

1 Estimation of the state and parameters in ice sheet model
2 using an ensemble Kalman filter and Observing System
3 Simulation Experiments
4 – Authors’ response (RC1) –

5 Youngmin CHOI et al.

6 August 8, 2025

7 *General comments*

8 *The manuscript by Choi et al. presents a data assimilation framework to improve the projection*
9 *capabilities of ice sheet models. Specifically, the performance of an Ensemble Adjustment Kalman*
10 *Filter in constraining the model state (ice thickness) and basal conditions (basal friction coefficient*
11 *and bed topography) of a 2D plan-view ice model is assessed. Their results indicate that assimila-*
12 *ting more observations generally increases the accuracy of model projections, with projections*
13 *for up to 200 years in close agreement with the reference simulation. The performance of the data*
14 *assimilation method is sensitive to the observational error as well as the cross-track spacing and*
15 *grid resolution of surface elevation data.*

16 *I believe the science behind this study is sound and aligns with the focus of The Cryosphere (TC).*
17 *However, the presentation of the methodology lacks clarity, at times adding avoidable confusion*
18 *(e.g., the introduction of both acronyms EnKF and EAKF). This overall issue is addressed in more*
19 *detail in the specific comments below, but I strongly suggest the addition of a flowchart outlining*
20 *the methodology (ice sheet model and data assimilation) and experimental design (twin experiment*
21 *and OSSEs).*

22 We thank the reviewer for reviewing the manuscript and constructive comments. We address spe-
23 cific comments below as clearly as possible. We appreciate the suggestion to include a flowchart
24 outlining the methodology and added it into the revised manuscript (Figure 1).

25 *Furthermore, parts of the experimental design are currently placed within the results section and*
26 *key aspects of the results are not addressed (e.g., why is the RMSE C for Grid_20 km_σ_h_20_σ_v_10*

27 *smaller than for the same grid resolution with smaller uncertainties as well as all 10 km grid reso-*
28 *lution experiments?). Considering the performance of the data assimilation method is sensitive to*
29 *the uncertainty in surface elevation observations, I believe also determining the effect of various*
30 *uncertainties in the velocity data would add further value to the manuscript (perhaps as supple-*
31 *mentary material). Finally, potential reasons/explanations for model results are often missing, e.g.,*
32 *why is the range of optimal localization radius (4 - 8 km) a lot smaller than suggested by previous*
33 *studies (4 - 120 km)? I recommend the authors also take my specific comments listed below into*
34 *account.*

35 We agree with the reviewer that the some of the results lack sufficient explanation. In the revised
36 manuscript, we added more explanation for our results by addressing specific comments listed
37 below. Also we added additional experiments varying uncertainties in the velocity data in the
38 appendix.

39 *Specific comments*

40 *L25: This sentence is very similar to the second sentence in the introduction. Instead, consider*
41 *opening with a sentence about the different DA methods (variational vs. methods leveraging time-*
42 *varying observations). Then proceed to discuss advantages/disadvantages of each.*

43 We revised this sentence. Now it starts with “Data assimilation (DA) methods for ice sheet mod-
44 eling generally fall into two categories: snapshot and transient inversions, which use single-time
45 observations and time series of observation, respectively.” and continues to explain two methods.

46 *L28: Double brackets.*

47 Fixed.

48 *L33: Consider starting a new paragraph before Alternatively.*

49 We revised it to start a new paragraph here.

50 *L37 – assimilation period: Readers unfamiliar with DA might not know what exactly you refer to*
51 *here. It becomes a lot clearer later on, but it would be nice to have a brief definition here (similar*
52 *for other DA-specific terms, e.g., data denial experiments in L61).*

53 We added brief definitions of several DA terms for clarity, including the assimilation period and
54 data denial experiments.

55 *L40: Move further up to the rest of the discussion on variational methods.*

56 We revised this sentence to include the both static and transient methods.

57 *L42: Consider introducing the term ensemble DA in general before describing the specific EnKF*
58 *(e.g., ensemble DA vs. variational methods).*

59 We added an introductory sentence for the ensemble DA in general.

60 *L54: Why are ensemble DA methods less commonly used in ice sheet modelling?*

61 We added the following new paragraph to the revised manuscript in response to this point:

62 *“The ice sheet modeling community has traditionally relied on snapshot inversion methods based*
63 *on adjoint-based techniques for parameter estimation, using time-invariant mosaics or composite*
64 *data (e.g., multi-year averaged surface velocity fields; Morlighem et al., 2010). Compared to these*
65 *methods, ensemble DA approaches have been less commonly used in ice sheet modeling, primar-*
66 *ily due to historical limitations in observational data, computational cost, and the challenges of*
67 *representing uncertainty in ice sheet models. Ensemble approaches rely on time-varying obser-*
68 *vations with well-characterized uncertainties, but surface observations for ice sheets have often*
69 *lacked reliable uncertainty estimates, making them less suitable for ensemble DA. Additionally,*
70 *ensemble methods typically require multiple forward model runs, making them more computationally*
71 *demanding than snapshot inversion approaches. Another limitation is that poorly understood*
72 *or unquantified errors in the ice flow model itself may limit the reliable estimation of covariances*
73 *using ensemble statics.”*

74 *L63 – (OSSEs)(OSSEs, ...).*

75 Fixed.

76 *L65 – appropriate observation error distribution: How do you determine if the distribution is*
77 *appropriate or not?*

78 We revised this to “a prescribed observation error distribution representative of real measurement
79 uncertainties”.

80 *L71: Although it is addressed in more detail in the next sentence, I believe adding (ice thickness)*
81 *just after model state would add clarity.*

82 Added.

83 *L75 – estimated state and parameters: For consistency, I recommend using estimated model state*
84 *and parameters throughout the manuscript.*

85 Now, we use consistent terminology (model state) throughout the manuscript.

86 *L75 – true reference values: At this point in the manuscript, it is not clear what the true reference*
87 *values are and how you obtain them.*

88 We revised it to “true values”.

89 *L83-84: Remove this sentence and add the reference to the description of the specific sections in*
90 *the text above.*

91 We removed this sentence and added reference to the description of the specific sections in the text.

92 *L86-88: Repetition of the text just before the Methods section.*

93 We revised this paragraph as follows:

94 *“This section describes the ice sheet model configuration (Section 2.1), the ensemble DA frame-*
95 *work (Section 2.2), and the experimental designs used in this study (Section 2.3 and 2.4). We first*
96 *outline the twin experiment setup, which tests the ability of the DA framework to recover the model*
97 *state and parameters under idealized conditions. We then describe the OSSEs, which explore the*
98 *effects of different observational strategies on model initialization. Our methods are summarized*
99 *in Figure 1.”*

