Glacier Movement Prediction with Attention-based Recurrent Neural Networks and Satellite Data

Jonas Müller, Raphael Braun, Hendrik PA Lensch, Nicole Ludwig
University of Tübingen
`nicole.ludwig@uni-tuebingen.de`

Abstract

Studying glacier movements is crucial because of their indications for global climate change and its effects on local land masses. Building on established methods for glacier movement prediction from Landsat-8 satellite imaging data, we develop an attention-based deep learning model for time series data prediction of glacier movements. In our approach, the Normalized Difference Snow Index is calculated from the Landsat-8 spectral reflectance bands for data of the Parvati Glacier (India) to quantify snow and ice in the scene images, which is then used for time series prediction. Based on this data, a newly developed Long-Short Term Memory Encoder-decoder neural network model is trained, incorporating a Multi-head Self Attention mechanism in the decoder. The model shows promising results, making the prediction of optical flow vectors from pure model predictions possible.

1 Introduction

Glaciers and their inherent dynamics are a well-documented indicator for global climate change [1, 31, 4, 9, 26, 24, 15, 23, 30, 5]. Together with ice sheets, they also cover 10% of the Earth's land area, holding about 69% of Earth's freshwater [11]. Their monitoring is crucial for local authorities, as glacier melting can have impacts on close-by regions like landslides, debris flows, rock slope failures or ice avalanches [26]. To study the impacts of glacial change, advanced remote sensing approaches have been developed, including, e.g., terrestrial laser scanning, radar satellite applications, unmanned aerial vehicles, automatic camera imaging, and *multispectral satellite imaging* [4].

Many models of glacier dynamics explicitly focus on the physical processes inside glaciers [8, 5]. This approach comes with the disadvantage of varying estimates of different models, uncertainties of different data sources, varying estimates of the same models by different research groups, and the need for explicit knowledge of the underlying processes for parameter estimation and data prediction. Therefore, deep learning methods have been deployed for glacier movement prediction from satellite data [16], where glacier dynamics are modelled indirectly through input-output relations. Recently, deep learning, and especially computer vision, has made large progress due to the ability to learn complex recurrent patterns, the emergence and improvement of advanced neural network architectures [27], as well as the use of higher computational resources in model training [3] and larger datasets [6]. Based on these advancements, the present study focuses on the application of recurrent deep-learning models to predict the movement of the Parvati glacier using satellite images.

Although recurrent neural networks (RNN) have been used to predict sea ice motion (e.g. Mu et al. [17], Zhai and Bitz [32], Petrou and Tian [19, 20]), most papers investigating glaciers focus on classification approaches (e.g. Chu et al. [7], Marochov et al. [13], Prieur et al. [22]) instead of predicting future glacier states with RNNs. Our approach is, therefore, inspired by sea ice motion prediction. We use scene patches, similar to Min et al. [16] and Petrou and Tian [19], apply a dense optical flow algorithm as Vonica et al. [28] and rely on an LSTM self-attention hybrid model inspired

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

by Mu et al. [17]. We compare our approach to an LSTM encoder-decoder structure, similar to Ali et al. [2], and a ConvLSTM model. To the best of our knowledge, we are the first to use an attention-based RNN for glacier movement prediction from satellite image time-series data. Therefore, this paper presents a unique opportunity to contribute to understanding large climate change detection indicators.

2 Data and Model

The generated datasets consist of Landsat-8 level 2 satellite images of the Parvati glacier accessed through the SpatioTemporal Asset Catalog (STAC) [1]. Because Landsat-8 imaging is not completely accurate [28], the images need to be aligned, for which we use the Enhanced Correlation Coefficient Maximization algorithm [10] to compensate for pixel shifts in images and produce more robust model predictions. We also filter the data for images within the glacier *scene* (77.4880°E, 31.9927°N, 77.8206°E, 32.2253°N) with less than 20% cloud coverage and less than 50% of missing pixel values and use two out of the 19 bands provided at 30m resolution to estimate the glacier movements: the green band (green, $0,53-0,59$ micrometers) and the shortwave infrared one band (swir1, $1.57-1.65$ micrometers). These bands still contain missing pixel values; thus, we use a 5x5 kernel to calculate a weighted average for the missing pixel in question [28]. We can then compute the Normalized Difference Snow Index NDSI $= (\text{green} - \text{swir1})/(\text{green} + \text{swir1})$, which quantifies snow or ice in $[-1, 1]$.

