 *the RGI first, then all of your calculations will be less computationally costly.*

766 Annual date ranges, or less, are the optimum time ranges to use. The tool can work for as many years
767 as a user wants but the outputs are affected, due to the way GERALDINE retrieves maximum debris
768 extent. As mentioned previously, any artefact is amplified into the final mosaic, so if run over 10 years,
769 there would be 10 years of artefacts in the final debris mosaic. We also see no reason to run over multiple
770 years, as the metadata cannot give a deposit date/time. Running over annual ranges not only improves
771 the visibility of supraglacial landslide deposits (due to less artefacts), it also narrows down the window
772 of occurrence, making it easier for a user to determine deposition dates with GEEDiT.

773 We are confident GEE is the correct platform for the tool because it is a familiar environment for a user
774 with no experience of programming, it is free for researchers, has a large data catalogue and has suitable
775 computational capabilities, allowing for further development. With regards to optimum images, these
776 would ruin the ability of the tool to detect deposits which occur in accumulation zones and are
777 consequently only visible in one image. See lines 563-571 of our responses for why a single image
778 method was not applied.

779 Although we welcome suggestions regarding tool efficiency, clipping to the RGI first is in fact much
780 more computationally costly. Earlier versions of GERALDINE processed images by clipping to the
781 RGI first, as we came to the same conclusion, but clipping is a memory intensive task in GEE, and the
782 RGI has thousands/millions of vertices. This made clipping every image to the RGI pre-analysis, 75%
783 more memory intensive and subsequently 60% more time intensive.

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

802 Please refer back to our responses on lines 543-598 about single images, large ROIs and automatic
803 detection of RAs. We disagree with the “user-specified date range” word changing because the tool has
804 two main uses: (1) finding supraglacial landslides on an annual/sub-annual basis; and (2) finding very
805 recent RAs. For example, if a RA derived seismic signal has been detected, this seismic signal can only
806 locate to within a 100 km² radius. A user can then utilise GERALDINE, specify a short date range, and
807 incorporate real-time Landsat imagery within this 100 km² radius, to identify its location. As mentioned
808 in our response on lines 572-598, we will amend and expand on the ROI size requirements, to make it
809 clear that larger areas are fine, and explain the caveats that come with increasing a ROI.

810 *L131: I can appreciate that the method used to assess cloud mask performance considers clouds in an*
811 *entire stack, thus incorporating a variety of cloud types in a simple run of your code. However, I would*
812 *like to see more direct evidence that clouds themselves are accurately mapped. From my experience*
813 *cloud mapping algorithms are unreliable in glacierized areas. Could you show a side by side image of*
814 *a raw satellite image and an overlay of the output of the cloud mask with scores 20%, perhaps one*
815 *where it worked well and a second where it was at its worst. I'm concerned that you're only mapping*
816 *60% of RA area. How were the studies that make up your validation dataset able to map 100% of the*
817 *RA area and you cannot? Surely with the stack methodology the aggregate over many images should,*
818 *together, capture 100% of RA area unless it's a particularly snowy or cloudy year. Does this suggest*
819 *you have a 40% error rate in detecting RAs?*

820 We have found the GEE in-built cloud mask to be surprisingly good in glacierised regions, but a side-
821 by-side comparison of a good and bad example is a great idea, which we will include in the
822 supplementary information. As mentioned in the manuscript (line 170), we use area as a proxy for how
823 easy it is for a user to identify these deposits. It is unlikely that a 100% deposit area detection could be
824 achieved because of the way supraglacial landslide deposits are often partially advected into the ice and
825 unpredictably entrain snow and ice during transport. If Landsat imagery does not image a deposit within
826 hours/a few days of occurrence, it is highly likely that a 100% deposit area detection is unachievable
827 with the available Landsat imagery. Some of the validation dataset utilised RAs of this nature, hence
828 the average 60% area detection. Many manual detections are not 100% of a supraglacial landslide
829 deposit, they are an interpretative map which is more difficult with increasing time from deposition,
830 and, we are attempting to validate against these often imperfect (not a criticism) data.

