Summary:

This article examines the potential impact of future observations on reducing uncertainty in projections for the Amery Ice Shelf catchment, using a Bayesian framework. It uses an emulator of the MALI ice-sheet model with six input parameters that modulate melting, friction, rheology, and calving, and produces sea-level rise as output. The study then uses present-day cumulative ice-mass change in the catchment to refine the projections, before examining how new idealized future observations might further reduce uncertainty in future projections. It shows that substantial reductions in both SLR projection uncertainty and parameter uncertainty occur when major changes in ice dynamics take place.

The idea is highly original and, to my knowledge, is applied here for the first time in an ice-sheet modelling context. Moreover, it raises an interesting question that is also policy-relevant. The introduction, results, and discussion are well written, and the overall structure is easy to follow. However, I have some general comments, mainly concerning the methodology.

Major comments:

I am not a specialist in GP emulators, although I understand the general idea behind them. The methods section was a bit hard to follow, especially for a non-specialist like me, and would benefit from clarification to make it accessible to a broader audience. Of course, I understand that you do not want to repeat all the details from Jantre et al. (2024), and I am not asking you to add more technical detail, but rather to be clearer, to properly define the terms you introduce, and to remain consistent throughout the text (including in the Results and Discussion sections). For example, in my minor comments, I refer to terms such as "best estimate" and "trajectory", the latter of which seems to be used in different contexts. I would also suggest adding a figure to illustrate the framework, which might help non-specialists understand and follow the methods section; see, for example, Fig. 2 in Hill et al. (2021). The rest of the manuscript, including the results and discussion, is easy to follow for non-specialists, so it is really only the methods section that needs clarification to make the article understandable to all ice-sheet modellers, if that is your goal.

I also have some concerns about C_phi and C_mu, which also apply to Jantre et al. (2024). I am not sure I understand their purpose or physical meaning. My understanding from Jantre et al. (2024) is that the phi and mu fields, representing rheology and friction, are first calibrated to match an observed velocity field. These calibrated fields are then multiplied by two spatially uniform parameters, C_phi and C_mu. However, doing so might produce velocity fields that differ substantially from observations. This may help fit mass changes when the C values differ from 1, but potentially for the wrong reasons, for example if mass loss occurs in the wrong locations compared with observed dh/dt fields. For instance, if the model does not produce enough mass loss, reducing friction everywhere could lead to mass loss in the interior, where it should not occur. Here are some examples that take into account uncertainty from calibration, which makes more physical sense to me: some from my own articles, Jager et al. (2024a,b), as well as other approaches such as Recinos et al. (2023) and Jakeman et al. (2025). I understand that this is not the core of the present paper, and that it is more something that should perhaps have been addressed in Jantre et al. (2024), where it seems to be mentioned only briefly in the limitations: "In particular, the ice stiffness and basal friction scaling factors are a highly simplified means of representing uncertainties in what are actually two-dimensional spatial fields." I am not sure what the best way is for the authors to address this comment, because removing this approach in order to better match current observations would require a lot of work. However, since this is not the core of the present paper, I see one possible way to improve or clarify this point, though it is not mandatory and the authors may prefer another approach. The authors could mention and discuss that these parameters do not have a specific physical meaning, but instead allow them to create an ensemble of synthetic, artificial, and non-physical friction and rheology fields. In this sense, they do not represent reality directly, but rather different possible alternative model realities.

A point related to the previous comment is that I am wondering whether it always makes sense to calibrate all the parameters considered here. For example, the average basal melt rate may change over time, so calibrating it to present-day conditions might be reasonable for the near future, but it would likely be different after 100 years. For q, I am quite confident that it can be kept fixed, whereas for other parameters, such as sigma_max and log(gamma_0), I do not know whether it makes sense to keep them constant, as this is beyond my expertise. For C_phi and C_mu, keeping them fixed might make sense if the interpretation suggested above is adopted, as the aim would then be to identify the most appropriate global friction and rheology fields.