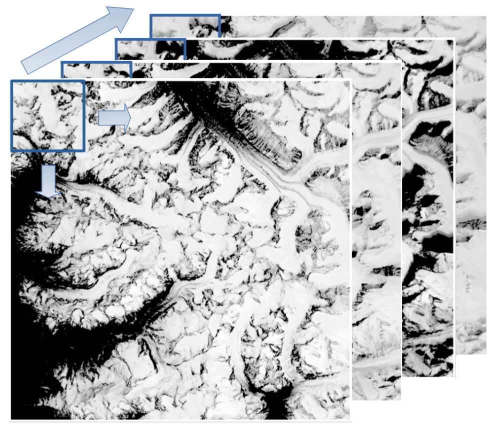

Figure 1. Glacier development over time. The processing patches are highlighted in blue.

This NDSI index is for example used to classify glaciers with a U-net architecture [12] with a multitude of different thresholds, while we quantify snow/ice in pixels with a generally adopted threshold of 0.3 [28], generating masks for each pixel. We use large regions, for the Parvati glacier to capture the glacier fully. The images are additionally averaged over three months because the intervals between images are not evenly spaced (e.g. clouds or missing data). The scene is sampled in patches of size 50×50 to decrease the computational load of the used models. To generate the final model input, patches of the same coordinates in different images are used over time, sampling the same region at different time points. For simplicity a sequence of n consecutive images of the scene at different time-points can be imagined, $S = \{\mathbf{S}_t \mid t \in T\}$, where $\mathbf{S}_t \in \mathbb{R}^{800\times800}$. Patches of overlapping or non-overlapping patches can be sampled from the scene, $P = \{\mathbf{p}_{k,t} \mid k \in K, t \in T\}$, where each $\mathbf{p}_{k,t}$ is $\in \mathbb{R}^{50\times50}$, $k \in K$ denotes the patch number and $t \in T$ the time index. We choose the n as 8, therefore taking 4 patches as input, $\{\mathbf{p}_{k,t},\ \mathbf{p}_{k,t+1},\ \mathbf{p}_{k,t+2},\ \mathbf{p}_{k,t+3}\}$, and predict the next 4 patches in the same coordinates, $\{\mathbf{p}_{k,t+4},\ \mathbf{p}_{k,t+5},\ \mathbf{p}_{k,t+6},\ \mathbf{p}_{k,t+7}\}$. Thus, the model input and target dimensions, ignoring the batch dimension, are tensors $\in \mathbb{R}^{4\times50\times50}$.

Our RNN-approach is a self-attention LSTM hybrid (SA-LSTM-H) model (as shown in Figure 2), inspired by Mu et al. [17]. The model processes the flattened input patches in an LSTM layer, with 3 LSTM cells to generate hidden state representations from the last LSTM cell. These are then fed into a multi-head self-attention layer [MHSA] [27] with five attention heads instead of an LSTM decoder [19], which improves the performance by weighting all the preprocessed hidden states from the encoder and produces a matrix of size 4×2500 (timesteps t to t_{+3}). This matrix is then flattened to a 1×10000 vector, fed through a linear layer and projected to a size of 1×2500, producing the output for timestep t_{+4}. This vector is then concatenated to the output of the hidden states of the last LSTM cell from the encoding layer, analogous to the concept of a first-in first-out stack. The decoder then recursively predicts the next hidden state outputs using timesteps t_{+1} until t_{+4}. This process

[1]Comprehensive documentation for Microsoft's Planetary Computer database can be found in the stackstac package for python [25]. Implementations can additionally be found here: `https://github.com/jonasmueler/Glaciers_NeurIPS`.

