https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development
Discussions

At-scale Model Output Statistics in mountain environments (AtsMOS v1.0)

Maximillian Van Wyk de Vries[1,2,3], Tom Matthews[4], L. Baker Perry[5], Nirakar Thapa[6], and Rob Wilby[7]

[1]Department of Geography, University of Cambridge, Cambridge CB2 3EL, UK.
[2]Department of Earth Sciences, University of Cambridge, Cambridge CB3 0EZ, UK.
[3]School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, UK.
[4]Department of Geography, Kings College London, London WC2B 4BG, UK.
[5]Department of Geography and Planning, Appalachian State University, Boone, North Carolina, USA.
[6]Department of Hydrology and Meteorology, Kathmandu, Nepal.
[7]Department of Geography and Environment, Loughborough University, Loughborough, UK

Correspondence: M. Van Wyk de Vries (msv27@cam.ac.uk)

Abstract. This paper introduces the AtsMOS workflow, designed to enhance mountain meteorology predictions through the downscaling of coarse numerical weather predictions using local observational data. AtsMOS provides a modular, open-source toolkit for local and large-scale forecasting of various meteorological variables through modified Model Output Statistics – and may be applied to data from a single station or an entire network. We demonstrate its effectiveness through an example application at the summit of Mt. Everest, where it improves the prediction of both meteorological variables (e.g. wind speed, temperature) and derivative variables (e.g. facial frostbite time) critical for mountaineering safety. As a bridge between numerical weather prediction models and ground observations, AtsMOS help produce insights for hazard mitigation, water resource management, and other weather-dependant issues in mountainous regions and beyond.

1 Introduction

Accurate mountain weather forecasts are critically important for society. They facilitate improved hazard mitigation for the 300 million mountain inhabitants worldwide and are important for effective resource management (e.g., Miner et al., 2020; Corbari et al., 2022). The latter is also relevant to the 1.6 billion who live downstream of mountains and, therefore, depend to varying extents on their supply of freshwater (Immerzeel et al., 2020). However, producing skilful forecasts in such environments is challenging. Major topographic variations cause similarly pronounced spatial variability in the weather, meaning that reality can diverge a long way from Numerical Weather Prediction (NWP) grid-point forecasts within typical grid-cell areas (hundreds to thousands of square kilometres) (Zhang et al., 2022). Whilst simple biases can be readily adjusted for (e.g., mismatches in elevation between forecast grid points and land surface locations of interest with knowledge of the lapse rate; Minder et al., 2010), the impact of unresolved processes – for instance local valley or glacier winds driven by surface heat fluxes (Khadka et al., 2022) – is harder to correct for a priori.

Although advances in NWP (e.g., finer grid resolutions and refinement of physical parameterisation schemes) may enhance forecast performance in mountainous terrain, progress can be costly and slow (Bauer et al., 2015). A cheaper, faster, and

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development
Discussions

more flexible option to improve forecasts for target locations is to statistically post-process NWP output through calibration to observations. Model Output Statistics (MOS) – which applies multiple linear regression to adjust forecast fields – has historically been the most popular method in this regard (Glahn and Lowry, 1972; Glahn, 2014), not least because it can be used to create forecasts of variables (predictands) not available in NWP model output (Rasp et al., 2020). Encouragingly, recent advancements in computational power have enabled machine learning to improve the performance of weather forecasts (Lam et al., 2023), including through post-processing (e.g., Lagerquist et al., 2017; Herman and Schumacher, 2018; Han et al., 2021; Grönquist et al., 2021).

As positive as such developments are, however, practical barriers may limit their take up at scale. For example, without template workflows to demonstrate the non-trivial tasks of pre-processing (very large) NWP datasets, training and evaluating appropriate machine-learning models, and automating real-time predictions, forecasts improved by machine learning are unlikely to reach the diverse range of potential end users in mountainous environments (Table 1). We also note that the benefits of highly accurate local weather predictions for use in other (e.g., hydrological) modelling chains may not be achieved if such forecasts not made available in an interoperable format that follows well-known (e.g., 'CF' – Climate and Forecast) conventions (Eaton et al., 2023).

Table 1. Examples of weather variables (predictands) and sectors in which highly accurate, site-specific forecasts may be desirable. [1]The term mountaineering is used to represent a wider set of similar activities – e.g., hiking, skiing and climbing.

Predictand	**Example sector(s)**
Precipitation amount and phase	Hazard forecasting (flood, avalanche); resource planning
Maximum wind gust	Aviation; [1]mountaineering; hazard (avalanche) forecasting
Ground temperature	(Road) transport; mountaineering
Wind chill temperature	Mountaineering
Cloud base and cloud top	Aviation; mountaineering
Probability of rime ice accretion	Communications
Facial frostbite time	Mountaineering

Hence, our paper aims to introduce a user-friendly, lightweight version of MOS to fill this gap. We describe modular Python code that calibrates and applies MOS, including state-of-the-art machine learning algorithms, to produce corrected forecasts in an interoperable format that can feed into other automated workflows to enable at-scale MOS. We anticipate that these features of AtsMOS will, combined with efforts to improve the availability of high-altitude weather observations worldwide (GEO Mountains 2022), offer a step change in the ability to forecast critical mountain weather variables.

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development
Discussions
Open Access
EGU

In Section 2 we describe the main features of AtsMOS, before illustrating its use in forecasting the weather on the summit of Mt. Everest, where highly accurate predictions can be the difference between life and death (Section 3). In Section 4 we discuss opportunities and challenges in using AtsMOS more broadly.

2 The AtsMOS workflow

AtsMOS is designed to be a computationally light and flexible template. It has (i) a flexible loading and preprocessing module, which draws in external data, deals with erroneous or missing data and prepares it for further analysis. Our code here is intended as a guide such that users may set up their own data loading and pre-processing as the need arises. (ii) A core processing module, comprising a modular suite of statistical and machine learning techniques to calibrate and perform data corrections, with XGBoost being the default and most advanced option. (iii) A post-processing module to calculate derivative variables and export the data in the self-describing and interoperable MDF format (GEO Mountains, 2022; Figure 1).