I finally have a general comment concerning the choice of observation, although this is less of an issue since the manuscript already includes a paragraph on this point in the limitations section. Using cumulative mass loss is clearly not the choice I would have made, especially for a study on the learning rate, because it is an observation that responds slowly and it takes time for a signal to emerge. Today, we already have observations of changes in velocity fields, grounding-line position, and elevation, which seem more appropriate for calibrating some parameters. For example, for friction and rheology parameters, evolving velocity fields would directly inform them and could possibly help resolve the equifinality problem they face (Arthern and Gudmundsson (2010)). Changes in these 2-D spatial fields would provide much more information, and it is not clear that cumulative mass loss can do so: multiple combinations of parameters might lead to the same global value but with different velocity and thickness fields; see Jager et al. (2024b) for a discussion of this global-target problem. It will also take more time for the signal to emerge: first, there are changes in geometry and velocity fields; at the same time, ice discharge changes; and only later do differences in cumulative mass loss between two ensemble members become visible. This leads to relatively late learning, after the massive mass loss of the catchment, which is likely too late for policymakers. I strongly recommend incorporating the points mentioned above into the discussion (the slow response of cumulative mass loss, the potential added value of velocity, grounding-line, and elevation observations, the risk of equifinality when using a global target, and the implications of delayed learning for policy relevance), as they would significantly enhance the manuscript. Furthermore, I believe it would be very beneficial to include a sensitivity test examining how the results vary with alternative targets, such as ice discharge or mass-change rate; this would not, in my view, require too much additional work, and while it is not strictly necessary, adding it in a supplement or appendix would be an excellent addition.

I hope these suggestions seem relevant to you and will help improve the manuscript. My major comments may be somewhat lengthy, but I would like to reiterate my congratulations on this excellent article, which is truly interesting. I have cited my previous work, but that doesn't necessarily mean you must cite it as well. I've also noted some minor corrections below, and at the end, you'll find the references I've mentioned. Hope it helps, and good luck with your revisions!

Eliot Jager, University of Helsinki

Minor comments:

L23: “While our perfect-model assumption and simplified likelihood structure…” → I suggest also mentioning the choice of observations here. In addition, I am not sure what you mean by “simplified likelihood structure”.

L52–53: I do not see why “little historical change” is an advantage here. With past dynamical changes, you would have been able to use real observations to study the learning rate, for example by starting in the past and adding observations one by one to examine the reduction in uncertainty.

L57–58: This is fine, but it may be worth discussing at some point whether the conclusions/results would be different for other sectors where dynamical changes have already been observed in the past, such as the Amundsen sector.

L75–77: It would be useful to add an appendix summarizing the model configuration from Jantre et al. (2024), including the definition, physical meaning, and effect on the model of each parameter. This would be especially useful for Sect. 4.1. It may also be helpful to describe the physical meaning of each parameter for non-ice-sheet modellers.

L83: “their SSP5 projection” → What does “their” refer to?

L92–93: In Jantre et al. (2024), I think there is also an initial calibration of the model using velocity fields to calibrate phi and mu. If you add an appendix on the model configuration, this first calibration should also be mentioned.

L95–97: To improve the stability of the rates, would it not have been possible to use a 15-year average mass-loss rate? I do not know enough about GP emulators, but does this not point to a limitation if they cannot use instantaneous rates?

L106: This part is a bit unclear to me. Do you have one emulator for each 15-year interval, and did you use the original calibration with present-day observations to obtain them?

L109–110: “Emulator skill is assessed via five-fold cross-validation (as in Jantre et al. (2024)); after validation, the emulators are refit on the full SSP5 ensemble for subsequent use.” → Similarly, it is unclear to me what exactly is being done here.

L104–113: The placement of this paragraph on the emulator is a bit surprising. Is the emulator not common to both observation generation and the rest of the analysis? If so, why is this paragraph placed here? Should it not be in a separate subsection on the emulator, or in the introductory paragraphs of Sect. 2?