831 *L132: What about cast shadows from topography? Herreid and Pellicciotti, accepted by Nature*
832 *Geoscience (available August 2020 at DOI: 10.1038/s41561-020-0615-0) found it necessary to remove*
833 *area in shadow in order to accurately map debris cover. The band ratio method is able to negotiate*
834 *some shading, but when a surface becomes too dark there is still the possibility for false positive debris*
835 *classification (e.g. Herreid and Pellicciotti, 2020 removed 760 km² of shaded glacier area in Alaska*
836 *and Western Canada).*

837 As mentioned at the start of our responses to reviewer 3, the method of GERALDINE is unsuitable for
838 accurately mapping all supraglacial debris cover. As we explain for cloud shadow in the manuscript,
839 masking shadow has minimal effect on the user's ability to identify supraglacial landslide deposits,
840 whilst greatly increasing processing complexity and time (L132 of the manuscript). Any shadow is
841 highly unlikely to be lobate and elongated; the typical characteristics a user would look for in a
842 supraglacial landslide deposit. After running GERALDINE for a 90,000 km² area of Alaska on an
843 annual basis from 1984-2019 (results not presented here), we have not had one instance (to date) of
844 supraglacial landslide misidentification because of topographic or cloud shadow.

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

851 We thank the reviewer for raising this interesting point. Based on these comments and on reflection, we
852 agree that mapping of supraglacial lakes/ponds is unnecessary. We originally implemented it to reduce
853 misclassification of new debris, but these are so small it is not necessary. A landslide deposit
854 would/could cover any lake/pond during deposition and lakes display no supraglacial landslide deposit
855 characteristics, so misclassification/misidentification is not an issue. We have determined it has no
856 effect on the detection results and will therefore remove this from the code and manuscript.

857 *L150: One of your inequality signs should include “or equal to”*

858 We will amend this to “and snow/ice (≥ 0.4)”.

859 *L158: There is a missing discussion on double counting translated debris features that deviate from a*
860 *flowline parallel the glacier valley walls. Also summed non-RA debris additions if the user defined time*
861 *period is not sufficiently short. Herreid and Truffer, 2016 established a very similar methodology to the*
862 *one presented here in order to detect glacier flow instabilities. In this study RA are identified but*
863 *considered an error in the context of the flow instability research question. For your work, RAs are*
864 *signal and the features identified by Herreid and Truffer, 2016 are errors. These should be discussed.*
865 *Herreid, Sam, and Martin Truffer. "Automated detection of unstable glacier flow and a spectrum of*
866 *speedup behavior in the Alaska Range." Journal of Geophysical Research: Earth Surface 121.1 (2016):*
867 *64-81.*

868 On line 277 we will add:

869 “We note other areas are flagged as ‘new debris’ in 2013 and 2014. These are typically where
870 downwasting has occurred exposing more of the valley walls, or where there has been temporal
871 evolution of the debris cover e.g. glacier flowline instabilities. These flow instabilities can cause double-
872 counting of debris when larger time windows are specified (described further in Herreid and Truffer,
873 2015). Both processes subsequently cause false classification as ‘new debris’; however, neither display
874 supraglacial landslide characteristics, so it is highly unlikely that a user would mistake them for one.”

875 Herreid, S. and Truffer, M. Automated detection of unstable glacier flow and a spectrum of speedup
876 behaviour in the Alaska Range, Journal of Geophysical Research: Earth Surface, 121(1), 64-81, doi:
877 10.1002/2015JF003502, 2016.

878 *L162: What do you mean by “Debris biased”?*

879 We think it is clear what is meant by this, as the previous two sentences explain how debris always
880 takes precedence over snow/ice in the final mosaics. We will amend so the user is directed once again
881 to Figure 2: “GERALDINE is therefore debris biased due to this processing step (Fig. 2)”.

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

884 The validation was undertaken in two stages, so we will reword to ‘A two-stage validation was
885 undertaken...’

886 *L172: RA already defined.*

887 We shall change to read “Validation was performed against the already defined supraglacially deposited
888 RA databases of...”

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

892 We shall update this to reflect the total number of RAs in these databases. We will reword this sentence
893 to:

894 “Forty-eight events out of a total of 325 met these criteria, their locations distributed across the
895 European Alps, Alaska, New Zealand, Canada, Russia and Iceland.”