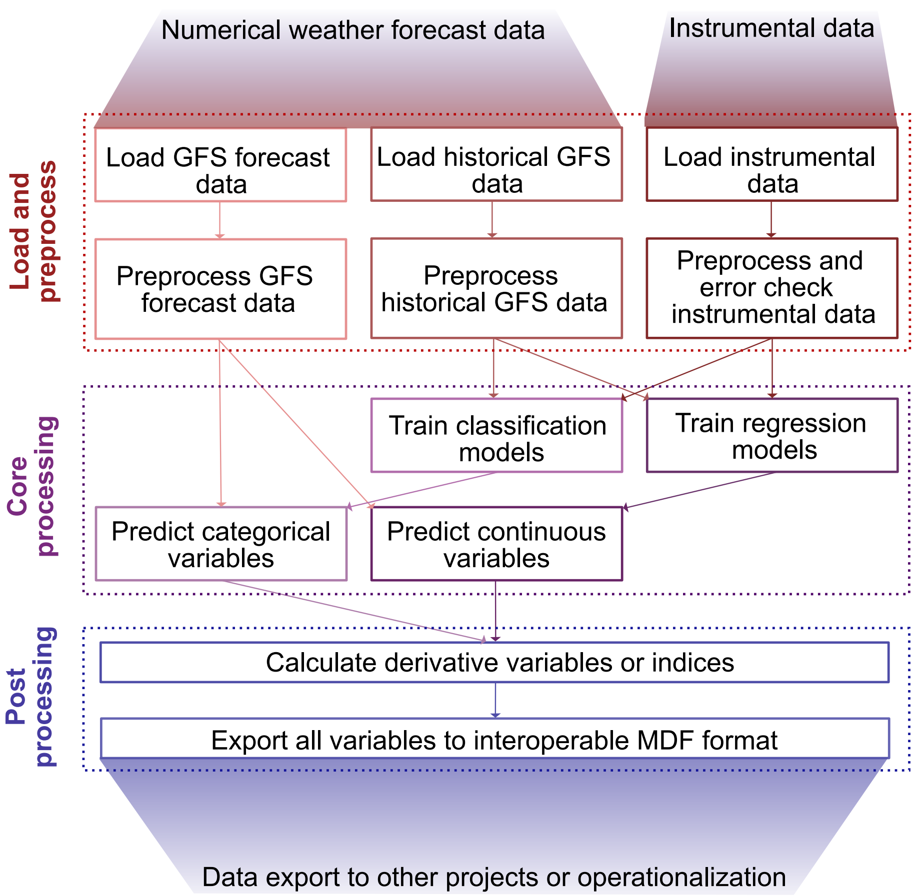

Figure 1. Diagram of the AtsMOS workflow. The loading of historical GFS data and loading and preprocessing of instrumental data are flexible components subject to user modification, while the others are fixed in this workflow.

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development Discussions

AtsMOS is currently designed to be used with data from the Global Forecasting System (GFS) data from the US National Centre for Atmospheric Research (https://rda.ucar.edu/datasets/ds084.1/), which is freely available on a global scale and real-time basis. GFS forecasts are computed every 6 h, with a lead time from 0 to 384 hours (16 days). Pressure-level data are generally preferred for mountain forecasting applications because the real-world surface in such regions is likely to be very different (e.g., in elevation and surface type) from the model surface (Mass et al., 2008), and hence we anticipate greater general predictability using data from the free atmosphere. We evaluated different methods for accessing GFS data and found that the web subsetting form (rda.ucar.edu/datasets/ds084.1/dataaccess/) is generally the most convenient for accessing historical archive data, while the online THREDDS server is best for real-time data. As such we include a preset module in AtsMOS for both automatically downloading and pre-processing real-time data, but only for pre-processing of historical archive data. Historical archive data need only be downloaded once for pre-trained MOS models to be created (see below), which may then be run on any real-time data. We include example scripts used to preprocess instrumental data, for instance, synchronizing measurement and NWP measurement frequencies and removing unreliable data, but note that these are heavily dependent on the type and location of the sensor. We encourage users to carefully consider what if any, processing steps are necessary to field data treated as 'ground truth', as any errors or biases remaining will be learnt by the model. We discuss this in further detail in our limitations section.

For the core processing, AtsMOS applies Model Output Statistics (MOS) to the GFS data, with a range of possible correction algorithms for the user to select from depending on predictand type (e.g., binary or continuous) and the weighting of interpretability versus performance (Table 2). In our illustration below (Section 3), we compare the results from applying simple linear regression and XGBoost (Chen and Guestrin, 2016). Linear regression works well when the relationship between the predictor and target variable is approximately linear. Its coefficients provide clear insights into the impact of each feature, making it valuable for tasks where interpretability is crucial. However, linear regression cannot resolve nonlinear relationships in the data (without transformations to the input variables) and is sensitive to data quality and outliers limiting its predictive performance in many real-world cases. XGBoost is at the other end of the complexity spectrum, combining decision trees with gradient boosting to improve computational efficiency and predictive performance, particularly in high-dimensional, nonlinear data scenarios. It has been shown to outperform most methods in terms of predictive accuracy (Chen and Guestrin, 2016) and is robust to overfitting, but its complexity can make it less suitable in cases where model transparency is essential and understanding the reasons behind incorrect predictions is key.

Once the appropriate ML model has been trained using the historical data, AtsMOS can process the real-time GFS forecast to produce corrected forecasts. The calibrated forecast may be a continuous variable (e.g. wind speed), a probability (e.g. probability of winds above a given threshold), or a binary categorized field (e.g. winds above or below a given threshold) depending on the processing choices made and project requirements. We also highlight that the flexible approach of AtsMOS enables prediction of any variable for which observations exist, and which are sensitive to the atmospheric state. We showcase this in Section 3, making predictions for facial frostbite time – an important variable for mountaineering, which is not available from NWP output.

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

As a final stage in AtsMOS, the corrected forecast variables are saved along with their metadata in the self-describing and interoperable MDF format (GEO Mountains 2022). This final export stage has three separate benefits: (i) it enables easy usage of the custom forecasts in other applications, or plotting dashboards; (ii) it ensures that the variables are saved with all necessary context for long-term archiving; and (iii) through standardized nomenclature, it enables easy comparison with other forecast datasets and external validation. Overall, the AtsMOS workflow is designed to be lightweight and flexible, while retaining sufficient complexity to add value to large-scale forecasts based on local observations.

3 Example application: Mt Everest summit meteorology

3.1 Background

As the highest peak on Earth, Mt. Everest sees hundreds of attempts of its 8850 m a.s.l summit each year. Fatalities are also common, including 17 fatalities in spring 2023 (Ellis-Petersen, 2023), and an overall mortality rate of around 1 % in recent years (Huey et al., 2020). The weather has been a significant contributor to these, historically playing a role in 25 % of deaths (Firth et al., 2008), consistent with the significant hazard from extremely low barometric pressure (low oxygen availability) and severe cold hazard that climbers may be exposed to (Moore and Semple, 2006; Matthews et al., 2020a, 2022). The latter is very sensitive to wind speed (Moore and Semple, 2011), which may also place climbers at risk of being blown from the mountain. It is for this reason that climbers limit their summit attempts to periods when the Subtropical Jet's retreat leaves lower wind speeds on the mountain. Therefore, accurately forecasting these periods of lighter winds is critical for the preservation of human life.

Whilst deciding on an acceptable wind speed threshold for summit attempts is subjective, physical considerations suggest that a human with an effective surface area (A_p) of 0.5 m^2 is at risk of being blown over if the wind force (F) exceeds 72 N (Hugenholtz and VanVeller, 2016; McIlveen, 2002). F is related to the wind speed (v) according to:

$$F = \frac{1}{2}\rho v^2 A_p C_D \tag{1}$$

where ρ is the air density (kg m$^-3$) and C_D is the drag coefficient (dimensionless). Using C_D=0.6 from McIlveen (2002), the critical wind speed (v_c) yielding 72 N can be evaluated:

$$v_c = \sqrt{\frac{144}{(0.3\rho)}} \tag{2}$$

At the altitude of the highest camp (the South Col: 7,945 m a.s.l: Fig 2) on Mt. Everest's main Nepal route – which marks the beginning of the 'death zone' – ρ (which depends on temperature and pressure) is, on average, 0.52 kg m$^-3$, translating to v_c = 30.3 m/s according to data from May 2019 until June 2023 (see data section below).