100 *L101: I am not familiar with this specific method, but a standard deviation of 500 m seems quite*
101 *large considering the bed varies only between $z_{b,deep} = -720$ m and ~ 500 m (really difficult to see*
102 *in Fig. 1a)*

103 The midpoint displacement method generates a 2D surface by iteratively subdividing a grid, assign-
104 ing random heights to corner points, and interpolating midpoints with added random displacement.
105 The magnitude of the displacement is scaled by a standard deviation that decreases with each iter-
106 ation as $2^{0.5H}$, where H is the roughness factor, set to 0.7 in this study. While the initial standard
107 deviation of 500 m may seem large relative to the vertical range of the bed topography, it is used
108 as a starting point in the midpoint displacement algorithm and is progressively reduced at each
109 iteration based on the roughness factor. This results in a realistic, spatially correlated roughness
110 pattern with limited high-amplitude variations. Additionally, the current value of $z_{b,deep} = 720m$
111 represents the base shape of the bed before roughness is added and the final bed elevation reaches
112 depths of approximately -1,500 m. We added these details to the text and revised the Fig. 1a as
113 suggested below.

114 *Fig. 1: I believe using two separate 2D plots instead of the 3D plot would make the identification*
115 *of certain details and interpretation of the plot a lot easier. As you are already showing the bed*
116 *topography in Fig. 2, I recommend combining Fig. 1 and 2 into a single plot with panels a: ice*
117 *surface elevation, b: bed topography, and c: ice velocity. Note that the rainbow colour scheme is*
118 *not in line with the journal guidelines. You can check all of your plots with the colour blindness*
119 *simulator (<https://www.color-blindness.com/coblis-color-blindness-simulator/>). The fonts in panel*
120 *b are too small and I recommend using a different colour for your contour lines.*

121 We revised this figure as suggested.

122 *Fig. 2: Why are you using such an asymmetrical (about $y=40$ km) bed topography compared to*
123 *the commonly used symmetrical approach in idealized studies? The y label in Fig. 2 indicates the*
124 *domain ranges from 20 to 100 km, whereas in Fig. 1 it is 0 to 80 km. You also might want to remove*
125 *the white margins at the top and bottom. How can the bed elevation be -1500 m when Eq. 1 limits*
126 *the bed to $z_{b,deep} = -720$ m?*

127 Applying the midpoint displacement method results in an asymmetrical bed topography, which
128 may better reflect realistic subglacial features, although we use an idealized twin experiment in this
129 study. We included this explanation and revised the figure as suggested.

130 *L103: Consider adding an additional panel (d) showing the triangular mesh to the new Fig. 1.*
131 *Are you using adaptive mesh refinement, e.g., following the grounding line?*

132 We used an adaptive mesh based on ice velocity and included a new figure (Fig. 1(d)) in the revised
133 manuscript.

134 *L104: I suggest adding another panel (e) for the basal friction coefficient to Fig. 1 or at least refer*
135 *to Fig. 6a here.*

136 We added a reference to Fig. 6a to avoid repetition.

137 *Eq. 5: In case you are working in LaTeX, I recommend using `left(` and `right)` to get brackets of*
138 *the correct size.*

139 Fixed the bracket.

140 *Eq. 6: Same as for Eq. 5*

141 Fixed.

142 *L114: Is C in Eq. 7 different from the one described in Eq. 4? If not, then remove C is a friction*
143 *coefficient.*

144 C is the friction coefficient, and C_x and C_y are the x and y components of C , respectively. We
145 added this to the text.

146 *L117: Remove equal.*

147 Removed.

148 *L124: Do you consider a melt rate of 200 m/yr realistic given that maximum present-day melt*
149 *rates are around 100 m/yr?*

150 We set the melt rate to 200 m/yr at a depth of 800 m, which results in an actual melt rate of
151 approximately 170 m/yr beneath the ice shelf. We agree that this melt rate exceeds the maximum
152 observed present-day basal melt rates. However, in this study, we chose this value to create a strong
153 dynamic response in the model over a 200-year forecast period, ensuring that the effects of data
154 assimilation could be clearly evaluated. The elevated melt rate is not meant to represent a realistic
155 present-day climate, but rather to serve as a diagnostic tool in the context of a twin experiment. We
156 clarified this in the revised manuscript.

157 *L125-127: This belongs into results.*

158 This describes the process of creating the reference run for the twin experiment rather than present-
159 ing model results. We revised the text to clarify it.

160 *L129: Sec. 2.3 and 2.4 are referenced before 2.2.*

161 We deleted this sentence.

162 *L134: modified version of the Ensemble Kalman Filter? As I mentioned above, using EnKF and*
163 *EAKF is confusing, especially since EAKF is introduced but only used within this paragraph.*

164 We revised this paragraph to clarify the use of data assimilation terminology throughout the manuscript.

165 *L136-143: This is where I think a flowchart would really help the reader to follow the details of*
166 *your method. Ideally, the flowchart should outline the details of the EAKF and how it relates to*
167 *your specific study. For example: How do ensemble members differ? What exactly is your model*
168 *forecasting? How is the observation window specified? Which ice sheet variables are considered*
169 *in the state vector? What are the state variables?*

170 We added a flowchart (Figure 1) and a reference to that in this paragraph.

171 *L142: How would adding extra variables, like surface velocity, to the state vector affect your*
172 *results?*

173 Surface velocity is a diagnostic variable; it is not part of the state vector. We added this in the
174 revised manuscript as follows:

175 *“The state vector is augmented to include both prognostic variables and model parameters to be*
176 *estimated. Under the stress balance of SSA, the velocity is a diagnostic variable, and due to the*
177 *flotation condition, ice thickness is the only prognostic variable (Gillet-Chaulet, 2020). In this*
178 *study, the state vector includes variables ice thickness (state variable), and basal friction coeffi-*
179 *cient and bed topography (model parameters), allowing joint estimation of the model state and*
180 *parameter fields through the DA process (Fig. 1).”*

181 *L143-144: EnKFs or EAKFs? Does this challenge arise in your study? What is the ensemble size?*
182 *What are the independently observed degrees of freedom in your case?*

183 Here, we explain the general case for the ensemble Kalman filter. We revised this paragraph to
184 improve clarity and better distinguish the general description from our specific implementation.

185 *L146: stability of the EnKF?*

186 We changed it.

187 *L150: Add more detail about what exactly you mean by sampling errors.*

188 We added *“Sampling errors occurs because the ensemble based covariance is only an approxima-*
189 *tion of the true covariance, and small ensembles may not adequately capture variability across the*
190 *full state space”* to clarify.