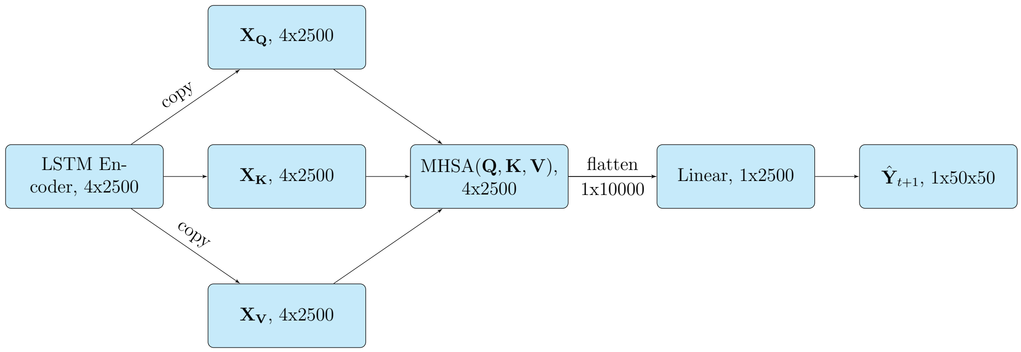

Figure 2. Overview of the SA-LSTM-H architecture with output dimensions represented in the nodes.

is repeated until all future timesteps are predicted, using encoder outputs and model predictions recursively.

3 Case study on Parvati Glacier

The Parvati glacier is well suited for the prediction of glacier movements because the Landsat-8 image catalogue provides a high temporal density of images that are almost evenly distributed over the months in a year, and, because of the high glacier density in the region and the surroundings. Therefore, a dataset is created in which images are averaged over a Δ_t of three months each year from January 1, 2013, until January 1, 2022. If no images are available for a given month, the missing months are interpolated linearly to prevent temporal bias in the data (one month in the whole sequence). The patches are extracted in a regio of interest of 800 $\times$ 800 pixels with a stride of 10, resulting in 161,728 sequences of four input and four corresponding target patches from 36 scenes. The data is then split into 80% training, where 10% of the training data is used for validation, and 20% test data. The dataset is used to test the new model architecture and to compare it to other state-of-the-art models on the test set, namely an LSTM Encoder-Decoder similar to Petrou and Tian [19] and a ConvLSTM similar to Petrou and Tian [20]. As regularization techniques, teacher forcing, dropout, early stopping, and weight decay are used. All used architecture specifications and hyperparameter settings are given in Table 2. All models capture the patches' spatial and temporal dynamics well. The SA-LSTM-H model (Figure 5) outperforms the LSTM encoder-decoder and ConvLSTM models on the test scores (Table 1), which is in line with the theoretical attributes of the SA-LSTM-H architecture that attending sequence elements in parallel increases the performance in comparison to an accumulation of recurrent information like standard LSTM cells [2]. The results indicate that the used approach can predict the scene NDSI masks over time (Figure 3). Additionally we can estimate the optical flow from consecutive predicted scene masks (Figure 4).

Table 1: *Comparison of test set performance*

	MSE	*MAE*
SA-LSTM-H	0.047	0.136
LSTM Encoder-Decoder	0.054	0.139
ConvLSTM	0.055	0.150

The self-attention mechanism of the SA-LSTM-H model increases the overall performance and enables the model to outperform the other two optimized LSTM-based approaches. The chosen hyperparameters for the LSTM-based models contribute to their good performance, especially the combination of relatively large batch sizes, generally resulting in stabilized gradients, [14], and dropout regularization, usually resulting in increased generalization performance [29], might be leading to convergence to an improved local optimum in the loss landscape. Interestingly, the SA-LSTM-H model improves the performance of the LSTM encoder-decoder model while using fewer parameters and outperforming the ConvLSTM model (Table 2). The reduced number of parameters is important as training large ML models also contributes to climate change through their electricity consumption [3]. Lastly, all the models show prediction errors on the glacier termini, with the least extreme effects for the SA-LSTM-H model. These results are in line with the empirical literature as glacier termini respond to many environmental variables, e.g. cloudiness, shading or wind-driven snow accumulation [18].