To illustrate the utility of AtsMOS to deliver improved, decision-critical forecasts we therefore use a new network of Mt. Everest weather stations (see below) to develop predictions of (1) absolute wind speed; and (2) the probability of speeds exceeding both 30 m/s and 20 m/s. The upper threshold is used to identify dangerous winds, whilst the lower we regard as potentially dangerous and hence a conservative threshold for identifying suitable weather for a summit attempt. We also showcase the flexibility of AtsMOS to directly forecast key variables such as windchill temperature and facial frostbite time.

3.2 Mount Everest weather data

In spring 2019, a network of five automated weather stations was installed on the Nepali side of Everest, known locally as Sagarmatha or Qomolangma, including three stations above the basecamp at Camp 2 (6464 m), the South Col (7945 m), and Balcony (8430 m; Matthews et al. (2020a)). Of these, the two highest stations: the 'South Col' (7,945 m a.s.l) and the 'Balcony' (8,430 m a.s.l) were positioned to monitor the potentially dangerous winds on the upper mountain. However, the Balcony's record is relatively short (due to wind damage), and considered unrepresentative of the upper mountain due to sheltering under common flow directions. A further station was installed at 8810 m altitude on the highest elevation exposed bedrock near the summit (the 'Bishop Rock') in Spring 2022, which is currently the highest altitude weather station in the world with publicly available data (Matthews et al., 2022). Note that another station was installed by a Chinese team at a similar altitude on the North side of the mountain in 2022, although its status and data availability are unknown.

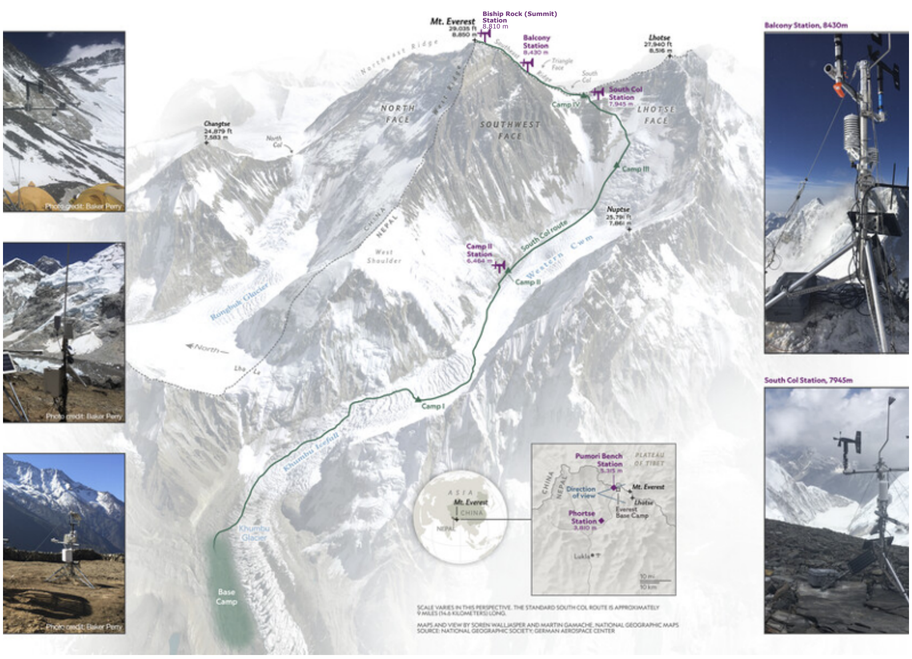

Figure 2. Location of the weather stations on Mt Everest. Modified after (Matthews et al., 2020b)

Three of the four weather stations (Bishop Rock or 'Summit', Balcony, and South Col) were installed with two separate wind speed sensors. The dual-sensors were installed for redundancy in the event of one failing, but are also invaluable for evaluating the reliability of wind speed observations. Recovery of destroyed monitoring equipment by the author team showed

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

that the wind-speed sensors can suffer mechanical failure (breakage of the anemometer cups) and growth of rime-ice that result in incorrect measurements, but that is not evident from a single time series. In the pre-processing stage of AtsMOS, we therefore apply a moving-window cross-correlation between the time-series of the two sensors to reveal periods of decorrelation and unreliable data. We use a minimum correlation threshold of 0.9 measured over a 14-day window for both the mean and maximum hourly wind speed (measured at 5-s intervals) to determine reliable data, and mask out data points falling below this threshold (Fig X top two rows).

Only the South Col station has a data record covering a period longer than a few months and across multiple years. While this station is located almost 1000 m below the summit, its position at the head of the Khumbu Valley with an open westerly aspect (the dominant wind direction) means that its wind speeds are very similar to the summit (Pearson's r-value = 0.85). We apply the dual-sensor correlation threshold (0.9), and filter out winds with a direction outside the range 270 $\pm$ 45 degrees due to the risk of topographic shielding outside this window. The lower elevation leads to a slight negative bias in wind speed at the South Col, which is on average 18% slower than summit winds. We linearly regress the remaining South Col wind record against the filtered summit record and use this to create a synthetic summit record. The resulting record contains just over 1 year of data, spread across two 6-month periods from 06/2019 to 01/2020 and 05/2022 to 01/2023 (Figure 4).

For the NWP component of the AtsMOS loading and pre-processing stage data were loaded and pre-processed from the Global Forecasting System (GFS) (https://rda.ucar.edu/datasets/ds084.1/) We downloaded all 10 variables: precipitation, temperature, relative humidity, N-S wind speed, E-W wind speed, vertical velocity, geopotential height, absolute vorticity, cloud mixing ratio, and ozone mixing ratio. We choose to include all variables (irrespective of whether physical connections to the predictand could be identified a priori) because (i) their variations could provide insight into relevant sub-grid scale process, and (ii) the default machine learning method we select (XGBoost) is robust to overfitting and collinearity. A user-supplied list of variable names can also be supplied to AtsMOS to limit the variables used in model fitting. The data were downloaded for a 9x9x3 data cube centred on the summit of Sagarmatha/Qomolangma, with 9 data points in each horizontal direction (from 27-29 degrees latitude and 86-88 degrees longitude at 0.25 degree spacing) and 3 vertical pressure levels (350, 400, and 450 hPa). We use the geopotential height from the three pressure levels to linearly interpolate or extrapolate all variables to a fourth vertical level, corresponding to the summit elevation at 8849 m. Finally, we calculate the horizontal and vertical gradients in the 9 variables, to further account for potential drivers of relevant sub-grid scale processes. A full list of all 172 variables and derivatives used is in the supplement. We separately download the GFS historical archive (via the NCAR web portal) and real-time GFS forecast (programmatically from the THREDDS server - see notebook), with the former used to calibrate our data corrections and the latter used to produce corrected forecasts.

For the core processing component of AtsMOS we use (simple) linear regression and (complex) XGBoost algorithms to improve the GFS forecast for the wind speed at the summit of Sagarmatha/Qomolangma. To avoid issues with temporal autocorrelation of training and validation data, we split our time series in half in January 2021. This test-train split provides us with 6 months of training data and 6 months of validation data from 06/2019 to 01/2020 and 05/2022 to 01/2023. We run each MOS algorithm twice, once training on data from 2019 and testing on data from 2022 and vice versa. Linear regression is applied using just the GFS model wind speed interpolated to the 8849 m summit altitude as the only predictor variable;

XGBoost, on the other hand, is trained using all 172 GFS variables and spatial derivatives. We reproduce the simple and complex MOS workflows for several GFS lead times: analysis (0 h nowcast), 24 h, 48 h, 120 h (5 day), and 240 h (10 days).