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

897 We agree that this would be a useful addition and shall add an additional figure in the supplementary
898 information with the locations of all validation RAs depicted on a world map.

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

901 Please see earlier comments (lines 543-598 and 802-830 of our response) as to why this method was
902 not used. We want to exploit all imagery, increasing our chances of detecting supraglacial landslide
903 deposits with short surficial residence times. These deposits may only be visible in one Landsat image
904 and are commonly missed by manual imagery analysis. This is due to time constraints only allowing
905 analysis of one or two optimum images in a year, particularly over large areas. Over 20+ images can be
906 discarded annually because of this, all of which may contain new supraglacial landslide deposits. This
907 is particularly true in images with high-percentage cloud cover that are commonly discarded from
908 manual analysis, but in the rare cloud-free areas of the image, may contain new, unknown deposits,
909 with short surficial visibility to optical remote sensing. GERALDINE can exploit all data from these
910 images that are typically discarded, making it a valuable time-saving tool for a user identifying
911 supraglacial landslide deposits.

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

914 We shall amend this sentence to read “However, a true 100 % detection rate for supraglacial landslide
915 events on glaciers is unlikely, due to some deposits running out over existing debris cover, and some
916 having high snow/ice content or entraining large amounts of snow/ice during events, which can be
917 common for supraglacial landslides deposited onto glaciers.”

918 And as mentioned in our response on lines 381-392, we shall add a sentence at L228 addressing multiple
919 failures:

920 “GERALDINE can also not detect landslide debris deposition onto an existing debris cover. Therefore,
921 if a supraglacial landslide consists of multiple failures, a GERALDINE output map would only detect
922 one event, with the deposit extent being the combined total of all failures. It would be highly beneficial
923 to combine GERALDINE with seismic detection to help delineate the amount of failures that occur.”

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

926 We shall amend this sentence to read “This is particularly applicable to small (<0.5 km²) glaciers, where
927 the overall significance of a single pixel increases.”

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

930 We refer to earlier responses explaining that GERALDINE is not a tool to accurately map all debris
931 cover. Despite this, noise is apparent in regions where temporary surficial debris cover is likely (so
932 classification as debris would be correct) i.e. at the bare ice-debris interface, and/or there are
933 discrepancies with the RGI i.e. where surface lowering has occurred, exposing more nunatak/valley
934 highlighting “new debris” around them.

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

941 Please see earlier comments (lines 899-911 of our response) as to why this method was not used. Two
942 optimal satellite images would certainly provide a prominent signal but then you lose a large amount of
943 temporal data, which is crucial for detecting supraglacial landslides in accumulation zones that may
944 only appear in one Landsat image. We have developed the tool to use every available cloud-free pixel,
945 to extract the maximum amount of potential debris from an image stack. The image stacking method is
946 what makes the tool unique and allows it to extract the maximum amount of debris information. We
947 disagree that this creates a ‘muddy’ signal because supraglacial landslide deposits are always easily
948 identifiable to a user. We therefore stand by this statement of justification.

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

954 This is true. However, as explained previously, for any one location there are 22+ raw Landsat images
955 in a year. Some of these may be neglected by a user because of high cloud cover, but these neglected
956 images may contain new supraglacial landslide deposits that have been advected into the ice by the time
957 the next image has been captured. The purpose of GERALDINE is to allow a user to aggregate and
958 extract all supraglacial debris information from every Landsat image, within their timeframe of interest.
959 As above (line 572-598 and 746-754 of our responses), we shall add a paragraph to better define the
960 spatial domain of the tool and how it can be run over much larger areas, depending on certain variables.
961 From our experience of using the tool in Alaska, we know that it allows a user to drastically improve
962 upon existing RA inventories, with current underestimation from initial analysis suggesting 50% of
963 supraglacial landslides are not found by manual analysis of raw images (manuscript in prep.). The major
964 point also remains, the time and computing capacity (both processing and storage) saved in looking
965 through GERALDINE outputs versus raw image investigation is considerable, and, is a large scientific
966 justification.