191 *L151: What localization and inflation parameters are you examining?*

192 We added a range of both localization radii (2 to 20 km) and inflation factors (1.00 to 1.20).

193 *Sec. 2.3: You either need to embed this information into the previous section or clearly outline at*
194 *the beginning of the methods section that you are first describing the EnKF in general and then how*
195 *this general structure relates to your specific setup (with references to sections). Again, a flowchart*
196 *linking the general structure to your experiments would be helpful.*

197 We outlined the description of the methods at the beginning of the Methods section with a flowchart.

198 *L154: EnKF or EAKF? If EnKF, then why bother introducing EAKF?*

199 EnKF is the general term, and EAKF is the specific approach we use in this study. We clarified this
200 in the revised manuscript.

201 *L164: Why did you decide to use lower standard deviations? What is the plausible range?*

202 We selected lower standard deviation values (5 m for surface elevation and 10 m/yr for velocity)
203 to provide a simple and conservative baseline for the twin experiment. While these values are
204 lower than those used in Gillet-Chaulet (2020), they are still within the plausible observational
205 uncertainty ranges reported in recent literature. For example, Dai and Howat (2017) report vertical
206 elevation uncertainties below 5 m in well-constrained regions, and Mouginot et al. (2017) report
207 horizontal velocity uncertainties ranging from 5–20 m/yr depending on the region. We chose values
208 at the lower end of these ranges to isolate the performance of the DA framework under favorable
209 conditions, and we explore sensitivity to larger uncertainties in the OSSEs presented in Section 3.2.
210 We clarified this in the revised manuscript.

211 *L178-179: Are you assuming that the friction coefficient and bed topography are uncorrelated?*

212 While we do not prescribe a prior correlation between them, the EAKF uses the ensemble-based
213 cross-covariance to update both fields during the assimilation process. We included this explanation
214 in the revised manuscript.

215 *L184: What radii did you explore?*

216 The radii explored in the experiment ranged from 2 km to 20 km. We added this information to the
217 text.

218 *L186: Ensemble size of 30 to 100, but what steps exactly?*

219 We tested ensemble sizes of 30, 50, and 100, and specified this information in the revised text.

220 *Fig. 3: The font size is too small. This is generally the case for a lot of plots and I will refrain from*
221 *mentioning it again afterwards. Otherwise, I think this is a great figure supporting the description*
222 *of your OSSEs.*

223 We increased the font size for figures.

224 *L221 – reference model mesh: Do you mean the mesh used in the reference simulation?*

225 Yes, we clarified this in the text.

226 *L230-231: This information needs to come earlier.*

227 We moved this information to the Method section.

228 *L234: Can you provide any insight as to why these values lead to the minimum RMSEs?*

229 The localization radius is determined through a set of sensitivity experiments and is based on
230 the expected spatial correlation length scale of the parameters, which may depend on the size of
231 flow features or stress balance regimes. We added some discussion on this point in the revised
232 Discussion section.

233 *L238: Why is that expected?*

234 Larger ensembles generally provide better approximations of the true error covariance, reducing
235 the need for artificially inflating the covariance to compensate for sampling errors. We clarified
236 this explanation in the revised manuscript.

237 *L240: You are using inflation parameters in the text but inflation factor in Fig.4.*

238 We revised it to “inflation factor” for consistency.

239 *L243-245: For ensemble size 100, the RMSE for friction coefficient does NOT continue to decrease*
240 *steadily (increase at t=7a). The other two ensemble sizes also show a small increase just after 5*
241 *years. Similar peaks are also visible for the bed topography and ice thickness. What is causing this*
242 *increase in RMSE?*

243 We examined the small increase in RMSE in early assimilation years and found that it is likely
244 due to a temporary mismatch between the model forecast and the observations during this period,
245 potentially caused by transient model dynamics or nonlinearities in the response to assimilated
246 observations. As the assimilation continues, the filter gradually corrects these discrepancies, which
247 leads to a subsequent reduction in RMSE. These fluctuations are not uncommon in ensemble data
248 assimilation systems, especially in complex, nonlinear models where localized error growth can
249 temporarily degrade performance. We added this behavior in the revised manuscript (Discussion).

250 *L245: A larger ensemble size (100 vs. 50) actually leads to a larger RMSE after $t=15$ a for all*
251 *panels in Fig. 5. Why do you think this is the case? And what does it mean for the design of future*
252 *experiments?*

253 While larger ensemble sizes generally improve the accuracy of error covariance estimates, they can
254 also increase the sensitivity of the filter to model errors or sampling noise if not properly tuned.
255 In our experiments, it is possible that the inflation and localization parameters used for the 100-
256 member ensemble were not optimal for later assimilation periods, leading to slightly degraded
257 performance after year 15. This suggests that filter performance does not necessarily scale linearly
258 with ensemble size. It also emphasizes the importance of adaptive inflation/localization techniques
259 or diagnostics for dynamically adjusting filter settings. We added this in the Discussion section of
260 the revised manuscript as a new paragraph:

261 *“Larger ensemble sizes could improve data assimilation performance but may also introduce chal-*
262 *lenges that must be carefully managed, particularly in long assimilation periods or highly nonlin-*
263 *ear systems, as in this study. In our experiments, it is possible that the inflation and localization*
264 *parameters used for the 100-member ensemble were not optimal for later assimilation periods,*
265 *leading to slightly degraded performance after year 15. This suggests that filter performance does*
266 *not necessarily scale linearly with ensemble size and highlights the importance of adaptive infla-*
267 *tion/localization techniques or diagnostics for dynamically adjusting filter settings.”*

268 *L248: Why did you choose a localization radius of 4 km when bed topography and ice thickness*
269 *showed a minimum RMSE for 6 km with a significant increase for smaller radii?*

270 The minimum RMSE for friction occurs at a localization radius of 4 km and this radius produces
271 decent results for bed topography and ice thickness as well. We chose 4 km for illustrative purposes
272 and added this in the text.

273 *L250: Somewhere you should state explicitly that a localization radius of 4 km and an inflation*
274 *parameter of 1.12 are your optimal DA configuration.*

275 The optimal inflation factor and localization radius depend on the parameter being estimated. We
276 chose 4 km and 1.12 for illustrative purposes and clarified this in the text.

277 *L254-256: You describe the results for the friction coefficient and bed topography, but what about*
278 *the ice thickness?*

279 The pattern in the ice thickness results is very similar to that of bed topography. In our model setup,
280 surface elevation is defined as the sum of ice thickness and bed topography (surface = thickness
281 + bed). Therefore, as surface observations are assimilated, improvements in bed estimates are
282 reflected in the estimated thickness field. We clarified this relationship in the revised manuscript.