While predicting Sagarmatha/Qomolangma wind speed as a continuous variable is scientifically interesting, a categorical prediction of dangerous versus safe winds may be of more use to the majority of potential end users (Sherpas and mountaineers). We therefore employ a wind speed threshold of 30 m.s-1 to classify our synthetic wind time series into a time series of dangerous winds. We also use a second, lower threshold of 20 m.s-1 to classify potentially dangerous winds. A wind speed of 30 m.s-1 corresponds approximately to the wind speed required to blow a human off their feet at Sagarmatha/Qomolangma summit conditions (Section 3.1) We intentionally do not call winds below this threshold 'safe' as they can still be hazardous in a range of ways (including slowing ascents and increasing exposure), but they correspond to conditions during which – at least in principle – a typical climber should not be in danger of being blown from the mountain. We use the same XGBoost MOS to run the categorical forecast, using GFS lead times of 0 h (analysis), 48 h, and 240 h (10 days). For the 0 h and 48 h lead times we forecast dangerous winds at the native GFS 6 h temporal resolution. For the 240 h (10 days) lead time, however, we inverse the problem and classify 48 h (2 days) periods during which all winds are below the given threshold. The objective of classifying low-wind periods with a 10-day lead time is to enable earlier identification of favourable summit weather conditions and a better distribution of climbs throughout the season to prevent potentially dangerous overcrowding.

3.3 Results and model evaluation

We test the AtsMOS dataset by training it on data from the first period (2019-2020) and predicting data over the second period (2021-2022) and vice versa. This enables a more robust validation than random test-train splitting of the dataset, by reducing inflation of model ability caused by meteorological time series temporal autocorrelation. We evaluate three different learning techniques: simple linear regression, Random Forest, and XGBoost (Figures 5, 6, 7).

Linear regression produces a reasonable overall fit to the data (Figure 5), with a model wind speed-field measured wind speed R2 of 0.87, a root mean squared error (RMSE) of 10.59 m.s-1, and a mean absolute error (MAE) of 7.87 m.s-1. The Kling-Gupta efficiency of these datasets is 0.73, evaluating a combination of their correlation, relative variation, and mean bias and with higher values reflecting a better fit (Gupta et al., 2009). In particular, linear regression successfully matches the magnitude of winds during the majority of the low-wind (monsoon) season from July to October. However, it fails to match the magnitude of the highest wind-speed events, with a clear overestimate evident.

Random Forest produces a good overall fit to the data, with a model wind speed-field measured wind speed R2 of 0.92 and a root mean squared error (RMSE) of 8.52 m.s-1, and a mean absolute error (MAE) of 6.33 m.s-1. The Kling-Gupta efficiency of these datasets is 0.77. There are three notable improvements of the model trained with Random Forest regression relative to standard linear regression: the estimated are more closely clustered along the 1:1 model-data line, the timings of high-wind episodes in the model better match those observed in the data, and the magnitude of high-wind peaks better matches across both datasets – although a small bias towards higher winds than reality remains.

XGBoost produces a good fit to the data, with a model wind speed-field measured wind speed R2 of 0.93 and a root mean squared error (RMSE) of 7.95 m.s-1, and a mean absolute error (MAE) of 5.97 m.s-1. The Kling-Gupta efficiency of these

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

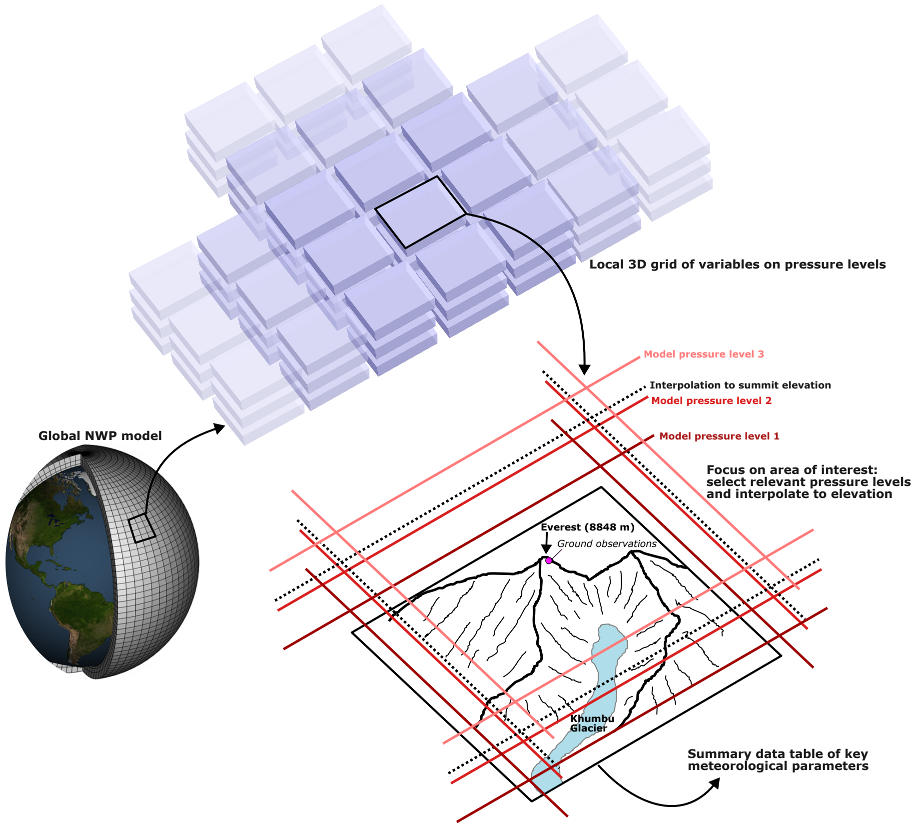

Figure 3. Module for extracting key meteorological parameters from the global Numerical Weather Prediction (NWP) model around a point of interest (typically, the location of the ground observations). Both horizontal and vertical derivatives are calculated from the NWP data to supplement the ML training dataset.

datasets is 0.79. The overall performance of the model trained with XGBoost is similar to that trained with Random Forest, with a slightly improved fit across all metrics. The timing of high-wind events is well predicted and, while the model still tends to overestimate the magnitude of high-wind events, the bias is lower (bias score: 0.86 for XGBoost, relative to 0.84 for Random Forest and 0.81 for linear regression).