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

969 Please see L177-182 of the manuscript. We shall reword to:

970 “GERALDINE was run for the year of the event using Landsat tier 1 imagery; the new debris vector
971 output file was exported into a GIS and after an initial qualitative step to see if the user would flag the
972 RA from the GERALDINE output, the area of the deposit it detected was calculated within the GIS.”

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

976 No user interpretation was involved with isolating the RA area. We utilised the select by location tool
977 in QGIS, to select any pixels/pixel clusters within/intersecting the digitised RA polygon, and clipped
978 these pixels to the RA polygon, before calculating their area. We shall amend L180 to read “We utilised
979 the select by location tool in QGIS, to select any pixels/pixel clusters within/intersecting an outline of
980 the RA manually-digitised from a Landsat image using the Google Earth Engine Digitisation Tool
981 (GEEDiT) (Lea, 2018). We clipped selected pixels to the manually digitised RA outline and calculated
982 the area of these selected pixels.”

983 We refer back to line 471-504 of our reviewer responses but shall reiterate again that we agree about
984 detection confidence; it would be optimal to do a validation in which we could quantify all true/false
985 positives/negatives, with an error matrix and associated statistics. However, due to the way the tool gets
986 a maximum debris extent using the image stacking method (if just one pixel in the image stack is debris,

987 that pixel in the final mosaic will be debris), there is no dataset we can use to perform such a validation.
988 All existing datasets rely on an average or singular image to calculate debris coverage, which is
989 completely unsuitable for validating GERALDINE outputs against. We have confidence in outputs
990 though because the underlying image classification methods (cloud removal and band ratio algorithms)
991 work, as they have been used and peer-reviewed elsewhere. We have therefore undertaken a validation
992 in this way to provide some measure of RA detection accuracy and believe it is suitable for these
993 purposes.

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

995 Topographic shading is likely to be masked out of composites, as mentioned on L227, so it would not
996 influence new debris detection. If topographic shading was to be classified as new debris detection, as
997 mentioned above we do not mask it due to any artefact it produces (which is minimal), not displaying
998 supraglacial landslide characteristics and therefore not being flagged as a false positive. We found NDSI
999 to perform sufficiently well in shaded areas. In addition, it requires additional computational capacity
1000 and subsequently increases analysis time for very little benefit with regards to supraglacial landslide
1001 detection.

1002 *L247: Your method has a high potential to detect all events [add: that are deposited onto initially bare*
1003 *glacier ice]. E.g. a hypothetical second event at the same scarp on the glacier east of Maclaren Glacier*
1004 *that deposited a slightly smaller volume of rock would be entirely missed by your method.*

1005 And as mentioned in our response on lines 381-392, we shall add a sentence at L228 addressing multiple
1006 failures; “GERALDINE can also not detect landslide debris deposition onto an existing debris cover.
1007 Therefore if a supraglacial landslide consists of multiple failures, GERALDINE would only indicate
1008 one event, with the deposit extent being the combined total of all failures.” This is of course a problem
1009 any manual identification of deposits will have from remote sensing. Running GERALDINE alongside
1010 seismic detections has the best chance of resolving this, seismic noise will likely be high for a landslide
1011 overrunning another rough, angular deposit.

1012 *L250-256: I find this to be significant conditionality and required prior knowledge for an automated*
1013 *tool. Your method doesn't automatically run for multiple years sequentially, so how would someone*
1014 *new to the area know where to start? Reading your Figure 4 alone suggests the BRG RAs were*
1015 *deposited between 2017 and 2018, this is misleading. The mapped Lituya RA in Fig. 5 also appears*
1016 *patchy, should the logic of L254 be followed and this area be dismissed as erroneous?*

1017 We believe it is unlikely anyone would find/use the tool without first seeing the manuscript. If they
1018 have not seen the manuscript, we link and advise reading it on the welcome screen that greets a user
1019 when they run GERALDINE. We shall change the figure caption of Fig. 4 to “d) 2018 erroneous tool
1020 detection of Black Rapids glacier RA deposits, which were deposited as a cause of the 2002 Denali
1021 earthquake (Jibson et al. 2006).” With regards to Fig. 5, the user has identified that a RA has occurred
1022 here and can be confident of its down-glacier movement, as the leading edge is typical of a RA, being
1023 both elongated and lobate. There is also no noise around the deposit leading edge which could
1024 compromise these measurements. GERALDINE debris maps are no different to most other products
1025 utilised at user discretion. This is just another example of how GERALDINE outputs can be used, and
1026 we explain the errors associated with it.

