# Predicting calving events in Antarctica using Machine Learning

Anonymous Full Paper
Submission 17

## Abstract

Monitoring the calving dynamics of the Antarctic ice shelves is central to understanding a major driver for the changes to ocean levels on our planet. Several physical models have been proposed as calving laws, with varying predictive power. We propose an approach using Machine Learning (ML) to identify key variables and parameters that may be used in future models of the ice shelf calving dynamics. As part of an ongoing project, we have trained a U-Net on samples from a set of Gaussian Random Field (GRF)-represented Essential Climate Variables (ECVs). Ablation studies establish a few of the selected variables as having high correlation with calving events, with an F1 score above 0.9. Our first study site is the Larsen C Ice Shelf, on the northwest part of the Weddel Sea, where in 2017 there was a massive calving event. We have found strong correlations between the changes in ice-velocity leading up to this event, which are further improved when accounting for basal melt rates in the area.

## 1 Introduction

The Antarctic continent is covered by a sheet of ice, known as the Antarctic Ice Sheet (AIS). Surrounding the outlets of the AIS are a series of ice shelves which are crucial indicators of the response of Antarctica to a changing climate, as well as being a major buttressing factor securing the grounded ice on the Antarctic [1, 2]. Collapsing ice shelves, where the whole shelf rapidly disappears, have a major effect on changing both the dynamics of the grounded ice as well as contributing to multitudes of downstream effects [3]. Predicting large scale events remains elusive to physics-based, process models, including the prediction of calving events. As part of a project, we have used a combined approach with a data cube of monthly GRF representations and machine learning to predict calving events, with the aim of providing a data-driven approach to forecasting ice shelf calving.

Glaciers and ice shelves have been a field of study since the latter part of the 18th century: from the sliding dynamics of alpine glaciers under the force of gravity to looking at mechanical tensor descriptions of strain-fields, guided primarily by the search for physical laws determining glacier and ice sheet dynamics, and using numerical models to simulate the physical models since the 1950s and onwards [4]. Many have proposed physical laws to describe the process of calving [5], but, as noted by Wilner et al. [6], a majority of the proposed physical laws have remained unvalidated in the Antarctic.

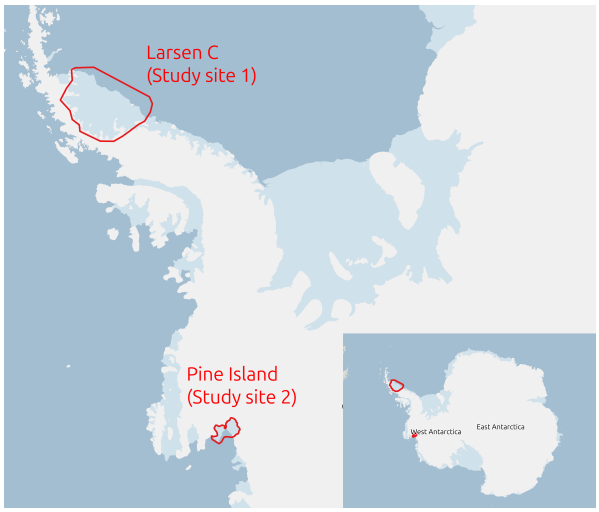


**Figure 1.** Overview of the two ice shelves selected as study sites; Larsen C and Pine Island.

Since the beginning of the 21st century, we have seen massive shedding of ice from the Antarctic ice shelves. Findings from the IMBIE assessment [7] report this shedding to be around 115 gigatonne per annum (Gt/a) in the period 2017-2020, with a peak discharge rate of 150 Gt/a in the prior 4-year period.

Following the standard definitions for general dynamical systems [8–10] we consider the physical system of ice dynamics as a tuple $(\chi, \tau, f)$, where $\chi$ is the state space of all possible states, $\tau$ is a set of times, and $f$ is the system's dynamical map:

$$f : \chi \times \tau \to \chi \tag{1}$$

Fixing time, $t$, we can observe $f$ as a time-dependent map:

$$f_t : \chi \to \chi,$$

which maps the transitions of states. Since the nature of our available data was historical, the above time-dependent map is assumed as a single path through the dynamical system. This is stated as our

set of samples:

$$\{f_t(x_i)|t \in [0,\ldots,T]\} \subset \chi, \tag{2}$$

which is a subset of the set of all allowed states, $\chi$, containing only the historical record of visited states. Another modification of assumptions was to account for stochasticity in our model of the underlying dynamical system, where the dynamical map is replaced with a tuple, $(\mathcal{C}, \tau, \Gamma, p, \mathcal{A})$, replacing the initial system states with the analogous system configurations, $\mathcal{C}$, which allows for mixed states, and consisting of a probabilistic map, $p$,

$$p : \chi \times \tau \to [0,1] \subset \mathbb{R},$$

a stochastic transition map, $\Gamma$,

$$\Gamma : \chi^2 \times \tau \to [0,1] \subset \mathbb{R},$$

where $\Gamma_{ij}(t) = p(x_i \times t | x_j \times 0)$ is the map of conditional probabilities for the system to transition into configuration $x_i$, at time $t$, from configuration $x_j$, at time 0, following the definition of conditional probabilities. Finally we also formally require an algebra of random variables[1], $\mathcal{A}$, which denotes a commutative algebra of maps for each random variable $A \in \mathcal{A}$

$$A : \chi \times \tau \to \mathbb{R},$$

in which the values $a_{x_i}(t) := A(x_i, t)$ tracks the magnitudes of the random variables.

In the mixed assumption of both historical and probable future events, we have used ML to predict the probability of some future state, $p_i(t')$, given information about the state $f_t(j)$, and propose that what results is an approximation [11] of the transition map, $\Gamma_{ij}(t)$. What was therefore implicitly approximated was the underlying dynamical system, as a stochastic process, by an approximation that assigns probabilities to future states of the actual system. Both the general dynamical system and the stochastic system were here assumed to be models which contained causality. Specifically we expected our data cube to contain approximate, or noisy, samples from real, continuous causal chains.

This project has focused on the prediction of anomalous or extreme dynamics on the AIS, combining the use of a GRF-resampled data cube, and an ML model for approximating the transitions from slices of time to future states of known calving events on the Larsen C site (see figure 1).