We then apply the AtsMOS workflow on a real-time case study for the approximately two-week (384-hour) period from 20 July 2023 to 05 September 2023 using GFS forecast data as described in the methods. As well as calculating the wind speed, temperature, and precipitation, we compute forecasts of two derivative variables: wind chill temperature and facial frostbite time (Moore and Semple, 2011). Both wind chill temperature and facial frostbite time are calculated based on wind speed and temperature forecasts according to the formulas of Moore and Semple (2011). Wind chill temperature reaches as low as -45 degrees celcius on 03/09/2023, also aligning with the shortest facial frostbite time of less than 7 minutes (Figure 8). Forecast wind speeds do not exceed 20 m.s-1, but reach more than 19 m.s-1 on the night of 20-21 July 2023, with the short forecast lead reducing the uncertainty for the forecast. The facial frostbit time briefly falls below 10 minutes this night also driven the the

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

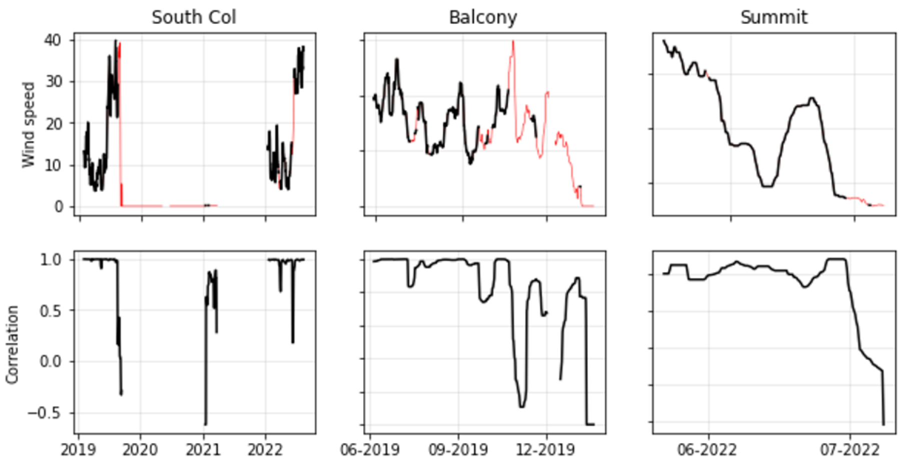

Figure 4. Validation of observational time series.

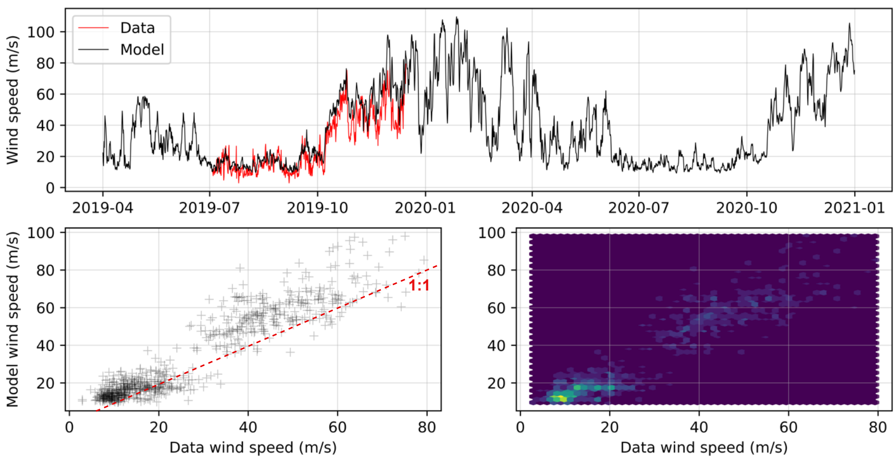

Figure 5. Observed (Red) and modelled (black) wind speed for the first observational period at the summit, with model (here, linear regression) training using only the second period (2019).

high wind speeds, and the wind chill temperature fluctuates between -35 and -40 C – well below the air temperature (-20C), highlighting the importance of the wind speed in modulating the cold hazard, and thereby the value of computing this derived variable with AtsMOS.

4 Discussion and broader applicability

The AtsMOS workflow builds on advancements in machine learning and data accessibility to improve mountain weather forecasts by downscaling coarse numerical model outputs to specific locations of high value. Through a case study focusing on

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development Discussions
Open Access
EGU

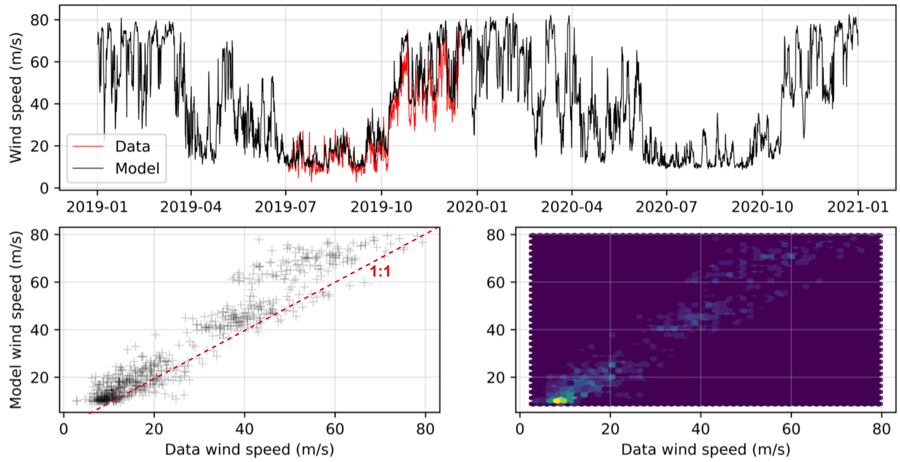

Figure 6. Observed (Red) and modelled (black) wind speed for the first observational period at the summit, with model (here, Random Forest) training using only the second period (2019).

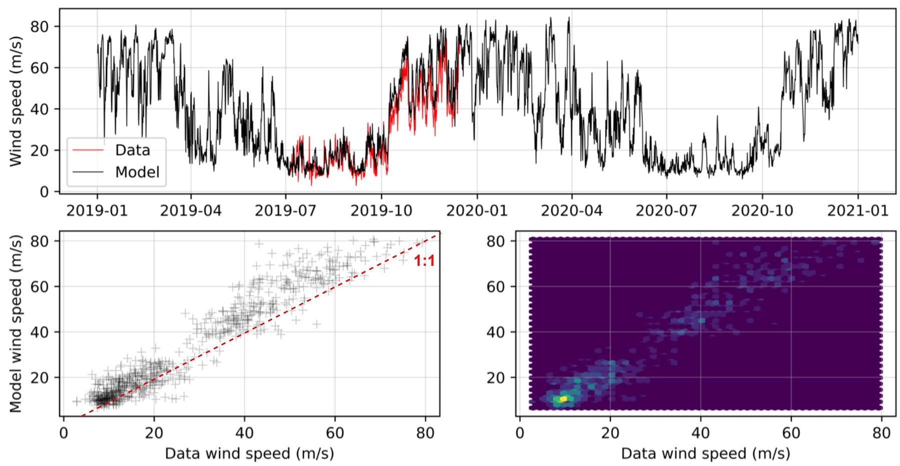

Figure 7. Observed (Red) and modelled (black) wind speed for the first observational period at the summit, with model (here, XGBoost) training using only the second period (2019).

Mt. Everest summit meteorology, we demonstrate the effectiveness of AtsMOS in refining wind speed (and wind chill temperatures) critical for assessing risks for mountaineering. This workflow is open-source, extremely flexible, and computationally cheap – features which should enable the accuracy of mountain weather predictions to be improved across many different environments.