1027 *L257: While translated features are present in your output (also translated features from flow*
1028 *instabilities, see Herreid and Truffer, 2016) and are scientifically useful, these are errors with respect*
1029 *to your intended tool function. If you can automatically differentiate feature translation from feature*
1030 *deposition then this can be a nice side component to your study, otherwise I think you need to treat this*
1031 *as error. Herreid, Sam, and Martin Truffer. "Automated detection of unstable glacier flow and a*
1032 *spectrum of speedup behavior in the Alaska Range." Journal of Geophysical Research: Earth Surface*
1033 *121.1 (2016): 64-81.*

1034 As mentioned previously, this is caused by the tool calculating a maximum debris extent, which we
1035 believe is the optimum method using this workflow, for detection of all supraglacially landslides (those
1036 deposited both in the accumulation and ablation zones). It is not the optimum method for accurately
1037 mapping a glaciers debris cover. It would be difficult to remove these translated features without
1038 inhibiting the performance and subsequent usability of the tool. See above comments and changes
1039 associated with line 277:

1040 “We note other areas are flagged as ‘new debris’ in 2013 and 2014. These are typically where
1041 downwasting has occurred exposing more of the valley walls, or where there has been temporal
1042 evolution of the debris cover e.g. glacier flowline instabilities. These flow instabilities can cause double-
1043 counting of debris when larger time windows are specified (described further in Herreid and Truffer,
1044 2015). Both processes subsequently cause false classification as ‘new debris’, however, neither display
1045 supraglacial characteristics, so it is highly unlikely a user would mistake them for one.”

1046 *L275: How does reduced ablation over one year around the ELA, where ablation rates are generally*
1047 *low, increase surface velocities?*

1048 We should be clearer that we mean reduced ablation under the deposit, causing debris expansion. We
1049 shall amend to “We suggest that the higher RA deposit velocities between 2012 and 2013 are a result
1050 of the immediate response of the glacier to reduced ablation rates directly beneath the debris, causing
1051 an ice-pedestal to form, from which debris is redistributed through avalanching off the sides, expanding
1052 debris coverage (Reznichenko et al., 2011).”

1053 *L281: RAs on bare glacier ice in ablation zones are easy to identify from one recent image and your*
1054 *method also requires manual inspection. Here I think you should highlight your tool’s ability to*
1055 *potentially catch events in the accumulation zone that have only a very short residence time.*

1056 We shall reword the conclusion introduction to “GERALDINE is the first free to use resource that can
1057 rapidly highlight new supraglacial landslide deposits onto clean ice for a user-specified time and
1058 location. It can aggregate hundreds of Landsat images, utilising every available cloud-free pixel, to
1059 create maps of new supraglacial debris additions. Using the output maps produced, GERALDINE gives
1060 an objective starting point from which a user can identify new debris inputs, eliminating the time-
1061 intensive process of manually downloading, processing and inspecting numerous satellite images. The
1062 method allows user identification of mass movements deposited in glacier accumulation zones, which
1063 have very short residence times due to rapid advection into the ice. This is a process that has not
1064 previously been quantified.”

1065 *L284: This is the first mention of 100% successful identification which should first appear in the results*
1066 *section, but I also think it is incorrect. By considering only true positive area, a map that is entirely*
1067 *“new debris additions” will also have a 100% successful identification rate but is clearly meaningless.*
1068 *You need to score your success against false positive and false negative area.*

1069 100% detection accuracy post-1991 is mentioned on L185 but we shall make it clearer by rewording
1070 L185 to “False negatives all pre-date 1991 (Figure 3), giving 100% successful user identification post-
1071 1991” See the response (lines 973-993 of our responses) above as to why false positives and negatives
1072 were not calculated. We understand that a map of 100% new debris additions would also have a
1073 successful identification, but as can be seen from our examples in Figures 4 and 5 that this is clearly not
1074 the case.