L123–124: “because we cannot know the true future trajectory, we consider the 100 sampled future trajectories based on the Bayesian calibration of model parameters to the present-day observables” → I do not understand this sentence.

Sect. 2.1: After reading this paragraph, I am not fully sure how the observations are obtained. It sounds as if, every 15 years, an observation is drawn from all possible trajectories, for example a dark-blue point in 2195 and a cyan point in 2210. I do not think this is what you did, so please clarify. If this is what you did, it seems quite unusual.

Sigma_c,t,i: I am not sure how this quantity is obtained, or what its value is.

Sigma_o,t: This linear growth does not make sense to me. If you look at IMBIE3, for example Otosaka et al. (2023), the total/cumulative mass change uncertainty does not grow linearly.

L136–138: “we assume independence among the 2015 observables” → This is clearly a strong assumption. For a mass-change rate, I would agree that correlations between two observations can be neglected, since the signal may be somewhat chaotic, but for cumulative mass change the observations are clearly not independent.

L142–146: This part is quite unclear to me; perhaps a schematic would help. For each sequence of observables, do you obtain 100 sampled parameter vectors and projections?

L171–172: “plume-colored trajectory” → In Fig. 3, it would be nice to make this more distinguishable, perhaps by using a star instead of a diamond.

Fig. 3: To better distinguish the different confidence intervals, you could add thin lines for the upper and lower bounds, and perhaps use different hatching (see Jager et al. (2024b) for an example).

Results: It would be useful to easily identify the Fig. 2 trajectory, as well as the four others shown in Fig. 3, across the different result plots. One suggestion would be to give them names, such as trajectory 1, trajectory 2, etc., and use these names in the legends. You could show these trajectories in Fig. 1, and then show only these four trajectories in Figs. 4 and 6.

L187: The reduction for sigma_max is actually quite small. Would it have been possible for uncertainty to increase?

L207–208: This sentence is quite unclear to me. I am not sure what you mean by “recalibrated MALI parameters”. It would be useful to define important terms in the introduction or methods section, such as “analysis”, “recalibrated parameters”, “trajectory”, and “observation”. I am still not sure whether “trajectory” is equivalent to “observation” in the manuscript.

L211-213: “Across those trajectories, SLE projection uncertainties at 2100 (Fig. 5a) remain indiscernible as information from more future years is introduced due to overlapping very-likely ranges” → This sentence is very unclear. I think the explanation is that learning is weak because there are almost no changes in ice dynamics, so each new observation provides little new information.

L214: Similarly, I do not think that simply having more observations is what leads to the reduction; the key is to have more distinct or informative observations. However, your assumption of independence for cumulative mass loss may hide this important point. In a Bayesian framework, what matters is new information, not simply new observations, and correlations between observations are how one accounts for whether an observation is genuinely new. If it is not, it will have a high correlation with previous observations and will therefore not change the posterior very much.

L216: “across-trajectory mean” → Is this the mean of all observation trajectories? If so, I do not understand why it is called the best estimate. For me, in a Bayesian framework, “best estimate” usually refers to the mean of a posterior quantity.

Fig. 4 and 6: For both figures, I would show the global statistics, such as the mean and 66 % CI, together with only the four trajectories shown previously, so that the figures can be followed and compared more easily. Since you do not discuss individual trajectories in detail, I do not see why all of them need to be shown. To improve the explanation of these figures, you could first explain how you compute the width of the 90 % CI for each trajectory, using the four examples, and then explain how the figure generalizes this to all trajectories, which is why another interval, such as the 66 % CI and mean, is obtained.

L227: “best estimate” → I suggest using “mean”, as this would make the text easier to follow and easier to link with the figure legends. The phrase “across-trajectory best estimate of widths of the very-likely range” is rather heavy and difficult to follow.

L229-231: "the dominant control” → This seems a bit too strong, especially regarding ice-shelf buttressing. Grounding-line migration and changes in basal shear stress close to the grounding line are at least as important, if not more so. Moreover, Fig. 7 shows that the ice shelf remains quite stable until 2100, so I think another key mechanism could explain the importance of melting: melting plays a major role in grounding-line retreat, which reduces resistance and leads to an increase in ice discharge.

