Review of "egusphere-2025-1206"

Brian Kyanjo, PhD

May 17, 2025

Overall Assessment

The paper "egusphere-2025-1206" (https://doi.org/10.5194/egusphere-2025-1206) presents the Mass Balance Machine (MBM), a machine learning model for predicting glacier surface mass balance, trained on a robust dataset of 4,201 point mass balance measurements from 32 Norwegian glaciers (1962–2021) using the XGBoost algorithm. This work is a significant advancement in glaciology, offering high-resolution predictions that outperform traditional models like GloGEM, OGGM, and PyGEM, particularly for seasonal mass balances. Its potential for applications in climate change research and water resource management is substantial. However, minor refinements in data resolution, model transparency, and uncertainty analysis, along with clarifications in Appendix A, could elevate its impact. Below, I provide a detailed review of the paper's strengths, areas for improvement, and a focused analysis of Appendix A, including specific corrections and broader recommendations.

Strengths of the Paper

1. **Robust Dataset**: The dataset, sourced from the Norwegian Water Resources and Energy Directorate (NVE) database, spans nearly six decades and includes 4,201 point mass balance measurements across 32 glaciers. The thorough cleaning process, detailed in Appendix A, ensures reliability by addressing missing coordinates and outliers, resulting in 3,910 annual, 3,929 summer, and 3,751 winter measurements.

2. **Effective Methodology**: The use of XGBoost is well-suited for capturing complex relationships between weather, terrain, and glacier mass balance. The independent glacier-based train-test split enhances the model's generalizability, making results trustworthy.

3. **Superior Performance**: MBM demonstrates lower RMSE and bias compared to established models, particularly for seasonal predictions. This is evident in figures like Fig. 6 and Table D1, which effectively support the text.

4. **Practical Applications**: The model's high-resolution predictions at point and monthly scales are valuable for water resource planning and glacier flow modeling in a warming climate.

Areas for Improvement

1. **Data Resolution**: The ERA5-Land data (9 km resolution) may be too coarse for smaller glaciers. Exploring higher-resolution datasets or downscaling techniques could improve local accuracy.

2. **Model Transparency**: XGBoost's complexity warrants feature importance analysis or partial dependence plots to clarify key drivers of predictions, enhancing interpretability.

3. **Uncertainty Quantification**: While measurement uncertainties (0.08–0.26 m w.e. a^{-1}) are noted, their impact on model outputs is unclear. A sensitivity analysis would strengthen confidence in predictions.

4. **Global Applicability**: Testing MBM in diverse regions like the Alps or Himalayas would broaden its relevance. A discussion of transferability challenges would be valuable.

5. **Future Directions**: The mention of remote sensing data is promising but vague. Specifying datasets (e.g., satellite-derived albedo or surface temperature) would clarify future enhancements.

6. **Presentation Polish**: Minor typos and awkward sentences, particularly in Appendix A, need correction. Additionally, Fig. 10 requires clearer labels for improved readability.

Detailed Review of Appendix A

Appendix A details the data quality and cleaning processes for the MBM dataset, critical for establishing its reliability. It describes the handling of 4,201 point mass balance measurements from the NVE database (accessed 12 October 2022), including the removal of erroneous entries and verification of stake locations. Below, I identify specific typos and awkward sentences with approximate line numbers (based on sequential sentence or paragraph counting) and provide broader recommendations to enhance clarity.

Identified Typos and Awkward Sentences

1. **Line 570**: "The total contribution of such uncertainties have been quantified 0.08-0.26 m w.e. a^{-1}..."

    - *Issue*: Subject-verb agreement error; "have" should be "has" for the singular subject "The total contribution."
    - *Suggestion*: Revise to "The total contribution of such uncertainties has been quantified as 0.08–0.26 m w.e. a^{-1} for five glaciers in our dataset."

2. **Line 575**: "Prior to training MBM, we performed a thorough cleaning and quality check... including removal of erroneous values and points with missing location, and a quality check of stake locations."

    - *Issue*: Redundant use of "quality check."

- *Suggestion*: Streamline to "Prior to training MBM, we performed thorough cleaning and quality checks on the raw point mass balance dataset (4,201 entries, NVE database, accessed 12 October 2022), removing erroneous values, points with missing locations, and verifying stake location accuracy."