The results of our test at Mt. Everest showcase a local application of the AtsMOS workflow. Mt. Everest represents a specific environment in which mountaineers and Sherpas expose themselves to potentially deadly conditions (Moore and Semple,

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development
Discussions
Open Access
EGU

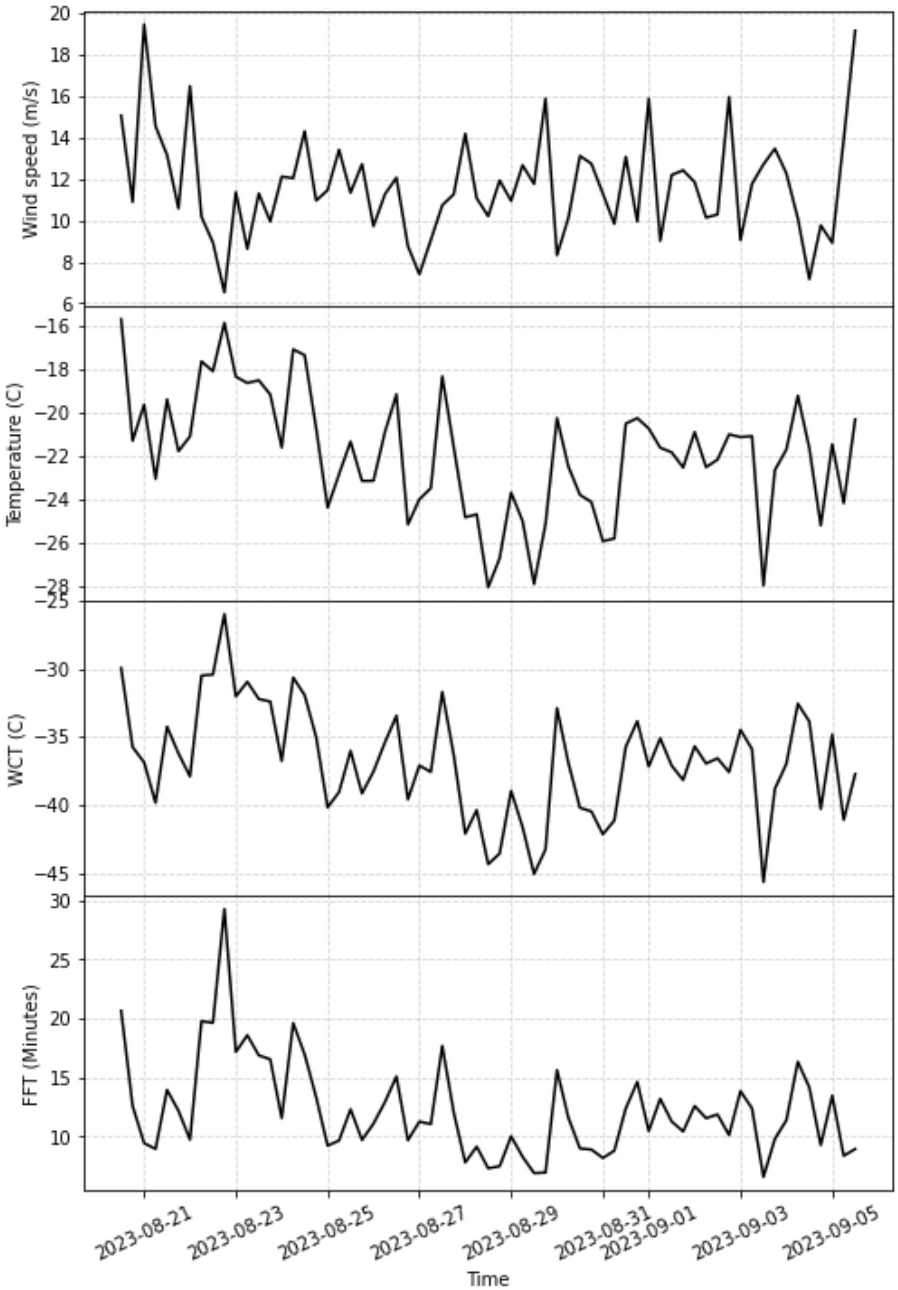

Figure 8. Example of real-time forecast for wind, temperature, and precipitation as well as derivate variables of wind chill temperature (WCT) and facial frostbite time (FFT; Moore and Semple, 2011).

2006, 2011; Matthews et al., 2020b). More precise meteorological forecasts are therefore critical for expedition planning in two significant ways. Firstly, by more accurately predicting wind speeds, our system enables expedition organizers to identify windows of potentially 'safe' (lower wind) conditions with approximately two weeks' notice, allowing the timing of trips to the upper mountain to be determined at an earlier date. This is invaluable for optimizing expedition scheduling, maximizing the likelihood of successful summit attempts, and potentially improving safety by preventing dangerous overcrowding from teams rushing to exploit weather windows at late notice. Secondly, the shorter lead time, and more precise AtsMOS forecasts assist in preventing climbs during times of dangerous weather, thereby enhancing safety. By providing reliable forecasts for both dangerous and potentially dangerous wind thresholds, our workflow empowers expedition leaders to make informed decisions, avoiding ascent attempts during periods of heightened risk.

The flexible nature of the workflow enables outputs with different levels of complexity, ranging from binary classifications ('dangerous/safe'), raw meteorological variable forecasts (wind speed, temperature, etc.), and derivative variables (e.g. facial frostbite time). This flexibility offers a wide range of possibilities to enable expedition planners who, armed with more information, should be able to plan safer climbs, thereby reducing the risk of attempting this iconic mountain. The evaluation of hazard probability with AtsMOS is seen as a particularly important feature for end-users. If properly calibrated, it more clearly aligns the forecast product with decision-making. Without the MOS approach here, expedition planners would likely need to consult ensemble forecasts (e.g., the Global Ensemble Forecasting System) to produce comparable probabilities, associated with a non-trivial increase in data processing for support teams; and/or burden on the expedition planner to interpret the forecast. Of course, we also note here for AtsMOS to be used within such ensemble forecasting systems – for example, propagating the ensemble members through the ML algorithms calibrated on the deterministic forecast to more fully explore uncertainty. This can ultimately combine the benefits of both the ensemble forecast and reduced bias from local calibration.

While the AtsMOS workflow's potential to improve local mountain meteorology forecasts is promising, it is important to acknowledge its limitations. The most significant constraint lies in the workflow's dependence on two separate data sources: numerical weather prediction data and field data. AtsMOS outputs therefore rely on the fundamental assumption that, while these datasets may contain uncertainties or noise, they also contain real information about local meteorological conditions. There are a number of scenarios in which this may not be the case for either dataset, for instance, large-scale NWP models missing key local processes (leading for example, to a poor representation of convection), or sensors may become degraded and record false data (for instance, a wind sensor covered in rime-ice). This limitation is present at both the training and prediction stages of the process. The effectiveness of the workflow is, therefore, highly dependent upon the availability and quality of ground observations, which are particularly rare in remote and high-altitude regions like Mt. Everest (Matthews et al., 2020a; Thornton et al., 2022). The applicability of AtsMOS may also be limited in regions with unique or extreme meteorological conditions not adequately captured by existing NWP models – even with the aid of machine learning to extract additional information. – which may be of concern if these regions are of particular interest for hazard mitigation. We note, however, that the latter may be guarded against by using near real-time (i.e., lagged) observations from the telemetry-enabled weather stations (Chkeir et al., 2023).