L228-234: I do not think the explanation “As anomalies in grounded ice discharge directly contribute to grounded mass change, our synthetic observations of cumulative grounded mass change strongly inform this parameter” is fully satisfactory. C_phi and C_mu also affect ice discharge, since lower friction or viscosity leads to higher ice discharge, so a change in ice discharge should also inform these parameters. Physically, I would suggest the following explanation. Basal melt acts as the trigger, because it affects the magnitude of grounding-line retreat. In Fig. 7, between 2020 and 2100, the grounding line does not always migrate to the same position, which is likely due to different melt rates near the grounding line associated with different log(gamma_0) values. This may occur relatively quickly, so the calibration can directly learn this parameter. I think this is also why learning for log(gamma_0) is faster than for m, because m controls the average melt rate and likely has less direct influence close to the grounding line. Then, for q, C_mu, and C_phi, some dynamical changes, such as changes in grounding-line position, are needed to identify the appropriate parameter values. Without changes in dynamics, no new information is added relative to the 2015 observations, so little learning occurs. This may explain why these parameters respond later or more slowly than log(gamma_0).

L242–244: Yes, I think this is because log(gamma_0) mostly affects grounding-line retreat; see the comment above.

L252–253: In Fig. 7, there is also a large change in grounding-line position over this period, so I think this is also a good candidate mechanism, if not a better one.

L277–279: As mentioned above, does m not affect the average melt rate, rather than specifically the melt rate at the grounding line?

L282: " This is might be due" → I am not sure, but is this formulation correct? Should it be "This might be due to…" instead?

L289–290: Once the system consists of tidewater glaciers, is there no further calving? Is there only melt at the front, or are there no further changes in ice-front position?

L280–291: Could the sigma_max range be too restrictive? In Jantre et al. (2024), sigma_max has almost no influence, but if the range were larger, we might see different frontal retreat and therefore different dynamics.

L294–302: This paragraph concerns projections to 2100. Should it therefore be in Sect. 4.2.1?

L297-299: "As seen in Fig. 5a, such a wide very-likely range around the posterior predictive mean of the 2100 SLE projections by 2090 would likely include 0 mm SLE, indicating that no sea-level contribution may occur by the end of the century." → I do not see the purpose of this sentence. More generally, this paragraph feels a bit strange, as it seems to describe results that have already been shown.

L308–309: As mentioned earlier, this may be because new observations are being added, but they do not necessarily provide new information, which is what matters for Bayesian calibration.

L323: One point I did not fully understand when comparing Sects. 4.2.1 and 4.2.2 is the following: if some learning for 2200 projections mostly occurs around 2075, why do the 2100 projections show almost no learning?

L342: This is related to the previous point. Why does the 22nd century still have large uncertainties if you have shown just before that uncertainty in 2200 projections can be reduced?

L352–353: As above, could this not be due to grounding-line migration?

L366–369: It would be useful to remind the reader which glacier area or system each of these studies focuses on.

L371: Is this not too late in terms of sea-level-rise management? This is also why I think other observations might provide an earlier signal, which would be more useful for SLR management.

L413–416: I do not particularly like citing my own papers, but I think it would be useful to also mention Jager et al. (2024b) here. We showed that using global variables as calibration targets, such as ice discharge or cumulative ice-mass loss, provides much weaker constraints than using spatial data, such as elevation or velocity changes. This is because global targets can hide compensating errors: if a model is able to reproduce velocity and elevation changes, it should also be able to reproduce ice discharge and ice-mass loss, but the opposite is not necessarily true. For example, an overestimated ice thickness can compensate for an underestimated velocity when computing ice discharge.

L441: “negative learning” → It would be useful to explain what this means, what can cause it, and what its consequences are.