Deep learning methods are especially developed to identify and leverage correlations between the input data and the output labels. Predictive models make a claim that the correlations have a causal relationship. To make such claims we must ensure at least the following: firstly, that the events or configurations are linked with a temporal offset so that one event can be the cause of another. Secondly, there must be a medium that permits the effects of the first event to propagate to the second, within that temporal offset. Finally, it should be verified that the proposed causal chain of events is rational, i.e. that we can map the proposed causal chain to an applicable system of dynamics. This final step can be as simple as an expert verification, where a suitably knowledgeable person considers whether the proposed link can be explained based on our current understanding.

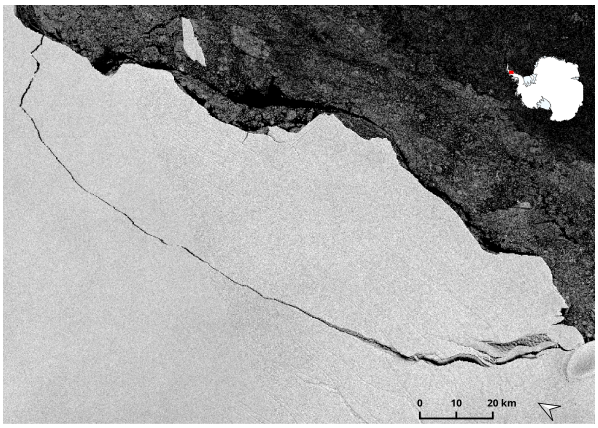


**Figure 2.** Sentinel-1 (SAR) image from 12/07/2017 of the A-68 iceberg calving event.

Under this paradigm, we can use ML to train a suitable neural network to identify correlations with future events across a cross-section of different input-variables. The Global Climate Observing System (GCOS) has assigned ECVs that pertain to AIS. These ECVs constitute a list of factors that are essential for modelling and understanding the climate. Of these ECVs, GCOS has isolated a subset that pertain to the cryosphere in general, and to Antarctica specifically. In this project we have used an enriched set of inputs in addition to the ECVs that are commonly ascribed to ice shelves, but we will in this paper refer to all of the inputs as ECVs.

It's important to note that the ECVs are assumed to have some level of correlation between them, which we implicitly use in our predictions. Such correlation would imply shared causality, which essentially means there is a pullback from the data-correlations to a common cause. We also make use of ablation studies to isolate individual ECVs, and identify whether these variables have strong correlations to calving events. Isolated variables in the ablations are then also explored with eXplainable AI (XAI) to highlight the saliency

[1]This terminology simply states that our measurable variables, $a_{x_i}(t)$, when the system is in a given state or configuration, $x_i$, are to some extent not completely deterministic.

of the inputs, as a further way to estimate the validity of the model predictions.

At the onset of this project a literature review was performed to find if similar approaches had been taken, but no such case could be found. What is presented here is therefore "Terra Nova", where the findings of this project may help guide future endeavours towards observations of the Antarctic.

# 2 Method

Our approach was structured in four stages;

- **Data Collection** Data was collected and reprocessed into a GRF representation, and associated with known calving events as an (`input`, `output`) tuple.
- **U-Net training** A U-Net was trained to correlate the different slices of the GRF-data cube with calving events within a moving window of future events.
- **Significance analysis** Ablation studies has been applied to identify the most significant contributions from the GRF-datacube, establishing the performance-metric correlation of each input. XAI will be used at a later stage to highlight an input-saliency map.
- **Validation & Verification** Results of the ablation studies have then verified by domain experts to ascertain the scientific validity of the predictions. Hold out validation data is used to track model performance (see figure 3).

## 2.1 Data Cube

A multivariate datacube was constructed comprising various datasets relevant to ice sheet physics selected to cover the years 2014-2022 inclusive. Our first study-site was the Larsen C Ice Shelf, figure 1, which had a large calving event in July 2017, where the approximately 1000 gigatonne (Gt) A-68 iceberg calved from the main ice shelf.

The datasets listed in table 1 have been fitted and resampled to a GRF representation to standardise spatial resolution and fill data gaps. The GRF was approximated at a set of irregular vertices on a mesh, before it was regridded as a contiguous surface of equal squares, i.e. as digital images. Samples taken from GRF have some process-based uncertainty, which was also produced as a gridded, spatially distributed output, and which will be quantified in the further analysis of the data cube, but which we for the purposes of our predictions have not taken into account.

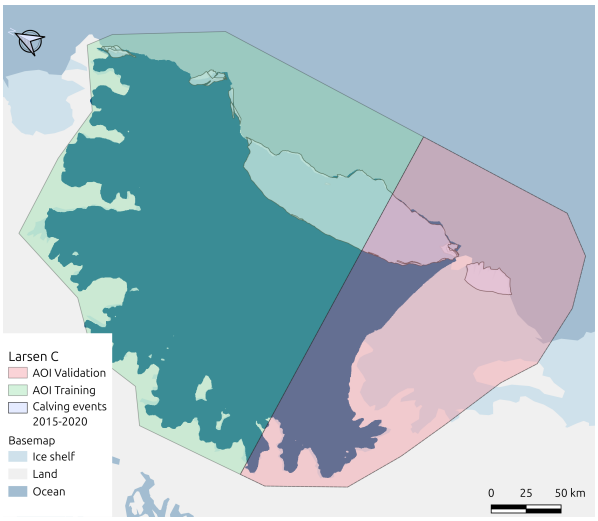


**Figure 3.** 2D surface map of our data. Showing data coverage (dark blue), training data (green) vs validation data (red), and calving events within 2015-2020 (bright areas). Calving events are here reproduced as vectorised data from Qi et al. [12], without spatial alignment and rasterisation.

**Table 1.** Table of datasets contained in the used version of Data Cube. All data is resampled from original resolution to 200 m resolution. The complete data cube will be published on a later date.

| Dataset (source) | Original Resolution |
|---|---|
| Ice Velocity [13] | $200 \times 200$ m |
| Surface mass balance [14] | $27 \times 27$ km |
| Firn thickness [15] | $27 \times 27$ km |
| Firn air content [15] | $27 \times 27$ km |
| Basal melt rate [16] | $1000 \times 1000$ m |
| Wind speed and direction [17] | $31 \times 31$ km |

From this resampled datacube we then sampled our input data, and paired the samples to the corresponding targets from a 15 year calving dataset by Qi et al. [12]. This calving dataset was rasterised to the same resolution as our input data, and assigned values with an indicator function over the majority class in the corresponding vector data,

$$\mathbf{1}_{calving}(t) = \begin{cases} 1\text{- if calving at } t \\ 0\text{- if } \underline{\text{not}} \text{ calving at } t \end{cases} \quad (3)$$

To establish a future-prediction from these data, the time at which the labels are sampled is offset compared to the slice of the data cube. This was determined using a sliding window of data that was set to 3 months, e.g. to validate prediction in July 2017 we provided data samples from April, May, and June of 2017 as our 3-month lead-time prediction[2]. Testing, and figures shown here (figures figs. 4 and 5) are from the month before the validation data, i.e. for predictions of an event

[2] ditto for 6 and 9 month lead-time prediction.

in July we test with samples from March. This ensures the data used for prediction is at least 3 months before any recorded event in July. This was also done to ensure against overfitting on data that have relatively low resolution (see 1).