283 *L257-262: This is a description of your experimental design and should be in the methods section.*
284

285 We moved this to the Methods section.

286 *L264: Just to make sure I understood it correctly. The deterministic forecast uses the mean*
287 *(across all ensemble members) basal friction coefficient and mean bed topography but is a single*
288 *simulation (compared to running a simulation for all ensemble members and then calculating the*
289 *ensemble mean).*

290 Yes, we clarified this in the text.

291 *L268-269: I think it is quite interesting that the deterministic forecast, which is based on the*
292 *ensemble mean basal conditions, follows the reference simulation relatively closely while most of*
293 *the individual ensemble member simulations show a much smaller ice volume change. I suspect*
294 *this is due to non-linearities in the system, but it might be worth having a closer look at this.*

295 We agree with the reviewer's point. While each ensemble member represents a physically plausible
296 realization of the basal parameters, small deviations from the true field can lead to large differences
297 in modeled ice volume due to non-linear feedbacks. However, the deterministic forecast, initialized
298 with the ensemble mean of the basal fields, appears to capture the overall structure of the true
299 conditions more effectively, reducing local extremes and yielding results that are more closer to the
300 reference simulation. We added this discussion to the revised manuscript (Discussion).

301 *L271-273: So assimilating more observations leads initially to a better agreement but increases*
302 *the difference in ice volume change at the end of the forecast period. This should be addressed in*
303 *more detail in the discussion.*

304 Assimilating more observations leads to better agreement throughout the forecast period, including
305 at its end. We included specific values for ice volume loss to support this comparison in the revised
306 manuscript.

307 *L273-274: Discussing the increase in the rate of mass loss after 100 a and its implication for*
308 *sea level rise projections over the next century (compared to 200 yr projections) could be a nice*
309 *additional takeaway.*

310 We added this to the Discussion section as follows:

311 *"Notably, the 200-year reference simulation includes a phase of accelerated volume loss after*
312 *130 years, which may represent a plausible sea level rise scenario for the coming century. Our*

313 results suggest that assimilating observations even before such nonlinear transitions can still re-
314 produce accurate long-term projections—provided that the model state and parameters are well
315 constrained.”

316 *Fig. 4: Why are some experiments diverging? Why does the diverging area shift to smaller*
317 *inflation factors as the ensemble size increases? What is the reason for the sharp increase in RMSE*
318 *for localization radii smaller than the minimum RMSE? Use friction coefficient, bed elevation, and*
319 *ice thickness as labels next to the colour bar.*

320 When the localization radius is too small, it overly restricts the influence of observations on the
321 state update. This can lead to underestimation of error covariances across space and result in filter
322 divergence. In our experiments, this is evident when the localization radius falls below the specific
323 threshold of each variable (e.g., 4 km for friction and 6 km for bed topography). We included this
324 explanation in the revised manuscript (Discussion). We also revised the figure as suggested.

325 *Fig. 5: I suggest using RMSE_C, RMSE_B, and RMSE_H as y-labels or adding friction coefficient,*
326 *bed elevation, and ice thickness as titles. The description of panel c is missing. The colour coding is*
327 *somewhat confusing because you are using the same colours as in Fig. 4 but they do not represent*
328 *the same thing (ensemble size vs. friction coefficient/bed elevation/ice thickness).*

329 We revised the figure as suggested.

330 *Fig. 6: Why did you choose a more or less symmetrical (about $y=40$ km) friction coefficient but a*
331 *very asymmetrical bed topography? The units are missing in the colour bar label. You might want*
332 *to increase the spacing between panels to make it clearer which text corresponds to which panel.*

333 We revised the figure as suggested.

334 *Fig. 7: What causes the sharp grounding line extent towards higher x values at $y=10$ km in the no*
335 *assimilation panel? Units are missing. Increase spacing between panels.*

336 We revised the figure as suggested.

337 *Fig. 8: Why did you not show the difference in Fig. 6 and 7? What causes the checkerboard*
338 *pattern?*

339 We chose to show the difference in ice thickness in Fig. 8 because changes in thickness are difficult
340 to detect visually from the similar figure as Fig. 6 and 7. The artifacts observed in the ice thickness
341 are the result of the conditional random fields generated using the Kriging method, which can

342 produce “bull’s eye” patterns commonly observed between observation points. We clarified this in
343 the revised manuscript.

344 *Fig. 9: I suggest adding a legend for the different lines. Change reference run to reference sim-*
345 *ulation. Change forecast simulation to deterministic forecast simulation (all of them are forecast*
346 *simulations).*

347 We revised the figure as suggested.

348 *Sec. 3.2: Do not use abbreviations as section titles.*

349 We changed the title.

350 *L285: Change 10 to 15 km to 15 km. Or did you also test, e.g., 13 km across-track spacing?*

351 We revised this as suggested.

352 *L286: I am not sure what you mean by the performance declines due to suboptimal choices for*
353 *inflation and localization parameters. You are examining the effects of these parameters here, so*
354 *shouldn’t you be able to determine the optimal choices? Do you mean the optimal choices are*
355 *outside your tested parameter ranges? If so, you need to show results supporting this claim.*

356 We removed the phrase “suboptimal choices for parameters”.

357 *L287: Start a new paragraph before ”For the gridded”.*

358 We revised it as suggested.

359 *L288: Add the reference to Fig. 11 to the end of this sentence (currently at the end of the para-*
360 *graph).*

361 We revised it as suggested.

362 *L289: range of localization.*

363 Done.

364 *L292: we conducted: Use present tense throughout the manuscript (e.g., same issue in L403*
365 *presented).*

366 We revised the manuscript to consistently use the present tense.

367 *Fig. 10: In panels (f) and (i), why is the RMSE of localization radius 6 km and inflation factor*
368 *1.04 much larger than the surrounding values?*

369 The RMSE values are calculated over the entire domain, and localized errors can increase the
370 overall RMSE. We included additional discussion on the localization and inflation factors in the
371 revised manuscript.

372 *L297-299: Add reference to Table 2.*

373 We added the reference to Table 2.

374 *L301: RMSE values continue to decrease until the end: Add reference to Fig. 12. Panels c, f, and*
375 *i show an increase in RMSE at the end.*

376 We added the reference to Fig. 12 and an increase in RMSE. This paragraph is about when the
377 elevation error is 5 m/yr which does not an increase in RMSE at the end.

378 *L301-303: Add reference to Table 3.*

379 We added the reference to Table 3.

380 *L304: marginal improvements ... after 10–15 years: Add reference to Fig. 13. Again, RMSE*
381 *actually increases in panels f and i. What do you think causes this increase?*

382 We added the reference to Fig. 13 and described the increase in RMSE for the 20 km grid data. We
383 also added some discussions on this point in the revised manuscript.

