We thank both reviewers for their comments and their positive assessment of the manuscript. We have made a number of changes to our manuscript in response to the recommendations, in particular focusing on expanding and clarifying the methods section. We have added a new, more comprehensive methods flowchart, and additional methods table for the core machine learning modules, and additional details about different elements of the pre and post processing. We have also reviewed the full manuscript text and made small adjustments and corrections where recommended, which are highlighted in the 'track-changes' manuscript. We respond to all reviewer comments between the lines below.

Reviewer 1

The paper introduces an MOS (Model Output Statistics) method using artificial intelligence (machine learning) suitable for at-scale rapid deployment, for correcting deviations of numerical weather prediction (NWP) in complex mountainous areas. It uses the Mount Everest climbing meteorological service in the Himalayas as a pilot study to validate the method's feasibility. The paper also discusses the advantages, potential issues, and risks of this method.

We thank the reviewer for these comments and have responded between the lines below.

The main review comments and suggestions are as follows:

1. As a manuscript intended for publication in GMD (Geoscientific Model Development), there needs to be a very detailed description of the described technical methods and model. In this manuscript, the technical details of AtsMOS need further refinement and organization. It is preferable to provide a more detailed flowchart than Figure 1, or to add detailed sub-flowcharts for each module, accompanied by text descriptions, especially for the implementation process, parameter settings of XGBoost and RF, etc., to enhance the practical reference value of this open-access paper.

We agree with this comment and have modified our manuscript so as to clarify these aspects. We acknowledge the need for a more detailed description of the technical methods and model. We have substantially expanded our methods section, adding a new figure (detailed workflow) and table (details of ML models implemented). We have divided this section into different parts and added further details where needed. Finally, we have added further references to our documented jupyer notebook, which we view as a parralel resource to the manuscript for users aiming for an in-depth understanding our our methods. These additions will ensure that readers have a comprehensive understanding of the AtsMOS workflow and can effectively replicate and apply the methodology in their own research.

2. The paper only conducts simulated comparative analysis and verification based on observations and forecasts of Mount Everest in the Himalayas. On the one hand, for the verification of weather transition stages (rapid temperature decrease or increase, rapid increase or decrease in wind speed), a detailed analysis is needed. On the other hand, if possible, more experiments and comparative analyses can be conducted with richer observations and AtsMOS forecasts in other mountainous regions around the world (such as the European Alps, the Rocky Mountains in the United States) to strengthen the reliability and universality validation of this method.

Thank you for your insightful comments. We chose to focus on Mount Everest as a pilot study due to its extreme meteorological conditions and clear need for skilled forecasts, which provide a useful testbed for validating the AtsMOS workflow. Future work conducting additional case studies in other mountainous regions, such as the European Alps or the Rocky Mountains, would indeed provide further validation of our method. For this manuscript we prioritise a thorough analysis of a single, challenging environment and reserve the application of AtsMOS to other mountainous regions for future work. While further validation is always valuable, we believe that the Mt Everest case study presented provides sufficient information to demonstrate that the technique is viable and is of sufficient interest to include in this GMD paper.