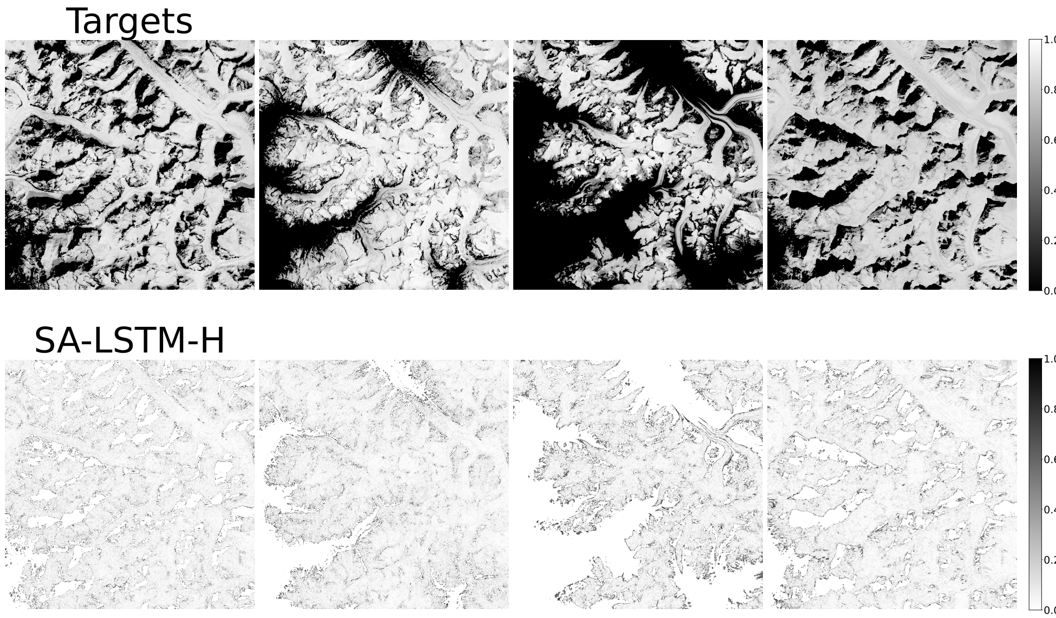

Figure 3. Ground truth target images (first row) and absolute differences between model predictions and targets (second row) on the test set images. Testset images are extracted for the full year 2021 with a Δt of 3 months.

4 Discussion

While the above-introduced approach is generally very promising, a few points exist where more investigation is necessary. Notably, the patch size differs from the one used, for example, by Petrou and Tian [20]. Therefore, it could be interesting to explore the use of different patch sizes and their influence on prediction accuracy, as there might be a trade-off between a smaller model due to decreased spatial dimensions and enough information density that still leads to valid inference. This trade-off could depend on the characteristics of a given glacier and can be estimated from features extracted from the glacier Furthermore, the NDSI is used with a threshold of 0.3 as a criterion for snow classification in the scene. While this approach is taken from Vonica et al. [28], it could be interesting to take other channel dimensions as model inputs or use a multitude of different thresholds from 0 to 0.7 as He et al. [12]. And lastly, taking the local dependency between patches into account could further improve the prediction accuracy of the used approach. This could be operationalized, for example, with a method that offers the model spatially close patches as input in addition to the patches from previous timesteps. For the borders of the images, padding could be used to counter the non-existing neighboring pixels.

5 Conclusion

Building on established methods in the literature (e.g. Vonica et al. [28] and Petrou and Tian [19, 20]), the present study developed a method to estimate glacier movement changes over time from Landsat-8 satellite image data. The experiments showed that it is possible to extract enough valid data from the STAC API to train deep learning models, align this data in a way to make patch prediction over time possible and that the SA-LSTM-H model can predict the patch sequences correctly to estimate optical flow vectors from pure model predictions. Therefore, the approach could be used further for monitoring glaciers and climate change effects in addition to in-situ methods to detect dangerous abnormal changes and their consequences. As the model could successfully predict glacier movements up to 12 months ahead, the pipeline could be implemented as a forecasting system using freely available Landsat-8 or similar [21] satellite data. Landsat-8 imaging spans the globe, making the approach scalable to many glaciers.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645 and supported by the Tübingen AI Center.