The data and labels were combined and gridded with overlapping grids to provide the (`input`, `label`) pairs which were randomly sampled by the models during training. All data was sampled at 200m per pixel, which translates to an XY-bounding box in the spatial domain of $1551 \times 1651$ pixels to cover the first study-site of Larsen C.

All the data were split into `Training` or `Validation` by intersection of a vectorised Area Of Interest (AOI) (see figure 3), where the major calving event of A-68 is roughly intersected to reserve $\frac{1}{3}$ of its total area for validation. The splitting of the ice-shelf remains the same throughout the time-dimension, to ensure no potential learning of the precursive configurations, or memorisation of the outlines carries over from training to validation data. The remaining ice-shelf is similarly split using vectorised areas, and masking areas that do not intersect with the target AOI

## 2.2 Model and Parameters

Since we wanted to inspect the resulting predictions from our network, we chose to use a U-Net [18] model. These have a long standing, empirical track-record classifying remote sensing data, and consist of encoder, decoder, and a bottleneck, with skip-connections to act as control-parameters for the reconstructed image of delineated features in the output map. Furthermore, as shown by Tai et al. [19], U-Nets are effectively solving a control problem. In our case, this means they approximate the transition maps inherent to our causal model, decoded as a probability-distribution over calving events. Our U-Net was trained using an AdamW optimiser [20], and following recent recommendations from the study on hyperparamters by Orvieto and Gower [21] we set $\beta_1 = \beta_2 = 0.9$

We have used Focal Loss [22] as our target for classification, with $\gamma = 3$, to account for some of the class imbalance. Sampling was performed from a random grid, and our target metric was the macro-averaged F1-score, which should avoid biasing the metric towards the `no-calving` label. Each sample also got a randomly applied augmentation, and was rotated within $\pm 180°$, and had a 50% chance of being flipped along each of the horizontal and vertical axes. The augmentations ensure a better generalisation of the model, and artificially increase the number of samples made available to the model.

## 2.3 Ablation & XAI

We have performed a set of ablations over the available dataset, focused on identifying the strength of the correlations between the available input data from the data cube and the labels from Qi et al. [12]. To do this we sliced the data cube into its constitute parts, and trained U-Nets for each variable in turn, under similar conditions. The ablations were performed across a selection of lead-times, divided into sections of 3-month periods, and predictions evaluated on the final time-slice that was not included in the training data. This was done to ensure the data was representative of the region trained on, but not included in the training data.

Additionally, we performed an experiment using the two most predictive subsets of the data: ice velocity and basal melt (see figure 4). It should be noted that the basal melt data is partially dependent on the ice velocity data. Combination of ablation and XAI is used to identify key input variables. To establish the significance of the different layers in the data cube, we divided the dataset. For the ablations we start with the individual ECVs, and measure the performance of them per se. Consecutive experiments look at the combination of variables within $\mathcal{P}(\text{ECV})$.

Now that we identified the contribution of each ECV to the final performance, we will use guided backprop, and a variation of GradCAM [23] to highlight the area of the input that is most salient for the predicted output. Verifying the areas the model finds salient acts as a constraint on the applicable mathematics that goes into considering whether the model seems to cover existing theory, or if it may be completely confused.

## 2.4 Validation

Predictions for each of the ablation experiments have been associated with their metric performance. Data and predictions was then prepared for the entire ice shelf, focusing primarily on the validation area of Larsen C. The quality of the predictions was then validated by domain experts from Centre of Excellence in Environmental Data Science (CEEDS) at Lancaster University (LU), who made a qualitiative verification of whether the causal relationship fits with our current understanding of glaciology. This form of validation was considered a more appropriate approach to the standard train-test split of data primarily due to the limited amount of calving events observed from Larsen C during the period of available data. Secondly, since a stated goal of the

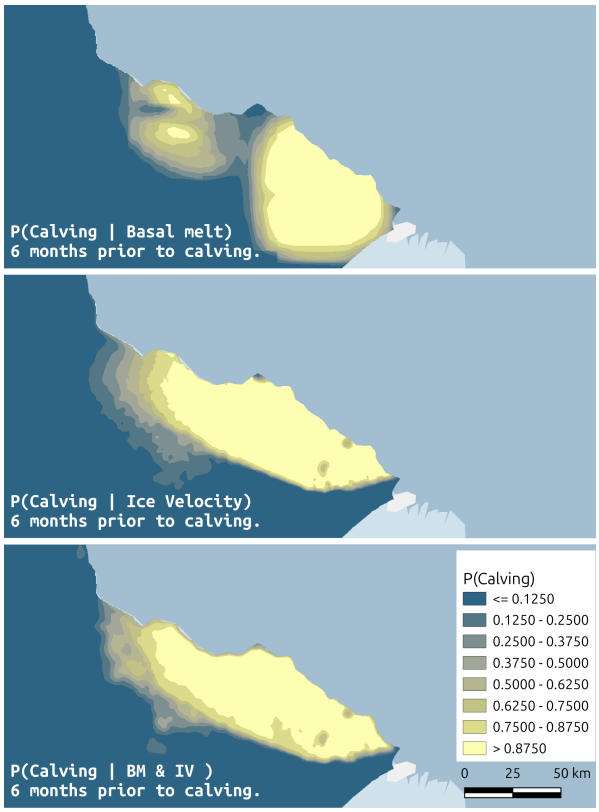


**Figure 4.** Predicted correlations of the three most performant subsets of data and the calving of the A-68 iceberg. Here we see predictions from a model trained on basal melt, one trained on ice velocity, and a model trained on the combined data of both basal melt and ice velocity.