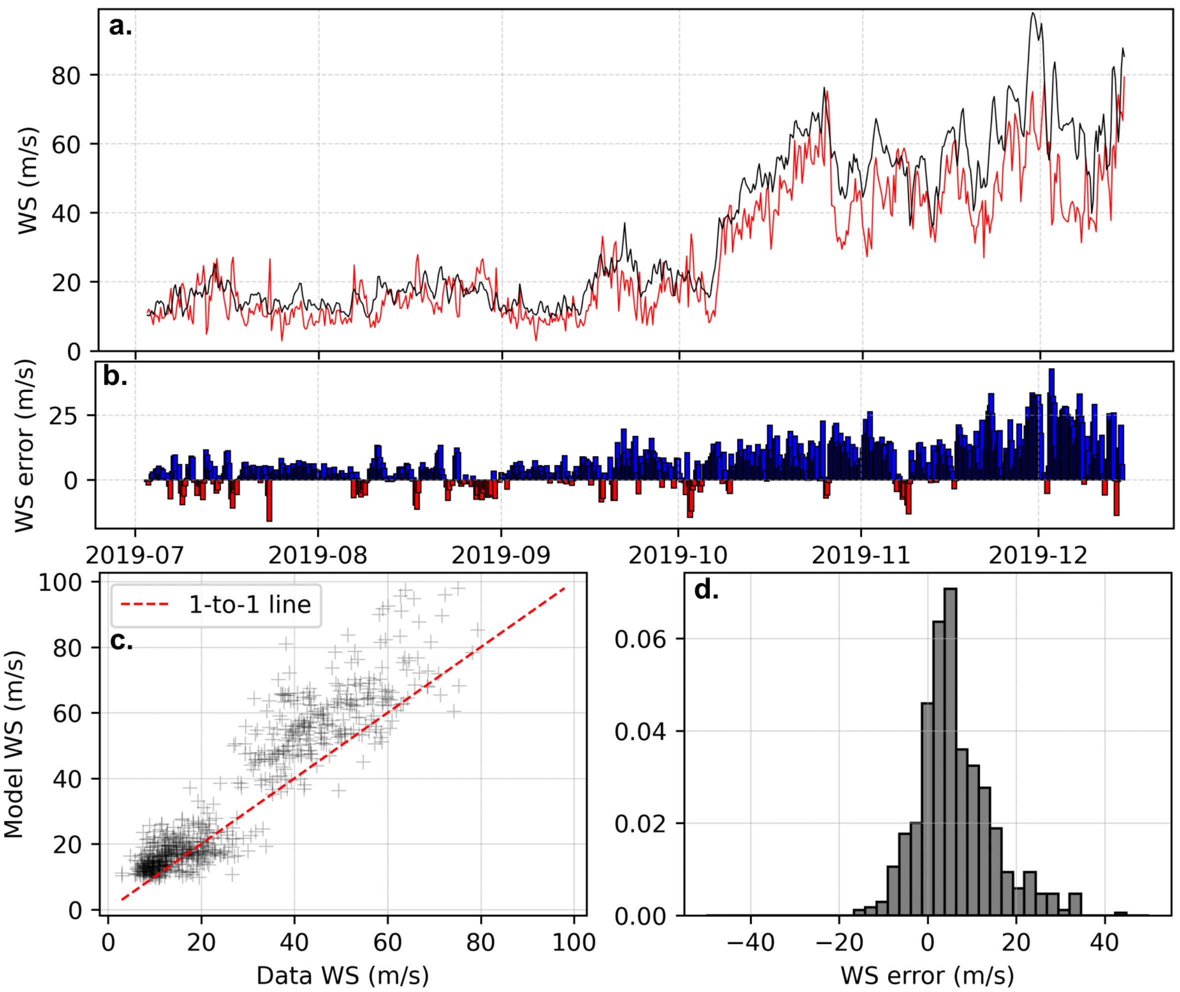
3. The textual presentation of the paper needs to be more rigorous. For example, some abbreviations need to be provided in full, or a list of abbreviations can be provided at the end of the paper. Sentence expressions need to be more rigorous, and writing needs to be standardized (such as subscript and superscript issues, unit measurement issues, meteorological professional expression issues, etc.).

Thank you for your suggestions regarding the textual presentation. We have thoroughly reviewed the manuscript to ensure that all abbreviations are provided in full upon first use. Additionally, we have standardized the writing to address issues with subscript and superscript formatting, unit measurements, and meteorological professional expressions where these could be identified. We hope that these changes improve the overall clarity of the manuscript.

4. For Figures 5-7, it is recommended to extract the data segment that simultaneously includes observation and forecast results and redraw clearer graphs (or add a curve showing the difference between the two ones). The current figures do not clearly show the specific differences.

We thank the reviewer for these suggestions, which we have implemented in the new version of these figures. We both cropped the first line graph to only the region of overlap between data and model, and provided an additional ‘running error’ plot to accompany this (b). As noted below, we also replaced the plot density graph with an error distribution histogram (d). We hope that these modified figures 5-7 (now, figures 6-8 in the new manuscript) better highlight the differences between the methods and capacity of the MOS technique to produce skillful forecasts.

Figure 6, showing the results for the linear regression:



5. The titles of all figures need to be further refined to increase clarity.

Thank you for your suggestion to refine the titles of all figures. We have revised the figure captions to increase clarity and ensure they succinctly describe the content of each figure.

6. The description of Kling-Gupta efficiency needs to be clarified. This evaluation index is mainly used in hydrology. Whether it is suitable for this work should be clearly explained.

We acknowledge the need to clarify the use of Kling-Gupta Efficiency (KGE) which, while traditionally used in hydrology, is a robust metric that combines correlation, bias, and variability, making it well-suited for evaluating the performance of meteorological models as well. We have added a new paragraph in our methods section on evaluation metrics to clarify our use of KGE and to explain its relevance and applicability to our study. Additionally, we continue to use other common metrics such as MAE and RMSE.

7. An interesting question is whether further verification and comparison can be conducted for forecast results similar to Figure 8. If the Everest climbing team(s) or guides have records or carry instruments with similar data, such verification and comparison can be conducted. I believe under such extreme geographical conditions in Everest, this serves as a meaningful validation and assessment of the AtsMOS method.

Thank you for your suggestion. Unfortunately, the Everest climbing teams or guides do not currently carry instruments that provide data comparable to our forecasts. However, we agree that such verification would be highly valuable. In the discussion section, we will include a recommendation that climbers in extreme locations like Everest report simple yes/no feedback on climbable conditions, if not more detailed meteorological information. This feedback could enhance forecast validation and assessment. The flexibility of our ML system allows it to predict such binary 'climbable/not climbable' outcomes from the GFS or other forecasting systems, bridging the gap between complex data and practical decision-making for climbers.