General comments

First of all we would like to thank all three reviewers for their useful comments which helped us to substantially improve our manuscript.

Upon reading the reviews we realize that in the submitted manuscript we did not succeed in formulating the scope of the paper, and the method of hindcasting, satisfactorily. Our assessment of the reviews is that the reviewers' main concerns arise due to this lack of clear formulation of the general purpose of our paper and a clear definition of the hindcasting method. In the revised manuscript we now clearly formulate the scope of the paper, and the method of hindcasting is clearly outlined. We rephrased the introduction along the lines of the text below.

Begin suggested introduction

Realistic projections of ice sheet response to a changing climate should be based on physical understanding of the processes involved, rather than trend extrapolation of historical observations (Arthern and Hindmarsh, 2006). Ice sheet models are multi-physics system models incorporating such physical process understanding. The main components are ice dynamics, surface and basal processes, and thermodynamics. Ideally, verification and validation should be done for each model component independently. Verification (i.e. the comparison of results from a numerical approximation to exact solutions of the same continuum model equations) is currently only possible for a few sub-systems (e.g. Bueler et al., 2005; Leng et al., 2013) and in a simple coupled system (Bueler et al., 2007). In engineering, validation is commonly defined as the process of comparing model results to a set of observations adequate to falsify a model (Roache, 1998). Such validation is challenging to apply in earth system modeling in general, and in ice sheet modeling in particular; nonetheless attempts have been made (Robison et al., 2010; Burton et al., 2012). Direct validation of substantial sub-systems such as basal hydrology, thermodynamics, ice dynamics is difficult or impossible as most or all observations available for validation are not linked to a single process, but are the consequence of a complex interplay between sub-systems.

However, ice sheet models can be viewed as part of earth system models, and may be evaluated in this context. In other words, we do not seek to isolate, e.g., the model of ice dynamics, and evaluate it independently; instead we evaluate how the *system* responds to a given forcing. We ask the question: "How successful is a state-of-the art ice sheet system model (i.e. the combination of physical models, their numerical approximations and implementations, and particular choices of boundary forcing and initial states) in reproducing observations of quantities such as ice thickness, and their temporal changes?" Even if all sub-systems could be verified and validated independently, testing of the system as a whole is indispensable. As a non-glaciological example, one may consider the grounding of Boeing 787 Dreamliner airplanes. While individually-tested sub-systems of the airplane received certification after testing, only testing of the airplane as a whole revealed flaws in the battery design (Wikipedia, 2013).

The aim of this study is to demonstrate that hindcasting, i.e. forcing a model with known or closely-estimated inputs for past events and comparing model results to time-dependent observations, is a viable method of assessing the performance of an ice sheet model viewed as a system model. The paper may be read as a tutorial. First we initialize an ice sheet model. Second the initialized model is integrated forward in time for a period where a wealth of data are available for validation. This produces a hindcast of this time period. Finally, the hindcasts are compared to available observations and the performance of the model system is evaluated. In an ideal case, when using appropriate model physics paired with realistic initial states and boundary conditions, a hindcast and observations should agree to within error bars associated with observations. Such an agreement with all available observations may not occur due to several reasons. Disagreement may arise from uncertainties in boundary conditions, unrealistic initial states, incomplete model physics, inadequate choices in numerical methods, to name a few. Using multiple data sets helps us pin down the source of disagreement. As models evolve the level of disagreement between observations and model simulation is expected to decrease.

Similar to short-term weather forecasts, a realistic initial state is essential for accurate simulations of the future evolution of an ice sheet (Arthern and Gudmundsson, 2010). In this study we use hindcasting to assess an ice sheet system model's sensitivity to its initial state. We show that it is possible to get good agreement with observations when too few data sets are used. One may wrongly put too much confidence in a given model simulation, not realizing that important properties are misrepresented. In particular, previous studies have often used ice volume (e.g. Stone et al., 2010; Rogozhina et al., 2011; Applegate et al., 2012) for validation of the initial state. Spatially-rich observations, especially rates of change, are, however, better metrics to evaluate the quality of initial states (Vaughan and Arthern, 2007). As an example, we show that good agreement between observed and simulated total mass loss can be achieved, and only by comparing observed vs. simulated ice discharge, it is possible to identify the good agreement of the former as being for the wrong reason.

Our manuscript is intended to serve as a guidance for future studies developing better practices in ice sheet model validation. We do not identify a preferred initialization procedure. Hindcasting is not limited to our particular choices of an ice sheet model, boundary conditions, and initial states; but is transferable to combinations. However, conclusions regarding the performance of initialization procedures may not be transferable. Hindcasting may also be helpful to study other aspects of the ice sheet system model's behavior. It is, for example, important to assess the sensitivity to climate forcing; but this is not within the scope of our study.

end suggested introduction

Consequently, changes to the abstract, methods, discussion, and conclusions were required.