384 *L304-305: This is discussion.*

385 We remove this sentence here since there is already a similar discussion point in the Discussion
386 section.

387 *L307: Add reference to Table 2. For RMSE_C, the smallest uncertainty leads to the second-largest*
388 *RMSE of all tested uncertainties. So DA performance does not necessarily decrease as uncertainty*
389 *increases!*

390 We agree with the reviewer that this paragraph is not well presented. We revised it to clearly
391 reference Table 2 and Fig. 12 and distinguish between the RMSE values for friction coefficient
392 (RMSE_C) and bed topography (RMSE_B) and ice thickness (RMSE_H).

393 *L308-309: Add reference to corresponding panel in Fig. 12. For the 10 km across-track data,*
394 *the maximum difference in RMSE_C is 15.43. For the 5 km data, it is only 8.65. Although the*
395 *maximum difference is much smaller in the 5 km case, you argue that it shows a decrease in per-*
396 *formance as uncertainty increases while the larger 10 km difference indicates a consistent per-*
397 *formance across different uncertainty levels. I believe your statement is primarily based on the*
398 *RMSE_B and RMSE_H results, but these details need to be spelled out!*

399 We added reference to corresponding panel in Fig. 12. In addition, we clarified the distinguish
400 between the RMSE values for friction coefficient (RMSE_C) and bed topography (RMSE_B) and
401 ice thickness (RMSE_H) in the revised manuscript.

402 *L309-311: Adding to my previous comment, if you compare Track_15km- σ_h -10- σ_v -10 to Track_15km- σ_h -15- σ_v -10,*
403 *and Track_15km- σ_h -20- σ_v -10, the DA performance increases as uncertainty increases for RMSE_C,*
404 *RMSE_B, and RMSE_H. This needs to be stated clearly and discussed in detail!*

405 We revised this paragraph to clarify.

406 *L311: Add a reference to the corresponding panel in Fig. 12. They actually show an increase in*
407 *RMSE after 15 to 20 years.*

408 We revised this paragraph to add this point.

409 *L311: Add new paragraph before “With the 1 km gridded”.*

410 As suggested, we separated the discussion of the 1 km gridded results into a new paragraph.

411 *L312: Add a reference to Table 3. What about the friction coefficient? For the friction coeffi-*
412 *cient and Track_15km, the highest uncertainty level has the smallest RMSE and, therefore, the best*
413 *performance. So DA performance does not necessarily decrease as uncertainty increases!*

414 Although the highest uncertainty level has the smallest RMSE, the lowest uncertainty level still has
415 smaller RMSE than the 10 m/yr and 15 m/yr uncertainty results. In addition to the RMSE pattern
416 during the assimilation window (Fig. 13a), we argue that the friction coefficient does not show
417 the clear pattern with varying uncertainty in surface elevation. We added this point in the revised
418 manuscript.

419 *L313: DA performance does not vary significantly across different uncertainty levels: I don't*
 420 *think I agree with this statement. Again, just looking at RMSE_C, the maximum difference across*
 421 *uncertainty levels at 1 km resolution is 2.84, while it is 10.28 at 10 km and 30.8 at 20 km. So if*
 422 *anything, the performance varies more significantly with coarser grid resolution. Additionally, the*
 423 *RMSE_C for the coarsest grid resolution (20 km) and highest uncertainty level is smaller than all*
 424 *other RMSE_C values at 20 km AND 10 km resolution! As the RMSE_C with the highest uncertainty*
 425 *and the coarsest across-track resolution is also smaller than all other values at this resolution, it*
 426 *seems unlikely that this is just a coincidence. So this really needs to be addressed.*

427 We apologize for the errors in the RMSE_C values for the 20 km grid data (caused by copy-and-
 428 paste mistakes in the latex document). We corrected these values in Table 3 and revised this para-
 429 graph accordingly.

430 *Fig. 12: Use RMSE_C, RMSE_B, and RMSE_H in y-labels. Panel labels in the second and third*
 431 *rows are the same. Actually, even the subplots themselves look the same. The panel labels seem to*
 432 *be just a copy/paste issue in your Python code, not sure about the actual data.*

433 We fixed the panel labels. The patters for RMSE_B and RMSE_H over assimilation time is very
 434 similar to each other since surface elevation is defined as the sum of ice thickness and bed topog-
 435 raphy (surface = thickness + bed). The figures in the second and third lows look very similar but
 436 not the same figures (note that values in y axis).

437 *Fig. 13: Same issues as for Fig. 12. What causes the increase in panel f after 20 years? Why is*
 438 *there such a rapid increase in RMSE in, e.g., panel d between 5 and 10 years? Why does this rapid*
 439 *increase get muted for coarser resolutions? A similar pattern occurs in Fig. 12.*

440 As mentioned above for other figures (e.g., Fig. 5), it is likely due to a temporary mismatch between
 441 the model forecast and the observations during this period, potentially caused by transient model
 442 dynamics or nonlinearities in the response to assimilated observations. Localized error growth can
 443 temporarily degrade performance. We included discussion on this point in the revised manuscript.

444 *L316: fast flowing regions: You haven't mentioned fast-flowing regions before. Are you referring*
 445 *to areas around the grounding line, where the signal-to-noise ratio of velocity is relatively high?*
 446 *I'd argue that large differences also occur for y=70-80 km and x=450-640 km, which seems to be*
 447 *a relatively slow-flowing region.*

448 We specified fast flowing regions as where velocity is larger than 100 m/yr, which corresponds to
 449 the around the grounding line in the revised manuscript.

450 *L319: fast flowing regions: Again, what about the region between y=70-80 km and x=450-640*
 451 *km?*

452 We specified fast flowing regions above.

453 *L322: “more” accurate.*

454 Done.

455 *L333: assumptions on the initial ensemble: Be more specific.*

456 We revised this to “*how the initial ensemble is generated*”.

457 *L337: relatively small ensemble size: Be precise.*

458 We specified the ensemble size.

459 *L337-338: previous studies: References are missing.*

460 We added the references.

461 *L342: a larger ensemble size could provide advantageous: Or not. Fig. 5e shows a larger RMSE*
462 *for 100 members than 50 members by $t=30$ years.*

463 We agree that increasing ensemble size does not always guarantee improved DA performance.
464 We revised the discussion to acknowledge that larger ensemble sizes can improve performance in
465 general but may also introduce challenges that must be carefully managed, particularly in long
466 assimilation periods or highly nonlinear systems.

467 *L343: What exactly should these studies investigate to identify the optimal approach?*

468 We removed this sentence, as paragraphs discussing future studies are now included only at the end
469 of the discussion section, as suggested by the reviewer.