3. **Line 578**: "Approximate locations are based on the approximate position and elevation of a given stake ID..."
   - *Issue*: Repetition of "approximate" is awkward.
   - *Suggestion*: Revise to "Approximate locations are derived from the estimated position and elevation of a given stake ID, whereas exact locations use GPS-measured position and elevation at the time of measurement."
4. **Line 581**: "Seven and 23 entries that were missing both exact and approximate elevation or geographical coordinates, respectively, were removed..."
   - *Issue*: Ambiguous phrasing regarding elevation and coordinates.
   - *Suggestion*: Clarify to "Seven entries missing both exact and approximate elevations and 23 entries missing both exact and approximate geographical coordinates were removed from the training dataset."
5. **Line 585**: "The mean ± standard deviation of the absolute difference between the exact and approximate coordinates and elevations is 166 ± 498 m and 24 ± 71 m, respectively."
   - *Issue*: Dense phrasing combines measurements, reducing clarity.
   - *Suggestion*: Split to "The mean ± standard deviation of the absolute difference between exact and approximate coordinates is 166 ± 498 m, while for elevations, it is 24 ± 71 m."
6. **Line 589**: "For stake locations where both summer, winter and annual mass balance measurements were available..."
   - *Issue*: List lacks an Oxford comma for clarity.
   - *Suggestion*: Revise to "For stake locations where summer, winter, and annual mass balance measurements were all available for a given year..."

Additional Recommendations for Appendix A

1. **Improve Transitions**: The shift from uncertainties to data cleaning is abrupt. Add a bridging sentence, e.g., "Ensuring dataset quality is crucial for MBM's accuracy, leading to the following cleaning procedures."
2. **Define Technical Terms**: Define "point mass balance" (e.g., "measurements of mass change at specific glacier locations") in a footnote or glossary for accessibility.
3. **Clarify Data Sources**: Specify that NVE is the Norwegian Water Resources and Energy Directorate to aid international readers.
4. **Quantify Cleaning Impact**: State the total entries removed, e.g., "After cleaning, the dataset was reduced from 4,201 to 4,170 stake locations (99.3% retained)."

5. **Explain Coordinate Conversion**: Justify the UTM to latitude/longitude conversion, e.g., "This conversion ensured compatibility with MBM's input requirements."

6. **Justify Erroneous Values**: Explain the removal of the 9.99 m w.e. measurement, e.g., "This value was unrealistically high for typical regional winter mass balances."

7. **Quantify Corrections**: If available, note the number of rounding error corrections, e.g., "In [X] instances, annual mass balances were corrected by summing summer and winter components."

Conclusion

The "egusphere-2025-1206" paper is a compelling contribution to glaciology, with MBM offering high-resolution, accurate predictions for glacier mass balance. Its robust dataset, effective methodology, and practical applications make it a valuable tool. Minor revisions, including addressing typos in Appendix A, improving data resolution, and enhancing model transparency, will further strengthen its impact. Appendix A effectively supports the dataset's reliability but can be polished with clearer transitions, defined terms, and quantified impacts. These changes require minimal effort but will significantly enhance the paper's clarity and global relevance.

Recommendations

- **Accept with Minor Revisions**.
- **Specific Actions**:
  - Correct the six typos and awkward sentences in Appendix A as suggested.
  - Explore higher-resolution weather data or downscaling for smaller glaciers.
  - Add feature importance or partial dependence plots for model transparency.
  - Conduct a sensitivity analysis to quantify uncertainty impacts.
  - Discuss testing MBM in other regions for global applicability.
  - Specify remote sensing datasets (e.g., albedo, surface temperature) for future work.
  - Add a transitional sentence in Appendix A between uncertainties and cleaning.
  - Define "point mass balance" in a footnote or glossary.
  - Clarify NVE as the Norwegian Water Resources and Energy Directorate.
  - Quantify total entries removed during cleaning (e.g., 4,201 to 4,170).
  - Justify UTM to latitude/longitude conversion and the 9.99 m w.e. removal.
  - Note the number of rounding error corrections, if available.
  - Improve Fig. 10 labels for clarity.
  - Include missing DOIs or URLs in the reference section.