Because in the revised manuscript, PISM and HIRHAM5 are not introduced in the Introduction anymore, we now do so at the beginning of the Methods section. To emphasize that our paper is about hindcasting as a method, we have deliberately chosen not to mention our particular choice of ice sheet system model in the Introduction.

To incorporate the reviewer's suggestion to discuss the limitations of hindcasting, the revised manuscript splits the discussion into two subsections. In the first we discuss the initial states and hindcasts relative to observations (this is the existing discussion). In the second we address the suitability of hindcasting as a method to evaluate the performance of an ice sheet system model. Below we include our suggested discussion on the suitability of hindcasting.

begin suggested addition to discussion

A comparison of observed and simulated rates of change requires a reference period covering both observations and simulations. Hindcasting provides simulated rates of change for this reference period. In other words, it adds a temporal dimension to validation efforts. Hindcasting enables a qualitative assessment of model performance relative to observed rates of change. This reduces the number of admissible initial states more rigorously than validation efforts that do not take advantage of observed rates of change. As an example, hindcasting allowed us to realize that our flux-correction method produces unrealistic surface elevation changes between 2003 and 2009.

At present there are both theoretical and practical limitations for using hindcasting as a method of assessing the performance of an ice sheet system model. Theoretically, the appropriate time-scale for hindcasting is unknown. Hindcasts are short (decades) compared to the time-scale associated with changes in energy (thousands of years). As a consequence, even a hindcast showing good agreement with all available observations may not capture the system's true behavior. Hence hindcasting does not identify the initial state representing the system's true state. Unfortunately the distribution of energy within an ice sheet cannot be measured directly. The age field, however, exhibits similar time-scales as energy and may thus serve as a surrogate. Practically, the duration of hindcasts is limited by the length of observational records. In consequence of those limitations, a quantitative assessment of model performance is currently not possible with hindcasting.

We have not used all available observations. Additional observations that can be used for validation are, for example, spatially-distributed mass loss estimates from GRACE (e.g. Luthcke et al., 2012) and temporal variation in ice surface velocity (Joughin et al., 2010; Moon et al., 2012). The aforementioned modeled age field could be compared to dated isochrones. Such a comparison provides a strong validation metric because the age field is, first, representative of long energy time-scales and, second, a three-dimensional field (in contrast to, for example, surface elevation). For those reasons an accidental agreement between simulated and observed age of the ice is quite unlikely. Fortunately the number of both remotely-sensed and in situ observations is constantly growing. Future studies should take advantage of these data sets for validation.

end suggested addition to discussion

Next, we updated the Conclusions section to reflect the changes to the Discussion. In particular we now provide an outlook at the end of the manuscript:

begin suggested addition to conclusions

While a direct validation in the engineering sense may not be applicable, our view of ice sheet models as part of earth system models allows validation of another type. Furthermore, hindcasting can be part of a concerted effort to validate ice sheet models. Other parts include formal sensitivity analyses to assess error propagation in forward models as, for example, carried out by Larour et al. (2012b,a). Ultimately, validation may be integrated in statistical frameworks to quantify uncertainties in ice sheet evolution due to different sources of model and observation uncertainty (c.f. Steinschneider et al., 2012, for an example in hydrologic modeling).

end suggested addition to conclusions

Finally, we rewrote the abstract to reflect the changes made to the manuscript.

begin suggested new abstract

Validation is a critical component of model development, yet notoriously challenging in ice sheet modeling. Here we view ice sheet models as part of earth system models, which allows us to evaluate how such an ice sheet system model responds to a given forcing. We show that hindcasting, i.e. forcing a model with known or closely-estimated inputs for past events to see how well the output matches observations, is a viable method of assessing model performance. By simulating the recent past of Greenland, and comparing to observations of ice thickness, ice discharge, surface speeds, mass loss and surface elevation changes for validation, we find that the short term model response is strongly influenced by the initial state. We show that the thermo-dynamical state can be misrepresented despite a good agreement with some observations, stressing the importance of using multiple observations. In particular we identify rates of change of spatially-rich observations as preferred validation metrics. Hindcasting enables a qualitative assessment of model performance relative to observed rates of change. It thereby reduces the number of admissible initial states more rigorously than validation efforts that do not take advantage of observed rates of change.

end suggested new abstract

References

Applegate, P. J., Kirchner, N., Stone, E. J., Keller, K., and Greve, R.: An assessment of key model parametric uncertainties in projections of Greenland Ice Sheet behavior, The Cryosphere, 6, 589–606, doi:10.5194/tc-6-589-2012, URL http://www.the-cryosphere.net/6/589/2012/, 2012.

Arthern, R. J. and Gudmundsson, G. H.: Initialization of ice-sheet forecasts viewed as an inverse Robin problem, J. Glaciol., 56, 527–533, 2010.