470 *L345: similar to values from earlier studies: References are missing. What are these values?*

471 We added the references with values from those studies.

472 *L347: 4-120 km is a lot wider range than your 4-8 km. What causes these differences?*

473 The differences in the optimal localization radius likely comes from the differences in model con-
474 figuration, dimensionality, and spatial resolution. Our study uses a 2D unstructured mesh with
475 relatively fine spatial resolution, whereas previous studies using flowline models (1D) with coarser
476 grids may require broader localization to account for longer correlation length scales. We added
477 this point in the revised manuscript.

478 *L351: ... assimilating more observations, i.e. more assimilation years, to estimate ...*

479 We revised it as suggested.

480 *L352: improves accuracy of model projections: In general, yes, but the difference in ice volume*
481 *change at t=200 years is larger in Fig. 9 panel b than panel a (between red and blue line).*

482 We added “reduced uncertainty” to specify the accuracy. As the reviewer mentioned, in general
483 and during the projection period, the determinist ice volume loss forecast in Fig. 9b shows better
484 agreement with the reference simulation than in Fig. 9a.

485 *L354: XX : Add numbers.*

486 Added.

487 *L358: in this study that: Replace with here.*

488 We revised this as suggested.

489 *L358: observations maintains: Replace with observations while maintaining*

490 We revised this sentence.

491 *L361: I suggest restating what OSSE means for readers quickly skimming through the manuscript.*

492

493 We added a brief definitions of OSSEs.

494 *L361: Remove “in this study”.*

495 Removed.

496 *L361: the capabilities of OSSEs: Replace with their capabilities.*

497 Done.

498 *L365: (Table 2 and 3): If you include references here then you should also include them in previous*
499 *paragraphs of the discussion.*

500 We removed these references.

501 *L365-366: As indicated previously, this is not the case for RMSE_C in Table 3.*

502 Again, we apologize errors in RMSE_C in Table 3. We revised this paragraph to clarify our points.

503 *L369-371: As mentioned above, additional data points can actually have a negative effect.*

504 We revised this sentence.

505 *L373-374: I disagree. The friction coefficient has the largest differences in RMSE across un-*
506 *certainty levels (for all resolutions in Table 2 and similar for Table 3), so it is actually the most*
507 *sensitive.*

508 We intended to compare DA performance in estimating bed topography (and ice thickness) versus
509 the friction coefficient. We revised it to ‘friction coefficient retrieval shows no clear pattern in
510 response to the prescribed surface elevation uncertainty’.

511 *L377: will or should?*

512 We revised it to “should”.

513 *L379-380: Consolidate all future study suggestions into one paragraph.*

514 We removed some mentions of future work, and now they are included in the final paragraphs of
515 the discussion section.

516 *L398-399: Your future studies should address sentences are spread out across the entire discus-*
517 *sion. I recommend bundling all of them into one single paragraph at the end of the discussion.*
518

519 We removed some mentions of future work, and now they are included in the final paragraphs of
520 the discussion section.

521 *L398-399: Again, this is not always the case.*

522 Except for the temporary decreases in DA performance observed during the assimilation period,
523 this statement generally holds true. We clarified this point in the revised manuscript.

524 *L400: What does great accuracy mean? Be precise.*

525 We agree with the reviewer and replaced “great accuracy” with a quantitative assessment.

526 *L404: Different levels of observational uncertainty: Do you mean smaller levels of uncertainty?*

527 It is not necessarily smaller uncertainty levels; it also depends on the resolution or the track spacing
528 of the data. We clarified this in the revised manuscript.

529 *L410: Will you also upload the data files?*

530 Yes, we uploaded the data files to run the scripts.

1 Estimation of the state and parameters in ice sheet model
2 using an ensemble Kalman filter and Observing System
3 Simulation Experiments
4 – Authors’ response (RC2) –

5 Youngmin CHOI et al.

6 August 6, 2025

7 *This is a review of “Estimation of the state and parameters in ice sheet model using an ensemble Kalman filter and Observing System Simulation Experiments” by Choi et al., submitted for publication to The Cryosphere. This manuscript describes the use of an ensemble-based data assimilation system, the Ensemble Kalman Filter (EnKF), to assimilate data into a 2D large-scale ice sheet models, for the purpose of better estimating parameter values and state variables during the historical period. It follows on other studies that have explored similar methods for 1D ice sheet models, and makes the crucial step of applying such methods to a model widely used for projections. This study also adds a novel “Observing System Simulation Experiment” in which different potential observing system configurations (resolution, track spacing, observational accuracy) are tested to determine their ability to improve accuracy in estimated parameters and state.*

17 *Overall, I think this is a pretty straightforward study using well-known tools in a new way with ice sheet models, advancing the state of the art in our field. My main suggestions are to further explore certain DA and modeling choices that are unexamined in the current version of the manuscript. I have detailed these suggestions and more minor ones below.*

21 We thank the reviewer for reviewing the manuscript and constructive comments. We revised the manuscript to include additional justification for key data assimilation and modeling choices, clarify methodological decisions, and expand the discussion on the implications and limitations of our approach. We address each specific comment in detail below and aim to improve the clarity.

25 *1. The manuscript briefly describes what the EnKF is, and then indicates that the EAKF version is chosen for this study. There are multiple different flavors of the EnKF available in DART, so it*

27 *is unclear why EAKF is chosen and whether the results would be any different if another filter was*
28 *chosen. My suggestion is to describe in some more detail what is done in an EnKF and how the*
29 *EAKF is different from the standard EnKF. Additionally, either some justification for why the EAKF*
30 *was chosen and some level of justification for why that is the preferred approach when others are*
31 *available.*

32 Thank you for the suggestion; this was also raised by another reviewer. We added more details
33 about the EnKF and EAKF, and the description about the distinctions between the two approaches
34 in the Data assimilation section as follows:

35 *“In contrast to the standard stochastic EnKF—which perturbs observation-space quantities ran-*
36 *domly for each ensemble member to account for observational uncertainty—the EAKF avoids addi-*
37 *tional perturbations and instead analytically adjusts the ensemble members to match the posterior*
38 *mean and covariance determined by the original Kalman filter equations (Anderson, 2001). This*
39 *approach improves numerical stability and reduces sampling noise over stochastic EnKFs, espe-*
40 *cially for small ensemble sizes (Whitaker and Hamill, 2002). In this study, we choose the EAKF*
41 *due to its reduced sensitivity to ensemble size and improved robustness in geophysical systems, as*
42 *demonstrated in previous studies using DART (Zubrow et al., 2008; Anderson et al., 2009).”*

43 *2. One thing that is unclear from your study design is the relative importance of assimilation win-*
44 *down (e.g. 5 vs 15 vs 30 years) as compared to number of assimilation cycles. You don’t change the*
45 *frequency of observations, which may be sensible given that annual observations are reasonable*
46 *for current observing platforms. However, it is then hard to understand as a reader whether there*
47 *is something fundamental about having 20-30 years of observations related to the time scales of*
48 *ice sheet response to adjustments, or whether it is having 20-30 assimilation cycles to improve. If*
49 *the observations were more frequent (e.g. an IceSAT2-like 90 days) would it takes less time for the*
50 *EnKF to improve to the level that you show here?*

51 This is a great point. In this study, we aimed to estimate two constant-in-time parameter fields
52 and the model state on an annual basis. Given the timescales associated with the model state and
53 parameters, as well as the capabilities of current observational platforms, we chose to use annual
54 observations for simplicity.