L449: “50–100 years of observation” → Please specify “50–100 years of mass-change observations”, because this timescale might be quite different for other types of observations.

L475–483: Here, you should discuss that, in your case, learning may occur too late for policy-relevant questions. Policymakers are likely most interested in the timing and magnitude of substantial mass loss. Performing a similar analysis with other types of observations, as mentioned above, might help address this issue. This would then raise the question of which observations are most relevant or important.

Fig. A1: To me, it is unclear what each point represents.

Figs. B1 and B2: These figures do not seem to provide new information; perhaps they could be removed 😊

References:

Hill, E. A., Rosier, S. H. R., Gudmundsson, G. H., and Collins, M.: Quantifying the potential future contribution to global mean sea level from the Filchner–Ronne basin, Antarctica, The Cryosphere, 15, 4675–4702, https://doi.org/10.5194/tc-15-4675-2021, 2021

Jager E, Gillet-Chaulet F, Mouginot J, Millan R. Validating ensemble historical simulations of Upernavik Isstrøm (1985–2019) using observations of surface velocity and elevation. *Journal of Glaciology*. 2024a;70:e36. doi:10.1017/jog.2024.10

Jager, E., Gillet-Chaulet, F., Champollion, N., Millan, R., Goelzer, H., and Mouginot, J.: The future of Upernavik Isstrøm through the ISMIP6 framework: sensitivity analysis and Bayesian calibration of ensemble prediction, The Cryosphere, 18, 5519–5550, https://doi.org/10.5194/tc-18-5519-2024, 2024b.

Recinos, B., Goldberg, D., Maddison, J. R., and Todd, J.: A framework for time-dependent ice sheet uncertainty quantification, applied to three West Antarctic ice streams, The Cryosphere, 17, 4241–4266, https://doi.org/10.5194/tc-17-4241-2023, 2023.

Jakeman, J. D., Perego, M., Seidl, D. T., Hartland, T. A., Hillebrand, T. R., Hoffman, M. J., and Price, S. F.: An evaluation of multi-fidelity methods for quantifying uncertainty in projections of ice-sheet mass change, Earth Syst. Dynam., 16, 513–544, https://doi.org/10.5194/esd-16-513-2025, 2025.

Arthern RJ, Gudmundsson GH. Initialization of ice-sheet forecasts viewed as an inverse Robin problem. *Journal of Glaciology*. 2010;56(197):527-533. doi:10.3189/002214310792447699

Otosaka, I. N., Shepherd, A., Ivins, E. R., Schlegel, N.-J., Amory, C., van den Broeke, M. R., Horwath, M., Joughin, I., King, M. D., Krinner, G., Nowicki, S., Payne, A. J., Rignot, E., Scambos, T., Simon, K. M., Smith, B. E., Sørensen, L. S., Velicogna, I., Whitehouse, P. L., A, G., Agosta, C., Ahlstrøm, A. P., Blazquez, A., Colgan, W., Engdahl, M. E., Fettweis, X., Forsberg, R., Gallée, H., Gardner, A., Gilbert, L., Gourmelen, N., Groh, A., Gunter, B. C., Harig, C., Helm, V., Khan, S. A., Kittel, C., Konrad, H., Langen, P. L., Lecavalier, B. S., Liang, C.-C., Loomis, B. D., McMillan, M., Melini, D., Mernild, S. H., Mottram, R., Mouginot, J., Nilsson, J., Noël, B., Pattle, M. E., Peltier, W. R., Pie, N., Roca, M., Sasgen, I., Save, H. V., Seo, K.-W., Scheuchl, B., Schrama, E. J. O., Schröder, L., Simonsen, S. B., Slater, T., Spada, G., Sutterley, T. C., Vishwakarma, B. D., van Wessem, J. M., Wiese, D., van der Wal, W., and Wouters, B.: Mass balance of the Greenland and Antarctic ice sheets from 1992 to 2020, Earth Syst. Sci. Data, 15, 1597–1616, https://doi.org/10.5194/essd-15-1597-2023, 2023.