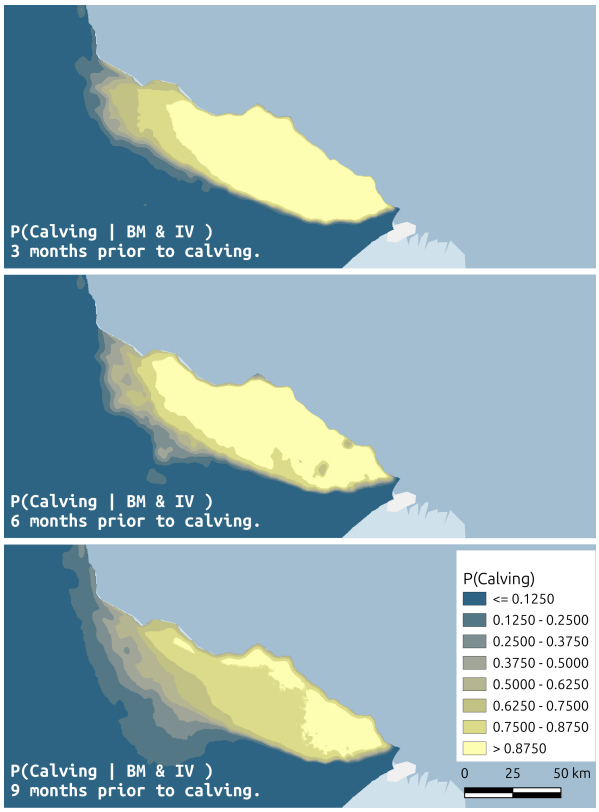


**Figure 5.** When the model is predicting events further into the future, we see a gradual increase in the uncertainty of the label. It is here clear that a 3 month lead time leads to a clear delineation of the calving when compared to predictions from 6 months or 9 months before.

project was to identify clear links between the data and the predictions of our ML approach, the validation of these causal predictions require some form of human-expert-in-the-loop. This provides necessary feedback to ensure that our models are more likely to rely on real physical relationships rather than data correlations that have no connection to the actual dynamics of the system. Finally, in the continuation of the project, we will be providing a GradCAM [23]-based XAI saliency map over the multivariate data cube. These are intended to inform domain experts about the relative weight the ML model places on the spatially distributed information of each channel. The validation of these also require familiarity with the inputs given to the model.

# 3 Results

## 3.1 Ablation

Table 2 lists the F1-scores of the different datasets with their respective lead times. We can here see that Ice Velocity (IV) both with and without the Basal Melt (BM) have strong correlations to the target labels. Note also that firn air content does not seem to have any discernable correlation to the calving events. Firn thickness has a slightly higher correlation as judged by the maximal F1 score, but this seems to be an artifact when looking at the metric through time 6.
In figure 6, we can see the early effect of including basal melt. The effect of including basal melt also seems to help models training with ice velocity to converge more rapidly, and at longer time scales surpass the performance of models only trained on ice velocity data.

Wind speed seems to have a higher than expected correlation, and we see the models trained on wind speed pick up more towards the later stages of training. However these models also seem to fluctuate much, and when predicting on the test-data the cor-

relation seems to disappear. For the maps presented to expert validation, the model predictions did not contain discernable or significant information. This may therefore indicate a case of overfitting, where the model associates the values of the wind speed directly to the label rather than generalising, a hypothesis further supported when considering the original resolution of the wind speed data (table 1).

## 3.2 Validation

Predictions from ablated models have been presented to domain experts from the polar science group at LU associated with the CEEDS and Centre for Polar Observation and Modelling (CPOM). Model performance is assessed with reference to experience with observed calving events and understanding of the physics behind the calving process [6, 24].

The results on the variables with most predictive power are consistent with an understanding of the physics of ice shelf calving. Ice velocity, the most significant predictor, has a direct relationship with calving; for floating ice closer to the calving front and further from the grounding line, high variability in velocity over short distances suggests it is associated with rifting, cracking, or other forms of disturbance to otherwise smooth ice flow which are significant in the process of icebergs detaching from the shelf.

Basal melt as the second most significant predictor has a major impact on ice shelf thickness, with significant thinning of the shelf making calving easier. It is not surprising that the other variables used do not appear to have major predictive power. They have a less direct impact on the calving process, so despite likely having some information to impart about ice shelf stability it is not expected that any of these datasets individually allows for confident prediction of imminent calving. Surface mass balance is not the dominant factor in mass changes for Antarctic ice shelves, with basal melt and calving much larger sources of mass loss. Changes in firn thickness and firn air content are also indications of surface processes and are not expected to have a large impact on overall flow. Wind speed may have some direct and indirect impact on ice shelf calving, but the coarse resolution of the input dataset and the fact that wind speed variations aren't likely to be localised to areas prone to calving explains the relative lack of predictive utility.

The spatial predictions of calving location (figure 5) matches well with the geometry of the A-68 calving event. The closer contours in the southern part of the domain (right of figure 5), suggesting higher confidence in the location, are consistent with the way the calving event played out, with the separation beginning in this area and propagating north (left of figure 5).

**Table 2.** Highest validation F1-score vs lead time. Overall highest F1-scores per lead time highlighted in **bold.**

| Subset \ Lead time | 3 months | 6 months | 9 months |
|---|---|---|---|
| IV | **0.938** | 0.936 | 0.919 |
| BM | 0.843 | 0.793 | 0.715 |
| IV + BM | **0.938** | **0.938** | **0.927** |
| Wind Speed (WS) | 0.780 | 0.739 | 0.753 |
| Surface Mass Balance (SMB) | 0.687 | 0.490 | 0.492 |
| firn air content | 0.482 | 0.471 | 0.477 |
| firn air thickness | 0.483 | 0.669 | 0.749 |

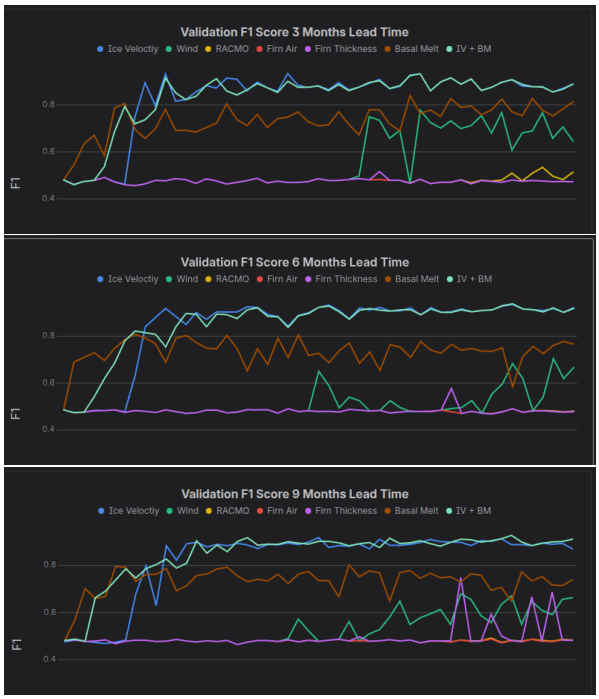


**Figure 6.** Validation F1 scores for the different ablations and lead times. Note that though the ice velocity converges at a higher level $\approx 0.93$ the basal melt contributes more in the earlier steps. The combination of ice velocity and basal melt reaches the highest overall peak at $\approx 0.938$, performing at parity or better than the ice velocity alone.