55 We agree that the distinction between the length of the assimilation window and the number of
56 assimilation cycles needs further investigation. To address this, we conducted an additional exper-
57 iment using semiannual observations to explore the relative impact of the number of assimilation
58 cycles versus the time period over which they are applied. We included the results and a discussion
59 in the revised manuscript as follows:

60 *“In this study, we focus on estimating two constant-in-time parameter fields and the model state*
61 *using annual observations over assimilation windows of varying lengths (5, 15, and 30 years). This*

choice is motivated by both the timescales associated with glacier dynamics and the current capabilities of observing platforms. However, the relative importance of the assimilation window length (i.e., total time span) versus the number of assimilation cycles (i.e., update frequency) remains an open question. To explore this, we conduct an additional experiment using semiannual observations under the same setup as the twin experiment (Fig. A1). The results suggest that semiannual observations lead to a faster reduction in RMSE for both the model state and parameters. However, the improvement at the end of the 30-year assimilation window, compared to annual assimilation, remains limited. This limited benefit is likely due to the nature of the parameters and state variables considered in this study—constant-in-time fields and annual-scale variability—which allow sufficient information to accumulate over time for a fixed target. Once sufficient assimilation cycles have passed, the parameters become well constrained, and more frequent updates offer little additional improvement. These findings suggest that, for slowly varying or static variables, increasing observation frequency can accelerate convergence toward the true state and parameter values, but may not yield additional improvement beyond a certain number of assimilation cycles. In contrast, if parameters or states change more rapidly or nonlinearly, a longer assimilation window or more complex update schemes might be needed to achieve similar improvements. Future work should explore the sensitivity of EnKF performance to both assimilation frequency and window length to identify optimal configurations for real glacier systems with time-varying parameters and limited observation periods.”

3. A big difference between your perfect model design and a real scenario where DA might be applied is that only two constant-in-time parameter fields are unknown. In reality, (e.g.) ice viscosity and climate forcing are also likely to be poorly known (though at least climate forcing is directly observable), and climate forcing (and basal friction) may vary in time. Two possibilities that would be helpful to run some experiments to assess are:

(a) if you are mistaken about the values of other parameters, but still only estimate basal friction and topography, will the estimate of basal friction compensate for these other errors (particularly for ice viscosity which trades off quite directly with basal friction in a depth-integrated model) - there is some evidence of such compensation already happening in your estimates, see below

(b) could this DART-ISSM configuration be used to estimate multiple parameters at once (I don’t see a reason why not, but the performance may not be the same as what is found for the single parameter estimation experiments explored currently).

I get that the design of these experiments are meant to mimic and compare directly to Gillet-Chaulet 2020, but it would be useful to also push beyond their design to get closer to a realistic case where DA might be used.

Thank you for this insightful comment. We agree that in real-world applications, other parameters such as ice viscosity and climate forcing are also poorly constrained and may vary in space and

time. In this study, we intentionally limited the number of unknowns to two constant-in-time parameter fields (basal friction and bed topography) to enable direct comparison with Gillet-Chaulet et al. (2020) and to isolate the behavior of the EnKF framework in a controlled setting.

Regarding (a), we recognize that compensation effects between parameters (e.g., between basal friction and ice viscosity) may arise if other sources of uncertainty are not properly accounted for. We added a discussion of this limitation and clarified that our experimental design assumes perfect knowledge of other parameters such as ice viscosity, which is not realistic.

Regarding (b), while we agree that the DART-ISSM framework is capable of estimating multiple parameters simultaneously, we chose to start with a reduced set of unknowns to rigorously test the baseline performance of the ensemble Kalman filter. We believe this simplified setup is appropriate for the scope of the current study, which aims to provide a foundational assessment of ensemble DA performance in an ice sheet modeling context. Extending the framework to include more unknown parameters—such as ice viscosity or time-varying climate forcing—will likely introduce new complexities. These directions are indeed important and are part of our ongoing and future work, but they are beyond the scope of this initial system evaluation. We have clarified this in the revised manuscript as follows and note that this study provides a foundational step toward that goal by demonstrating the feasibility and utility of ensemble DA with a simplified setup.

“Our experimental design also assumed perfect knowledge of all model parameters except for basal friction and bed topography. This choice was made to facilitate learning about the DA system in a controlled setting and to keep the experimental setup more tractable, while also allowing for direct comparison with Gillet-Chaulet (2020). However, this approach limits the realism of experiments. In practice, parameters such as ice viscosity and climate forcing are also poorly constrained and may vary in both space and time. For example, uncertainties in viscosity may interact with basal friction during assimilation, potentially leading to parameter compensation effects. Future sensitivity studies should explore how mis-specified background parameters (e.g., biased viscosity fields) affect the estimation of other parameters and whether such compensation leads to biased or unstable forecasts. Although this study focused on estimating two constant-in-time parameter fields (friction coefficient and bed topography), the DART-ISSM framework is well-suited for the joint estimation of multiple spatially or temporally varying parameters. Extending the current configuration to include additional unknowns – such as ice viscosity, accumulation rate, or time-varying boundary conditions – represents a valuable next step toward more realistic data assimilation in ice sheet modeling.”

4. At more than one point it is suggested that using variational methods is more computationally intensive than ensemble-based DA. However, there is no real direct proof of this as you don’t perform a direct comparison and to my knowledge this has not been done in the published literature. Given that ISSM has a variational DA option already implemented, it could be valuable to compare EnKF with ISSM to EnKF with ISSM in terms of core-hours for a simple standardized run. Short of that, it would be useful to have a sense for the DART overhead? If it is negligible then

136 *I would expect ensemble-based DA to have n times the computational expense of a conventional*
137 *ISSM run where n is the number of ensemble members. Additionally, giving a sense for how this*
138 *ensemble-based approach has (or can be) parallelized would be useful. In theory, ensemble DA is*
139 *highly amenable to parallelization, but this depends on how covariance matrices are constructed*
140 *and how shared memory parallelism is handled. More details on all the computational aspects of*
141 *this new method would be very useful to include.*

142 Thank you for the great suggestion. We agree that a computational comparison between the two
143 data assimilation methods would be valuable. However, due to the fundamental differences in the
144 core computational processes of variational and ensemble based approaches, a direct comparison
145 of their computational costs is challenging and is beyond the scope of this study.