References

[1] Nerilie Abram, Jean-Pierre Gattuso, Anjal Prakash, Lijing Cheng, María Paz Chidichimo, Susan Crate, Hiroyuki Enomoto, Matthias Garschagen, Nicolas Gruber, Sherilee Harper, et al. Framing and context of the report. *IPCC special report on the ocean and cryosphere in a changing climate*, pages 73–129, 2019.

[2] Sahara Ali, Yiyi Huang, Xin Huang, and Jianwu Wang. Sea ice forecasting using attention-based ensemble lstm. *arXiv preprint arXiv:2108.00853*, 2021.

[3] Lasse F Wolff Anthony, Benjamin Kanding, and Raghavendra Selvan. Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. *arXiv preprint arXiv:2007.03051*, 2020.

[4] Michael Avian, Christian Bauer, Matthias Schlögl, Barbara Widhalm, Karl-Heinz Gutjahr, Michael Paster, Christoph Hauer, Melina Frießenbichler, Anton Neureiter, Gernot Weyss, et al. The status of earth observation techniques in monitoring high mountain environments at the example of pasterze glacier, austria: Data, methods, accuracies, processes, and scales. *Remote Sensing*, 12(8):1251, 2020.

[5] Etienne Berthier, Dana Floricioiu, Alex S Gardner, Noel Gourmelen, Livia Jakob, Frank Paul, Désirée Treichler, Bert Wouters, Joaquin MC Belart, Amaury Dehecq, et al. Measuring glacier mass changes from space-a review. *Reports on Progress in Physics*, 2023.

[6] Junyi Chai, Hao Zeng, Anming Li, and Eric WT Ngai. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, 6:100134, 2021.

[7] Xinde Chu, Xiaojun Yao, Hongyu Duan, Cong Chen, Jing Li, and Wenlong Pang. Glacier extraction based on high-spatial-resolution remote-sensing images using a deep-learning approach with attention mechanism. *The Cryosphere*, 16(10):4273–4289, 2022.

[8] William Colgan, Harihar Rajaram, Waleed Abdalati, Cheryl McCutchan, Ruth Mottram, Mahsa S Moussavi, and Shane Grigsby. Glacier crevasses: Observations, models, and mass balance implications. *Reviews of Geophysics*, 54(1):119–161, 2016.

[9] Mark B Dyurgerov and Mark F Meier. Twentieth century climate change: evidence from small glaciers. *Proceedings of the National Academy of Sciences*, 97(4):1406–1411, 2000.

[10] Georgios D Evangelidis and Emmanouil Z Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE transactions on pattern analysis and machine intelligence*, 30(10):1858–1865, 2008.

[11] Peter H Gleick. Water resources. *Encyclopedia of climate, weather*, pages 817–823, 1996.

[12] Q He, Z Zhang, G Ma, and J Wu. Glacier identification from landsat8 oli imagery using deep u-net. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:381–386, 2020.

[13] Melanie Marochov, Patrice Carbonneau, and Chris Stokes. Automated image classification of greenlandic outlet glaciers using deep learning: A case study on helheim glacier. 2020.

[14] Dominic Masters and Carlo Luschi. Revisiting small batch training for deep neural networks. *arXiv preprint arXiv:1804.07612*, 2018.

[15] Nausheen Mazhar, Ali Iqtadar Mirza, Sohail Abbas, Muhammad Ameer Nawaz Akram, Muhammad Ali, and Kanwal Javid. Effects of climatic factors on the sedimentation trends of tarbela reservoir, pakistan. *SN Applied Sciences*, 3:1–9, 2021.

[16] Yimeng Min, S Karthik Mukkavilli, and Yoshua Bengio. Predicting ice flow using machine learning. *arXiv preprint arXiv:1910.08922*, 2019.