# 4 Conclusion

In this work we show that a U-Net can be used for predicting future events, acting as an approximation of the temporal transition map. While these predictions are now at a stage that shows clear correlations, it still remains to move this work into a more dynamic study site, namely Pine Island. It should also be noted that the correlations are currently statistical correlations, and not founded strongly in physical theory. A further review of proposed physical models for the calving dynamics

of the AIS can be made on the basis of these correlational models, where clear correlations such as the ice velocity are more closely considered.

We have found that our models can train to a high degree of correlation for long-reaching causal chains, as measured by the F1 scores for lead-time data. The question of generalisability remains open, with the model currently only being validated for a relatively stable ice sheet. Further modifications to the experiment design may provide more insight, and we are considering a change from the lead-time approach into a windowed gap-based approach.

Confirmation that the model performance remains consistent with current understandings of the physics of ice shelf calving is promising. This indicates that our central hypothesis of approximating dynamical systems remains unfalsified, and that we in the continuation of the work may be able to approximate the more active dynamics of Pine Island.

## 5 Future Work

Currently data is being collected for a second study site over Pine Island, which is a more dynamic ice shelf than Larsen C. We will here work with a similar GRF resampled data cube of ECVs. For the work on Pine Island we expect to perform similar ablation studies, and we will also investigate the transferability of the current models, as well as the inverse, i.e. training new models on Pine Island, and then transferring them back to Larsen C to compare with our current findings.

We expect that our currently trained models will be transferable, and we will seek to answer questions regarding performance vs generalisation on the domain of calving prediction.

## References

[1] T. K. Dupont and R. B. Alley. "Assessment of the Importance of Ice-Shelf Buttressing to Ice-Sheet Flow". In: *Geophysical Research Letters* 32.4 (Feb. 2005). ISSN: 0094-8276. DOI: 10.1029/2004GL022024.

[2] G. H. Gudmundsson. "Ice-Shelf Buttressing and the Stability of Marine Ice Sheets". In: *The Cryosphere* 7.2 (Apr. 2013), pp. 647–655. ISSN: 1994-0416. DOI: 10.5194/tc-7-647-2013.

[3] E. A. Hill, G. H. Gudmundsson, and D. M. Chandler. "Ocean Warming as a Trigger for Irreversible Retreat of the Antarctic Ice Sheet". In: *Nature Climate Change* 14.11 (Nov. 2024), pp. 1165–1171. ISSN: 1758-6798. DOI: 10.1038/s41558-024-02134-8.

[4] H. Blatter, R. Greve, and A. Abe-Ouchi. "A Short History of the Thermomechanical Theory and Modeling of Glaciers and Ice Sheets". In: *Journal of Glaciology* 56.200 (Jan. 2010), pp. 1087–1094. ISSN: 0022-1430, 1727-5652. DOI: 10.3189/002214311796406059.

[5] D. I. Benn, C. R. Warren, and R. H. Mottram. "Calving Processes and the Dynamics of Calving Glaciers". In: *Earth-Science Reviews* 82.3 (June 2007), pp. 143–179. ISSN: 0012-8252. DOI: 10.1016/j.earscirev.2007.02.002.

[6] J. A. Wilner, M. Morlighem, and G. Cheng. "Evaluation of Four Calving Laws for Antarctic Ice Shelves". In: *The Cryosphere* 17.11 (Nov. 2023), pp. 4889–4901. ISSN: 1994-0416. DOI: 10.5194/tc-17-4889-2023.

[7] I. N. Otosaka, A. Shepherd, E. R. Ivins, N.-J. Schlegel, C. Amory, M. R. van den Broeke, M. Horwath, I. Joughin, M. D. King, G. Krinner, S. Nowicki, A. J. Payne, E. Rignot, T. Scambos, K. M. Simon, B. E. Smith, L. S. Sørensen, I. Velicogna, P. L. Whitehouse, G. A, C. Agosta, A. P. Ahlstrøm, A. Blazquez, W. Colgan, M. E. Engdahl, X. Fettweis, R. Forsberg, H. Gallée, A. Gardner, L. Gilbert, N. Gourmelen, A. Groh, B. C. Gunter, C. Harig, V. Helm, S. A. Khan, C. Kittel, H. Konrad, P. L. Langen, B. S. Lecavalier, C.-C. Liang, B. D. Loomis, M. McMillan, D. Melini, S. H. Mernild, R. Mottram, J. Mouginot, J. Nilsson, B. Noël, M. E. Pattle, W. R. Peltier, N. Pie, M. Roca, I. Sasgen, H. V. Save, K.-W. Seo, B. Scheuchl, E. J. O. Schrama, L. Schröder, S. B. Simonsen, T. Slater, G. Spada, T. C. Sutterley, B. D. Vishwakarma, J. M. van Wessem, D. Wiese, W. van der Wal, and B. Wouters. "Mass Balance of the Greenland and Antarctic Ice Sheets from 1992 to 2020". In: *Earth System Science Data* 15.4 (Apr. 2023), pp. 1597–1616. ISSN: 1866-3508. DOI: 10.5194/essd-15-1597-2023.

[8] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Encyclopedia of Mathematics and Its Applications. Cambridge: Cambridge University Press, 1995. ISBN: 978-0-521-34187-5. DOI: 10.1017/CBO9780511809187.

[9] R. L. Devaney. *An Introduction to Chaotic Dynamical Systems*. 3rd ed. Boca Raton: Chapman and Hall/CRC, Oct. 2021. ISBN: 978-0-429-28080-1. DOI: 10.1201/9780429280801.

[10] J. A. Barandes. *The Stochastic-Quantum Theorem*. Sept. 2023. DOI: 10.48550/arXiv.2309.03085. arXiv: 2309.03085 [quant-ph].

[11] K. Hornik, M. Stinchcombe, and H. White. "Multilayer Feedforward Networks Are Universal Approximators". In: *Neural Networks* 2.5 (Jan. 1989), pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8.