146 Variational approaches using automatic differentiation (AD) tend to have higher memory demand,
147 while ensemble DA methods primarily increase computational cost through the need to run mul-
148 tiple forward simulations. Additionally, the two approaches are implemented using different tools
149 within ISSM – variational DA is built using the AD tool, while ensemble DA is integrated via
150 DART – which further complicates direct comparison. We have included this limitation in the re-
151 vised manuscript as follows and mentioned the need for future work to systematically assess the
152 computational trade-offs between these two methods.

153 *“Although ensemble-based data assimilation offers conceptual and practical advantages, its com-*
154 *putational cost is often considered a limiting factor. In this study, we did not perform a direct*
155 *computational comparison between ensemble and variational (transient) DA approaches. Such*
156 *a comparison is challenging due to their fundamentally different implementations. For example,*
157 *variational DA in ISSM relies on automatic differentiation (AD), which can be memory-intensive,*
158 *whereas ensemble DA increases computational cost primarily by requiring multiple forward simu-*
159 *lations. However, ensemble approaches can be parallelized, as each ensemble member’s forward*
160 *run can be distributed across separate cores or nodes, and the DA process here is managed through*
161 *DART, which supports parallel computing. While formal benchmarking was beyond the scope of*
162 *this study, it would be valuable in future work to quantify computational trade-offs across DA*
163 *methods in ice sheet modeling.”*

164 *L16: less numerical model re-development*

165 Done.

166 *L26: use a form of variational*

167 Done.

168 *L28: really on observational at a single time to*

169 Done.

170 *L30: it often introduces nonphysical artifacts into the*

171 Done.

172 *L35: The use of computational techniques such as automatic differentiation in ice sheet models*

173 Done.

174 *L69: to my knowledge this is the first ice sheet modeling paper to apply OSSE, so I think you can*
175 *be more direct about this sentence*

176 We revised this sentence as suggested.

177 *L86: an ensemble...for ice sheet model initialization*

178 We revised this paragraph.

179 *L88: on model initialization*

180 We revised this paragraph.

181 *L92: simulation of ice sheets*

182 Done.

183 *L100: explain what the random midpoint displacement method is*

184 We added the details on this method.

185 *L138: model simulations*

186 We revised this paragraph.

187 *L143: I am confused here because you don't include velocity in the state vector, but later you say*
188 *it is part of what is assimilated?*

189 Velocity is an observation being assimilated, not the part of the state vector. We clarified this in the
190 text.

191 *L184: does localization as implemented preserve covariance between different variables at the*
192 *same location in space or does it simply localize along the diagonal of the covariance matrix?*
193 *For example, there should be strong correlation between the ice thickness estimate and the bed*
194 *topography estimate, and so you would be losing a significant amount of your ability to assimilate*
195 *if covariances between these two variables at the same location in space were zeroed out by*
196 *localization.*

197 Localization is applied to joint observation-state-space covariance in DART, which means that co-
198 variance across model variables at the observation location will not be damped by the localization.
199 We added the following to the revised manuscript:

200 *“Localization is applied to reduce correlations between model states projected into observation*
201 *space and the unobserved state variables, which does not explicitly damp covariances across co-*
202 *located variables.”*

203 *L205: what would happen if you had no velocity observations? How much of the performance is*
204 *due to velocity observations vs thickness?*

205 We acknowledge that this trade-off between observation types is important for real-world applica-
206 tions. As similar points are raised by another reviewer, we conducted additional experiments with
207 varying uncertainties in the velocity data and included the new results in the revised manuscript.
208 We also added a discussion on the relative importance of velocity versus elevation uncertainty.

209 *L243: I think this should refer to Fig. 5*

210 Yes. We revised it to Fig. 5.

211 *L259: mean to initialization the deterministics...full ensemble to initialization the ensemble*

212 We revised it as suggested and moved this paragraph to the Method section as suggested by another
213 reviewer.

214 *Fig 4. In the caption you mention that highly localized experiments diverge. It would be helpful to*
215 *speak to why these experiments diverge in the main text.*

216 We added this point to the Discussion section to clarify that when the localization radius is too
217 small, it overly restricts the influence of observations on the state update. This can lead to underes-
218 timation of error covariances and result in filter divergence. In our experiments, this is evident when

219 the localization radius falls below the specific threshold of each variable (e.g., 4 km for friction and
220 6 km for bed topography).

221 *Fig. 7: There are artifacts in the bed topography and ice thickness estimates that correspond to the*
222 *basal friction estimate. Can you speak to this? Is it related to how the localization is performed?*

223 The artifacts in the bed topography and ice thickness are the result of the conditional random fields
224 generated using the Kriging method, which can produce “bull’s eye” patterns commonly observed
225 between observation points. We added this to the Results section.

226 *L276: can you quantify this change in spread? by eye it doesn’t seem to change much between 20*
227 *and 30 years of assimilation*

228 We added specific values for each spread in the revised manuscript.

229 *Figure 9: can you add a legend and plot the ensemble mean as well?*

230 We added a legend and included the ensemble mean in the plot.

231 *L290: It would help to discuss what this sentence means in practice. Is prediction accuracy*
232 *degraded for this case? Or can you achieve similar results with different localization and inflation*
233 *parameters?*

234 Although minimum RMSE values were identified for the coarser-resolution cases (10 km and 20
235 km), the overall prediction accuracy is still lower than that of higher-resolution cases (e.g., 1 km).
236 We added this in the revised manuscript.

237 *L325: initial estimates for the model parameters?*

238 We revised this sentence including the suggestion.

239 *L328: correlation between both parameters*

240 We meant the “establish spatial correlation within each parameter”. We revised the text.

241 *L332: what do you mean by “initial ensembles”*

242 We meant “initial ensemble of parameters”. We revised the text for clarity.

243 *L354: need to fill in values for XX*

244 Fixed.

245 *L358: transient changes in model state but not in parameters*

246 We revised it as suggested.

247 *L366: can you speak to the limitations on this? would having 100m resolution data be even better*
248 *or proportionally so?*

249 In principle, higher-resolution data (e.g., 100 m) could further improve data assimilation perfor-
250 mance by providing finer spatial detail on surface features and more precise constraints on model
251 parameters. However, the benefit of finer resolution may decrease beyond a certain threshold due
252 to increased observational noise, modeling uncertainties, and the inherent spatial correlation scale
253 of the parameters being estimated. We added this in the revised manuscript.

254 *L370: this is a very important point that is worth highlighting in the abstract*

255 We included this finding in the abstract.