[17] Bin Mu, Xiaodan Luo, Shijin Yuan, and Xi Liang. Icetft v 1.0. 0: Interpretable long-term prediction of arctic sea ice extent with deep learning. *Geoscientific Model Development Discussions*, pages 1–28, 2023.

[18] Osita Onyejekwe, Bryan Holman, and Nezamoddin N Kachouie. Multivariate models for predicting glacier termini. *Environmental Earth Sciences*, 76:1–10, 2017.

[19] Zisis I Petrou and YingLi Tian. Prediction of sea ice motion with recurrent neural networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5422–5425. IEEE, 2017.

[20] Zisis I Petrou and Yingli Tian. Prediction of sea ice motion with convolutional long short-term memory networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6865–6876, 2019.

[21] Darius Phiri, Matamyo Simwanda, Serajis Salekin, Vincent R Nyirenda, Yuji Murayama, and Manjula Ranagalage. Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, 12(14):2291, 2020.

[22] Colin Prieur, Antoine Rabatel, Jean-Baptiste Thomas, Ivar Farup, and Jocelyn Chanussot. Machine learning approaches to automatically detect glacier snow lines on multi-spectral satellite images. *Remote Sensing*, 14(16):3868, 2022.

[23] Ted A Scambos, Robin E Bell, Richard B Alley, S Anandakrishnan, DH Bromwich, K Brunt, K Christianson, T Creyts, SB Das, R DeConto, et al. How much, how fast?: A science review and outlook for research on the instability of antarctica's thwaites glacier in the 21st century. *Global and Planetary Change*, 153:16–34, 2017.

[24] Maheswor Shrestha, Toshio Koike, Yukiko Hirabayashi, Yongkang Xue, Lei Wang, Ghulam Rasul, and Bashir Ahmad. Integrated simulation of snow and glacier melt in water and energy balance-based, distributed hydrological modeling framework at hunza river basin of pakistan karakoram region. *Journal of Geophysical Research: Atmospheres*, 120(10):4889–4919, 2015.

[25] stackstac. Animated gif from stac items — stackstac documentation. `https://stackstac.readthedocs.io/en/latest/examples/gif.html`, 2021. Accessed [15 May 2023].

[26] Markus Stoffel and Christian Huggel. Effects of climate change on mass movements in mountain environments. *Progress in physical geography*, 36(3):421–439, 2012.

[27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[28] Maria-Minerva Vonica, Andrei Ancuta, and Marc Frincu. Glacier movement prediction through computer vision and satellite imagery. In *2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 113–120. IEEE, 2021.

[29] Stefan Wager, Sida Wang, and Percy S Liang. Dropout training as adaptive regularization. *Advances in neural information processing systems*, 26, 2013.

[30] Aijie Yu, Hongling Shi, Yifan Wang, Jin Yang, Chunchun Gao, and Yang Lu. A bibliometric and visualized analysis of remote sensing methods for glacier mass balance research. *Remote Sensing*, 15(5):1425, 2023.

[31] Michael Zemp. *Global glacier changes: facts and figures*. UNEP/Earthprint, 2008.

[32] Jun Zhai and Cecilia M Bitz. A machine learning model of arctic sea ice motions. *arXiv preprint arXiv:2108.10925*, 2021.