Arthern, R. J. and Hindmarsh, R. C. A.: Determining the contribution of Antarctica to sea-level rise using data assimilation methods., Phil. Trans. R. Soc. A, 364, 1841–65, doi:10.1098/rsta.2006.1801, 2006.

Bueler, E., Lingle, C. S., Kallen-Brown, J. A., Covey, D. N., and Bowman, L. N.: Exact solutions and verification of numerical models for isothermal ice sheets, J. Glaciol., 51, 291–306, doi:10.3189/172756505781829449, URL http://openurl.ingenta.com/content/xref?genre=article&issn=0022-1430&volume=51&issue=173&spage=291, 2005.

Bueler, E., Brown, J., and Lingle, C. S.: Exact solutions to the thermomechanically coupled shallow-ice approximation: effective tools for verification, J. Glaciol., 53, 499–516, doi:10.3189/002214307783258396, 2007.

Burton, J. C., Amundson, J. M., Abbot, D. S., Boghosian, A., Cathles, L. M., Correa-Legisos, S., Darnell, K. N., Guttenberg, N., Holland, D. M., and MacAyeal, D. R.: Laboratory investigations of iceberg capsize dynamics, energy dissipation and tsunamigenesis, J. Geophys. Res., 117, F01 007, doi:10.1029/2011JF002055, URL http://www.agu.org/pubs/crossref/2012/2011JF002055.shtml, 2012.

Joughin, I., Smith, B., Howat, I. M., Scambos, T., and Moon, T.: Greenland flow variability from ice-sheet-wide velocity mapping, J. Glaciol., 56, 415–430, 2010.

Larour, E., Morlighem, M., Seroussi, H., Schiermeier, J., and Rignot, E.: Ice flow sensitivity to geothermal heat flux of Pine Island Glacier, Antarctica, J. Geophys. Res., 117, F04 023, doi:10.1029/2012JF002371, URL http://www.agu.org/pubs/crossref/2012/2012JF002371.shtml, 2012a.

Larour, E., Schiermeier, J., Rignot, E., Seroussi, H., Morlighem, M., and Paden, J.: Sensitivity Analysis of Pine Island Glacier ice flow using ISSM and DAKOTA., J. Geophys. Res., 117, doi:10.1029/2011JF002146, 2012b.

Leng, W., Ju, L., Gunzburger, M., and Price, S.: Manufactured solutions and the verification of three-dimensional Stokes ice-sheet models, The Cryosphere, 7, 19–29, doi:10.5194/tc-7-19-2013, URL http://www.the-cryosphere.net/7/19/2013/, 2013.

Luthcke, S. B., Sabaka, T. J., Loomis, B. D., Arendt, A. A., McCarthy, J. J., and Camp, J.: Antarctica, Greenland and Gulf of Alaska land ice evolution from an iterated GRACE global mascon solution, J. Glaciol., submitted, 1–38, 2012.

Moon, T., Joughin, I., Smith, B., and Howat, I. M.: 21st-Century Evolution of Greenland Outlet Glacier Velocities, Science, 693, 576–578, doi:10.1126/science.1219985, 2012.

Roache, P.: Verification and Validation in Computational Science and Engineering, Hermosa Publishers, Albuquerque, New Mexico, 1998.

Robison, R. A. V., Huppert, H. E., and Worster, M. G.: Dynamics of viscous grounding lines, J. Fluid Mech., 648, 363, doi:10.1017/S0022112009993119, URL http://www.journals.cambridge.org/abstract_S0022112009993119, 2010.

Rogozhina, I., Martinec, Z., Hagedoorn, J. M., Thomas, M., and Fleming, K.: On the long-term memory of the Greenland Ice Sheet, J. Geophys. Res., 116, 1–16, doi:10.1029/2010JF001787, 2011.

Steinschneider, S., Polebitski, A., Brown, C., and Letcher, B. H.: Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change, Water Resources Research, 48, 1–16, doi:10.1029/2011WR011318, URL http://www.agu.org/pubs/crossref/2012/2011WR011318.shtml, 2012.

Stone, E. J., Lunt, D. J., Rutt, I. C., and Hanna, E.: Investigating the sensitivity of numerical model simulations of the modern state of the Greenland ice-sheet and its future response to climate change, The Cryosphere, 4, 397–417, doi:10.5194/tc-4-397-2010, 2010.

Vaughan, D. G. and Arthern, R. J.: Why Is It Hard to Predict the Future of Ice Sheets?, Science, 315, 1503–1504, doi:10.1126/science.1141111, 2007.

Wikipedia: Boeing 787 Dreamliner — Wikipedia, The Free Encyclopedia, URL http://en.wikipedia.org/wiki/787_Dreamliner, [Online; accessed 24 February 2013], 2013.