We also highlight caution in the application of machine learning algorithms. Whilst techniques like Random Forest and XGBoost can offer enhanced predictive capabilities, they may also introduce complexities in model interpretation and require careful validation to ensure robust performance. These limitations underscore the need for ongoing refinement and validation

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

of the workflow to optimize its utility and effectiveness in diverse mountainous environments. One specific concern in the usage of tree-based machine learning algorithms such as Random Forest or XGBoost is that they cannot reasonably extrapolate beyond the range captured in the training data. This is a particular concern in areas with strong seasonal variation, where training on one season alone may lead to failure to produce meaningful predictions in the other season. In the case of Everest, this limitation is mitigated by having data covering the transition from low to high wind season, but in areas where this is not possible alternative methods may need to be considered.

Another type of 'overfitting' may occur if machine learning inadvertently reproduces biases in the observations, for example due to instrumentation errors. This challenge should be taken seriously, as the error could be systematic and dangerous. For example, if icing of wind sensors occurred preferentially in conditions of low temperature and high winds (i.e., periods of greater cold stress), the machine learning, trained on the errors, would underestimate the hazard most when it was greatest. Such risks highlight the importance of thoroughly quality assuring the observations in the pre-processing stage of the AtsMOS workflow. We note that, on Mt. Everest, the station design enables the detection of such icing through the use of redundant wind sensors (Matthews et al., 2020a, 2022). We hope that ongoing efforts to develop a Universal High Altitude Observing Platform (to enhance mountain weather monitoring worldwide) also be designed with such challenges in mind (Napoli et al., 2023). More generally, we emphasise that the AtsMOS approach to forecast improvement differs from efforts to embed ML in NWP (e.g. Frnda et al., 2022). In this case, ML algorithms do not replace high-quality observational data; rather they emphasise the need for it and amplify any limitations of the data. By investing in data quality and instrumentation and leveraging ML alongside this, we increase our potential for accurate and actionable meteorological forecasts in mountainous regions.

In addition to ensuring the accuracy and reliability of sensor data, effective data management practices are crucial for maximizing the utility and impact of field datasets, particularly in the context of mountain meteorology. Good metadata, which provides detailed information about the characteristics and origins of the data, is essential for understanding and interpreting observational datasets. Interoperability, where data can be integrated and exchanged across different platforms and systems with minimal barriers, becomes increasingly important when considering the generalizability of findings and methodologies to other mountainous environments. While the specific challenges and characteristics of each mountain region may vary, the fundamental principles and approaches developed for mountain meteorology research can often be applied more broadly and insights and techniques developed in one region can inform and benefit studies in others. Promoting robust data management practices is key for both the effectiveness of individual research efforts and the broader advancement of mountain meteorology as a field.

Whilst we have demonstrated the added value of improving weather forecasts for Mt. Everest with AtsMO, we anticipate much greater benefits from this approach than just improving the safety of mountaineering expeditions. For instance, the ability to forecast thresholds for rainfall-triggered landslides, snow avalanches, or flooding relies heavily on accurate meteorological data and predictive models. By integrating high-resolution local meteorological data from AtsMOS into early warning systems, communities can better prepare for and respond to extreme weather events, reducing the risk of casualties and damage. Furthermore, on a regional or national scale, the integration of detailed mountain meteorology datasets into larger-scale networks enhances the effectiveness of early warning systems by providing comprehensive coverage of weather patterns and potential

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development Discussions
Open Access
EGU

hazards across diverse landscapes. Improved prediction of meteorological conditions in mountainous regions has far-reaching implications for promoting the resilience and safety of mountain communities and ecosystems and is an important component of effective early warning systems for many hazards.

5 Conclusion

In conclusion, the AtsMOS workflow represents a computationally efficient template for downscaling numerical model outputs using one or a small number of field observations. The template outlines a flexible, modular workflow, for custom preprocessing of field observations or numerical weather model outputs depending on the need, and provides several possible core learning algorithms ranging from simple linear regression to more complex Random Forest and XGBoost. We explore an example application at Mt. Everest, which demonstrates its practical utility in improving the prediction of critical weather parameters for mountaineering safety. There are limitations to this approach, including reliance on high-quality sensor data and potential biases inheritance in machine learning algorithms. Moving forward, continued research and observation network development hold promise for improving the accuracy and reliability of mountain meteorology forecasts, ultimately enhancing hazard mitigation efforts, and contributing to the resilience of communities living in these landscapes.

Code and data availability. The AtsMOS workflow is available at https://github.com/MaxVWDV/AtsMOS or https://zenodo.org/doi/10.5281/zenodo.10889509 (Van Wyk de Vries, 2024). Everest weather station data is available at https://www.nationalgeographic.org/society/everest-weather-data/.

Author contributions. MV and TM conceived the project and conducted the analyses with input from all authors. All authors commented on the analyses and final manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements.

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development
Discussions

References

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, https://doi.org/10.1038/nature14956, number: 7567 Publisher: Nature Publishing Group, 2015.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pp. 785–794, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/2939672.2939785, 2016.

Chkeir, S., Anesiadou, A., Mascitelli, A., and Biondi, R.: Nowcasting extreme rain and extreme wind speed with machine learning techniques applied to different input datasets, Atmospheric Research, 282, 106 548, https://doi.org/10.1016/j.atmosres.2022.106548, 2023.

Corbari, C., Ravazzani, G., Perotto, A., Lanzingher, G., Lombardi, G., Quadrio, M., Mancini, M., and Salerno, R.: Weekly Monitoring and Forecasting of Hydropower Production Coupling Meteo-Hydrological Modeling with Ground and Satellite Data in the Italian Alps, Hydrology, 9, 29, https://doi.org/10.3390/hydrology9020029, number: 2 Publisher: Multidisciplinary Digital Publishing Institute, 2022.

Ellis-Petersen, H.: Climate change to blame for up to 17 deaths on Mount Everest, experts say, The Guardian, https://www.theguardian.com/world/2023/may/30/climate-change-to-blame-for-up-to-17-deaths-on-mount-everest-experts-say, 2023.

Firth, P. G., Zheng, H., Windsor, J. S., Sutherland, A. I., Imray, C. H., Moore, G. W. K., Semple, J. L., Roach, R. C., and Salisbury, R. A.: Mortality on Mount Everest, 1921-2006: descriptive study, BMJ, 337, a2654, https://doi.org/10.1136/bmj.a2654, publisher: British Medical Journal Publishing Group Section: Research, 2008.

Frnda, J., Durica, M., Rozhon, J., Vojtekova, M., Nedoma, J., and Martinek, R.: ECMWF short-term prediction accuracy improvement by deep learning, Scientific Reports, 12, 7898, https://doi.org/10.1038/s41598-022-11936-9, number: 1 Publisher: Nature Publishing Group, 2022.

Glahn, B.: Determining an Optimal Decay Factor for Bias-Correcting MOS Temperature and Dewpoint Forecasts, Weather and Forecasting, 29, 1076–1090, https://doi.org/10.1175/WAF-D-13-00123.1, publisher: American Meteorological Society Section: Weather and Forecasting, 2014.

Glahn, H. R. and Lowry, D. A.: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting, Journal of Applied Meteorology and Climatology, 11, 1203–1211, https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2, publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, 1972.