[12] M. Qi, Y. Liu, J. Liu, X. Cheng, Y. Lin, Q. Feng, Q. Shen, and Z. Yu. "A 15-Year Circum-Antarctic Iceberg Calving Dataset Derived from Continuous Satellite Observations". In: *Earth System Science Data* 13.9 (2021), pp. 4583–4601. DOI: 10.5194/essd-13-4583-2021.

[13] ENVEO, J. Wuite, M. Hetzenecker, T. Nagler, and S. Scheiblauer. *ESA Antarctic Ice Sheet Climate Change Initiative (Antarctic_Ice_Sheet_cci): Antarctic Ice Sheet Monthly Velocity from 2017 to 2020, Derived from Sentinel-1, V1*. 2021. DOI: 10.5285/00FE090EFC58446E8980992A617F632F.

[14] C. T. Van Dalum, W. J. Van De Berg, S. N. Gadde, M. Van Tiggelen, T. Van Der Drift, E. Van Meijgaard, L. H. Van Ulft, and M. R. Van Den Broeke. "First Results of the Polar Regional Climate Model RACMO2.4". In: *The Cryosphere* 18.9 (Sept. 2024), pp. 4065–4088. ISSN: 1994-0424. DOI: 10.5194/tc-18-4065-2024.

[15] S. B. M. Veldhuijsen, W. J. Van De Berg, M. Brils, P. Kuipers Munneke, and M. R. Van Den Broeke. "Characteristics of the 1979–2020 Antarctic Firn Layer Simulated with IMAU-FDM v1.2A". In: *The Cryosphere* 17.4 (Apr. 2023), pp. 1675–1696. ISSN: 1994-0424. DOI: 10.5194/tc-17-1675-2023.

[16] B. Davison. *Data and Code for: "Annual Mass Budget of Antarctic Ice Shelves from 1997 to 2021"*. June 2023. DOI: 10.5281/ZENODO.8052518.

[17] Copernicus Climate Change Service. *ERA5 Monthly Averaged Data on Single Levels from 1940 to Present*. 2019. DOI: 10.24381/CDS.F17050D7.

[18] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[19] X.-C. Tai, H. Liu, R. H. Chan, and L. Li. *A Mathematical Explanation of UNet*. Oct. 2024. DOI: 10.48550/arXiv.2410.04434. arXiv: 2410.04434 [cs].

[20] I. Loshchilov and F. Hutter. *Decoupled Weight Decay Regularization*. Jan. 2019. DOI: 10.48550/arXiv.1711.05101. arXiv: 1711.05101 [cs].

[21] A. Orvieto and R. Gower. *In Search of Adam's Secret Sauce*. May 2025. DOI: 10.48550/arXiv.2505.21829. arXiv: 2505.21829 [cs].

[22] I. Newton, T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. "Focal Loss for Dense Object Detection". In: *Proceedings of the IEEE international conference on computer vision* 1 (2017), pp. 2980–2988. DOI: 10.1109/TPAMI.2018.2858826.

[23] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization". In: *International Journal of Computer Vision* 128.2 (Feb. 2020), pp. 336–359. ISSN: 0920-5691, 1573-1405. DOI: 10.1007/s11263-019-01228-7. arXiv: 1610.02391 [cs].

[24] D. I. Benn and J. A. Åström. "Calving Glaciers and Ice Shelves". In: *Advances in Physics: X* 3.1 (Jan. 2018), p. 1513819. ISSN: null. DOI: 10.1080/23746149.2018.1513819.