Appendix

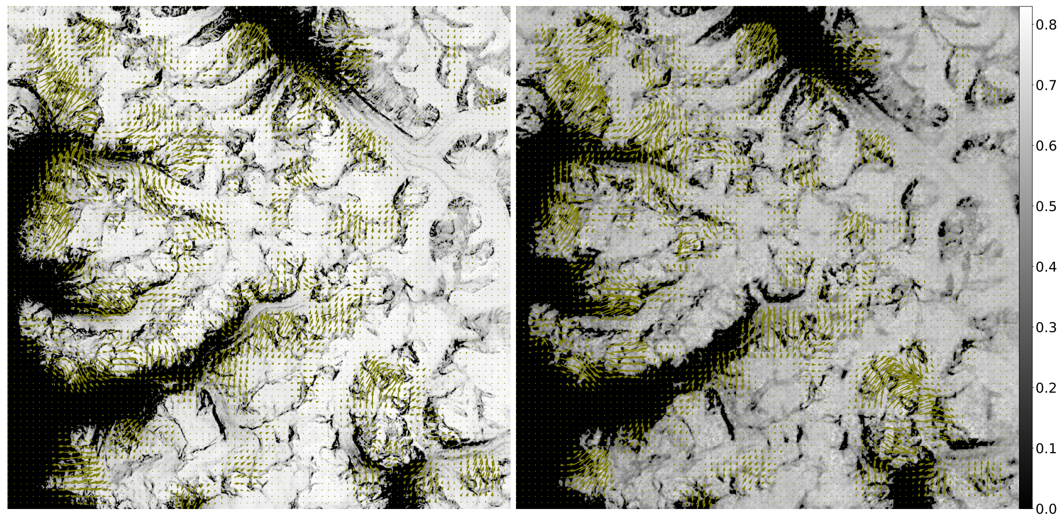

Figure 4. Optical flow estimations based on ground truth masks (left) and predictions of the SA-LSTM-H model (right) for averaged masks of April, May, and June 2021 (first mask) and Juli, August, and September 2021 (second mask).

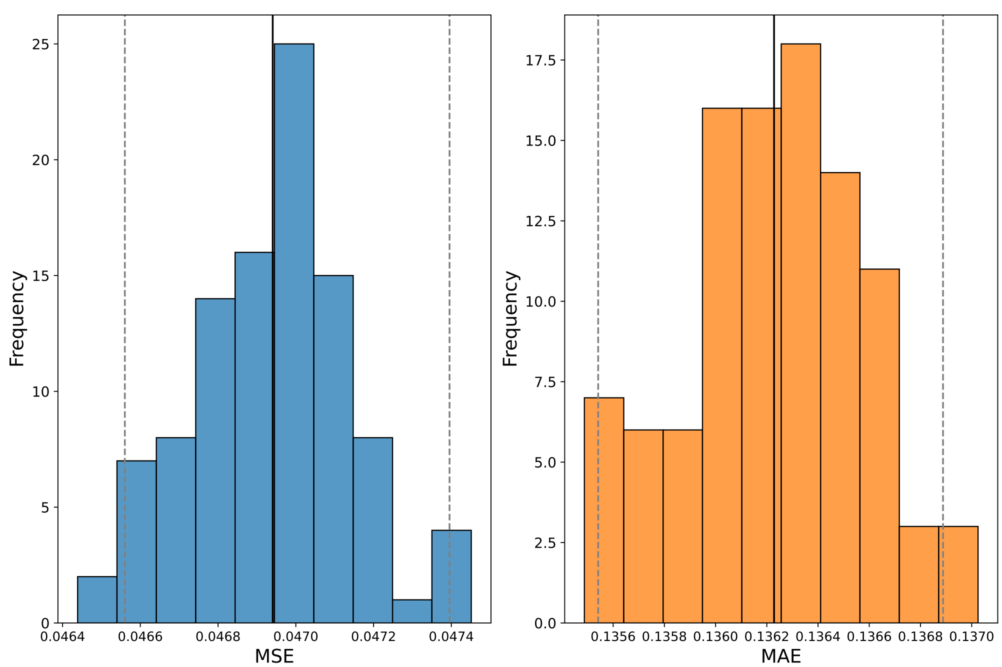

Figure 5. Boostrapped mean squared and mean absolute error of the SA-LSTM-H model on the test set (100 bootstrap iterations). Dotted lines indicate 95% confidence bounds, solid lines indicate mean values.

Table 2: *Model Parameters*

Model	LR	WD	Dr.	Epoch	BS	Opt.	Param.
LSTM	.0001	.001	0.1	40	100	adamW	300 120 000
SA-LSTM-H	.0001	.001	0.1	40	100	adamW	200 072 500
ConvLSTM	.0001	.001	-	35	100	adamW	842 233

Note. LR is the learning rate, WD is the used weight decay, Dr. is the dropout parameter, Epoch is the epoch number, BS is the batch size, Opt. is the optimizer, and Param. is the parameter count used.