# A Appendix: Figures

## A.1 3 month predictions

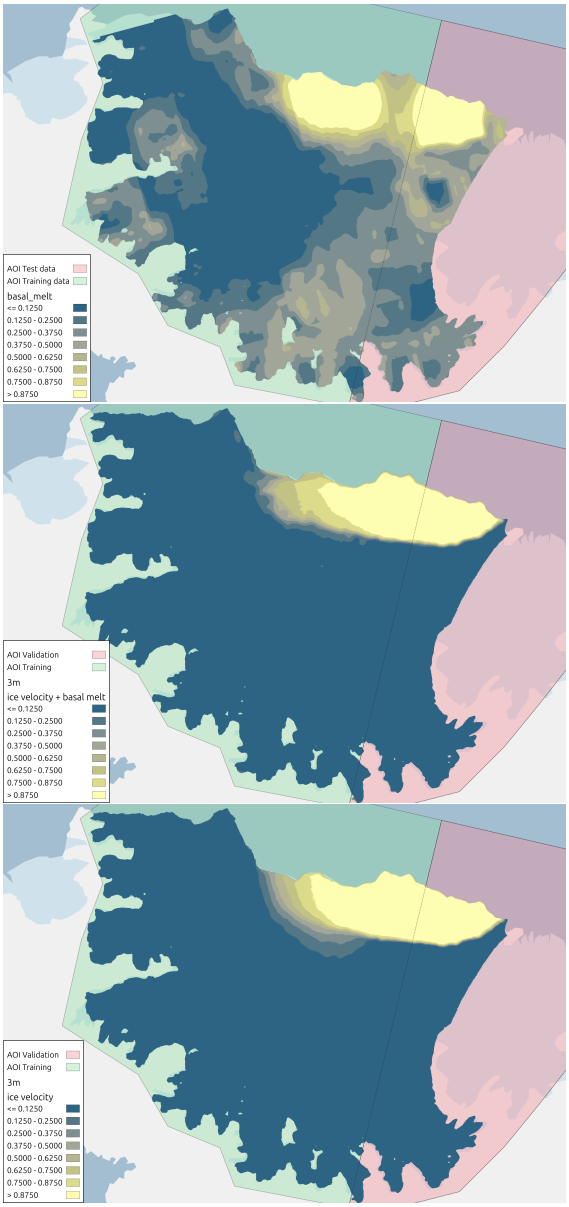


**Figure A.1.** 3 month; basal melt, ice velocity & basal melt, ice velocity.

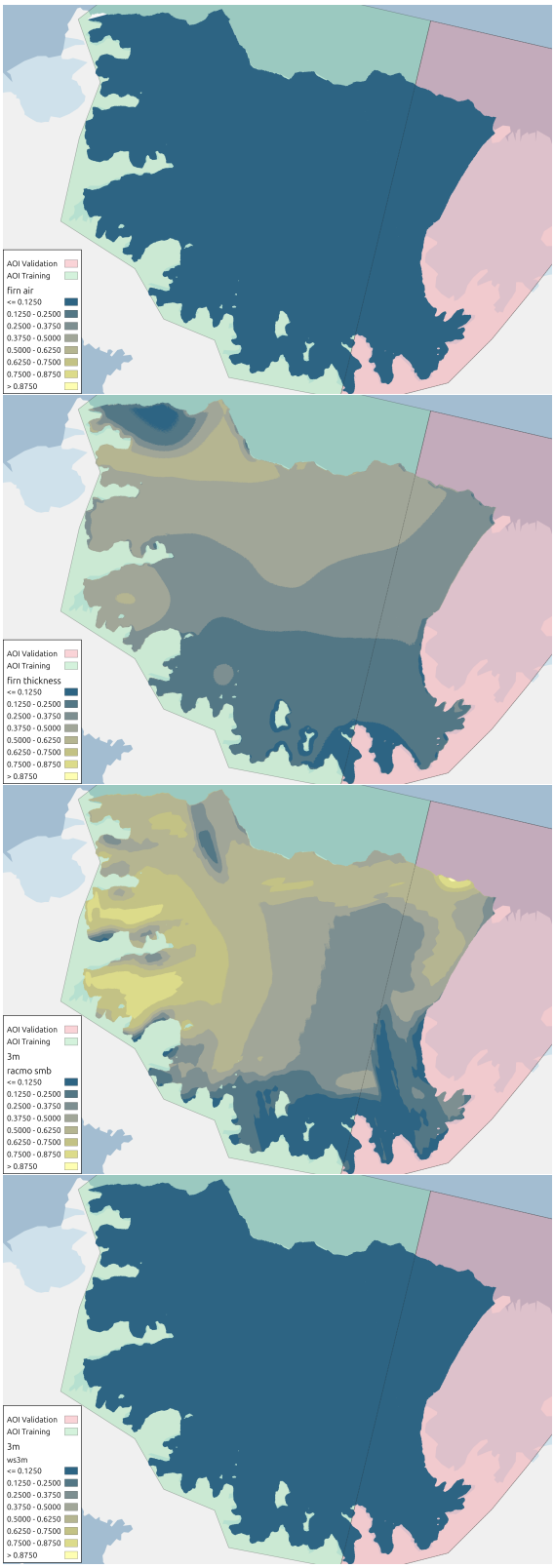


**Figure A.2.** 3 month; firn air, firn thickness, surface mass balance, wind speed

## A.2 6 month predictions

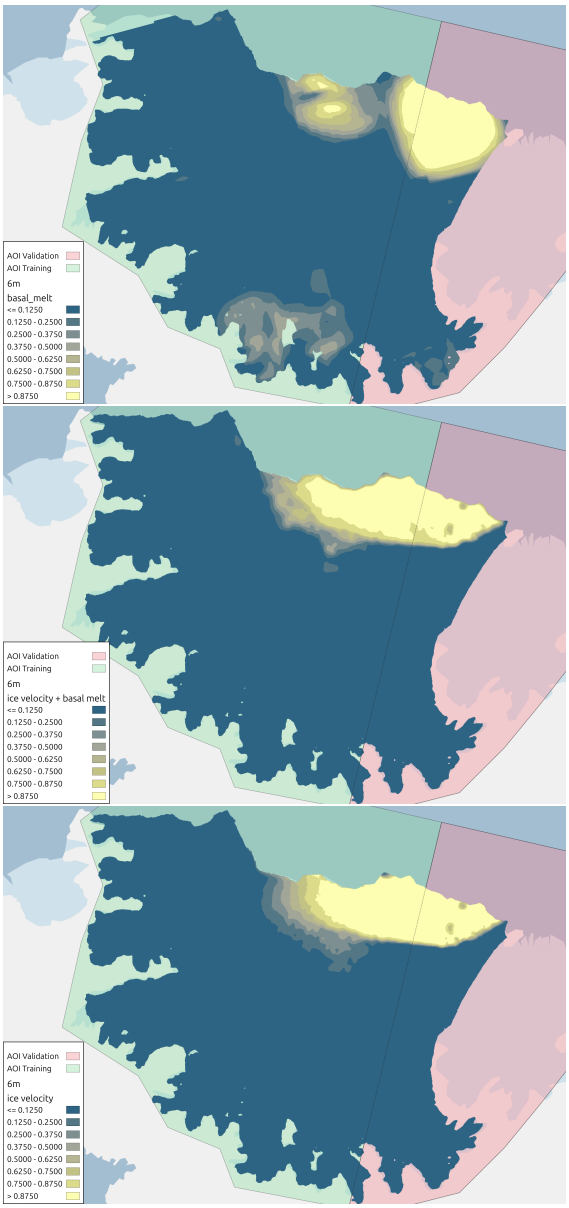


**Figure A.3.** 6 month; basal melt, ice velocity & basal melt, ice velocity.

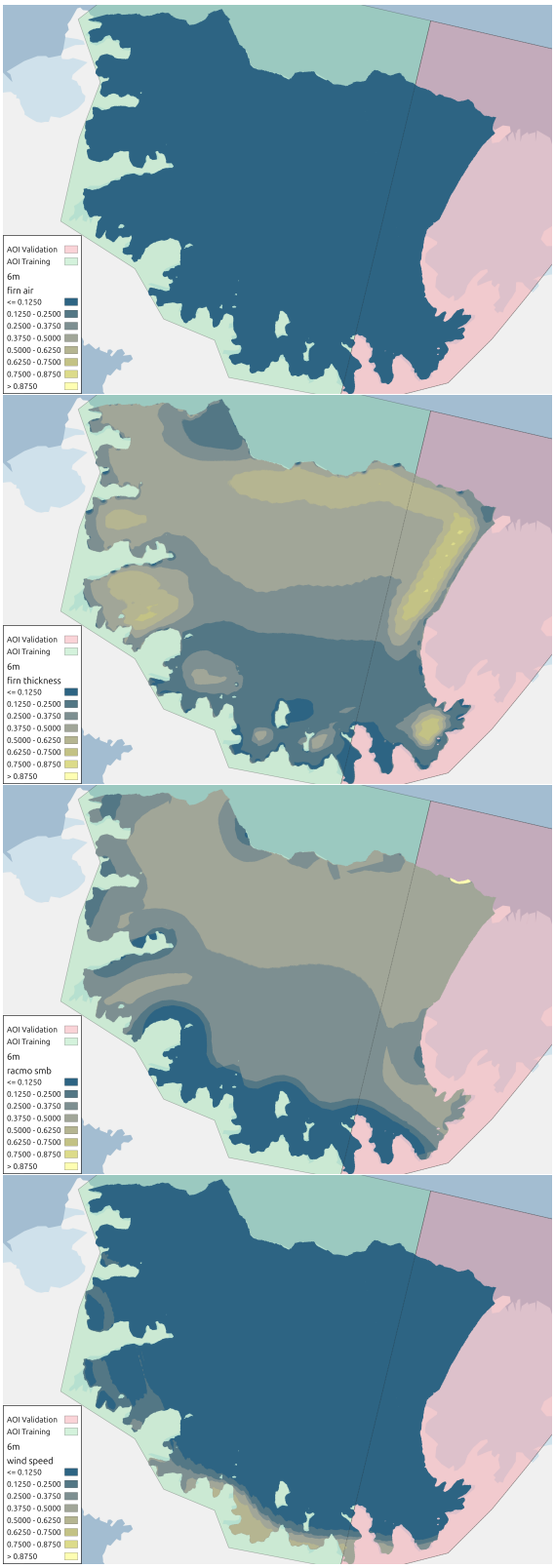


**Figure A.4.** 6 month; firn air, firn thickness, surface mass balance, wind speed

## A.3 9 month predictions

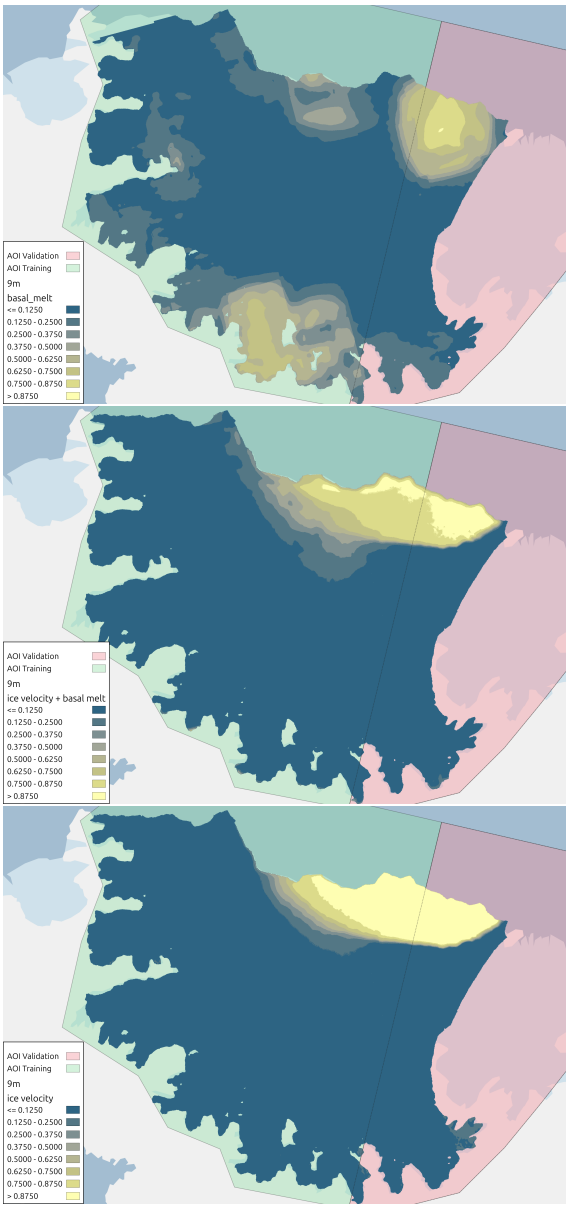


**Figure A.5.** 9 month; basal melt, ice velocity & basal melt, ice velocity.

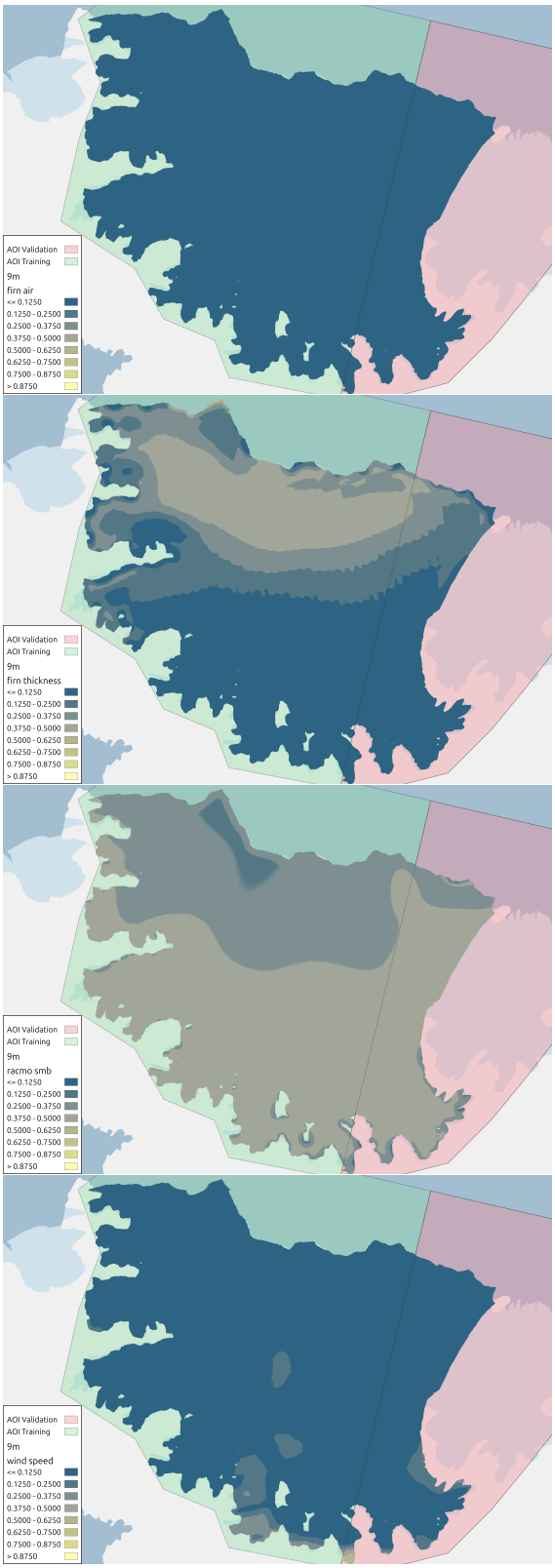


**Figure A.6.** 9 month; firn air, firn thickness, surface mass balance, wind speed