Grönquist, P., Yao, C., Ben-Nun, T., Dryden, N., Dueben, P., Li, S., and Hoefler, T.: Deep learning for post-processing ensemble weather forecasts, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379, 20200 092, https://doi.org/10.1098/rsta.2020.0092, publisher: Royal Society, 2021.

Han, L., Chen, M., Chen, K., Chen, H., Zhang, Y., Lu, B., Song, L., and Qin, R.: A Deep Learning Method for Bias Correction of ECMWF 24–240 h Forecasts, Advances in Atmospheric Sciences, 38, 1444–1459, https://doi.org/10.1007/s00376-021-0215-y, 2021.

Herman, G. R. and Schumacher, R. S.: "Dendrology" in Numerical Weather Prediction: What Random Forests and Logistic Regression Tell Us about Forecasting Extreme Precipitation, Monthly Weather Review, 146, 1785–1812, https://doi.org/10.1175/MWR-D-17-0307.1, publisher: American Meteorological Society Section: Monthly Weather Review, 2018.

Huey, R. B., Carroll, C., Salisbury, R., and Wang, J.-L.: Mountaineers on Mount Everest: Effects of age, sex, experience, and crowding on rates of success and death, PLOS ONE, 15, e0236 919, https://doi.org/10.1371/journal.pone.0236919, publisher: Public Library of Science, 2020.

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development
Discussions

Hugenholtz, C. H. and VanVeller, G. S.: Wind hazard in the alpine zone: a case study in Alberta, Canada, Weather, 71, 27–31, https://doi.org/10.1002/wea.2567, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wea.2567, 2016.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., Nepal, S., Pacheco, P., Painter, T. H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A. B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., and Baillie, J. E. M.: Importance and vulnerability of the world's water towers, Nature, 577, 364–369, https://doi.org/10.1038/s41586-019-1822-y, number: 7790 Publisher: Nature Publishing Group, 2020.

Khadka, A., Wagnon, P., Brun, F., Shrestha, D., Lejeune, Y., and Arnaud, Y.: Evaluation of ERA5-Land and HARv2 Reanalysis Data at High Elevation in the Upper Dudh Koshi Basin (Everest Region, Nepal), Journal of Applied Meteorology and Climatology, 61, 931–954, https://doi.org/10.1175/JAMC-D-21-0091.1, publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, 2022.

Lagerquist, R., McGovern, A., and Smith, T.: Machine Learning for Real-Time Prediction of Damaging Straight-Line Convective Wind, Weather and Forecasting, 32, 2175–2193, https://doi.org/10.1175/WAF-D-17-0038.1, publisher: American Meteorological Society Section: Weather and Forecasting, 2017.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, https://doi.org/10.1126/science.adi2336, publisher: American Association for the Advancement of Science, 2023.

Mass, C. F., Baars, J., Wedam, G., Grimit, E., and Steed, R.: Removal of Systematic Model Bias on a Model Grid, Weather and Forecasting, 23, 438–459, https://doi.org/10.1175/2007WAF2006117.1, publisher: American Meteorological Society Section: Weather and Forecasting, 2008.

Matthews, T., Perry, L. B., Koch, I., Aryal, D., Khadka, A., Shrestha, D., Abernathy, K., Elmore, A. C., Seimon, A., Tait, A., Elvin, S., Tuladhar, S., Baidya, S. K., Potocki, M., Birkel, S. D., Kang, S., Sherpa, T. C., Gajurel, A., and Mayewski, P. A.: Going to Extremes: Installing the World's Highest Weather Stations on Mount Everest, Bulletin of the American Meteorological Society, 101, E1870–E1890, https://doi.org/10.1175/BAMS-D-19-0198.1, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2020a.

Matthews, T., Perry, L. B., Lane, T. P., Elmore, A. C., Khadka, A., Aryal, D., Shrestha, D., Tuladhar, S., Baidya, S. K., Gajurel, A., Potocki, M., and Mayewski, P. A.: Into Thick(er) Air? Oxygen Availability at Humans' Physiological Frontier on Mount Everest, iScience, 23, 101 718, https://doi.org/10.1016/j.isci.2020.101718, 2020b.

Matthews, T., Perry, B., Khadka, A., Sherpa, T. G., Shrestha, D., Aryal, D., Tuldahar, S., Thapa, N., Pradhananga, N., Athans, P., Sherpa, D. Y., Guy, H., Seimon, A., Elmore, A., Li, K., and Alexiev, N.: Weather Observations Reach the Summit of Mount Everest, Bulletin of the American Meteorological Society, 103, E2827–E2835, https://doi.org/10.1175/BAMS-D-22-0120.1, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2022.

McIlveen, J. F. R.: The everyday effects of wind drag on people, Weather, 57, 410–413, https://doi.org/10.1256/wea.29.02, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1256/wea.29.02, 2002.

Minder, J. R., Mote, P. W., and Lundquist, J. D.: Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD013493, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009JD013493, 2010.

https://doi.org/10.5194/gmd-2024-36
Preprint. Discussion started: 3 April 2024

Geoscientific Model Development Discussions
Open Access EGU

Miner, K. R., Mayewski, P. A., Baidya, S. K., Broad, K., Clifford, H., Elmore, A., Gajurel, A. P., Giri, B., Guilford, S., Hubbard, M., Jaskolski, C., Koldewey, H., Li, W., Matthews, T., Napper, I., Perry, L. B., Potocki, M., Priscu, J. C., Tait, A., Thompson, R., and Tuladhar, S.: An Overview of Physical Risks in the Mt. Everest Region, One Earth, 3, 547–550, https://doi.org/10.1016/j.oneear.2020.10.008, 2020.

Moore, G. and Semple, J.: Freezing and Frostbite on Mount Everest: New Insights into Wind Chill and Freezing Times at Extreme Altitude, High Altitude Medicine & Biology, 12, 271–275, https://doi.org/10.1089/ham.2011.0008, publisher: Mary Ann Liebert, Inc., publishers, 2011.

Moore, G. W. K. and Semple, J. L.: Weather And Death On Mount Everest: An Analysis Of The Into Thin Air Storm, Bulletin of the American Meteorological Society, 87, 465–480, https://doi.org/10.1175/BAMS-87-4-465, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2006.

Napoli, A., Pepin, N., Palazzi, E., and Zardi, D.: A Workshop on Advances in Our Understanding of Elevation Dependent Climate Change, Bulletin of the American Meteorological Society, 104, E928–E934, https://doi.org/10.1175/BAMS-D-23-0043.1, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2023.

Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., and Thuerey, N.: WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 203, https://doi.org/10.1029/2020MS002203, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002203, 2020.

Thornton, J. M., Pepin, N., Shahgedanova, M., and Adler, C.: Coverage of In Situ Climatological Observations in the World's Mountains, Frontiers in Climate, 4, https://www.frontiersin.org/articles/10.3389/fclim.2022.814181, 2022.

Van Wyk de Vries, M.: AtsMOS: AtsMOS 1.0, https://doi.org/10.5281/zenodo.10889510, 2024.

Zhang, G., Zhu, S., Zhang, N., Zhang, G., and Xu, Y.: Downscaling Hourly Air Temperature of WRF Simulations Over Complex Topography: A Case Study of Chongli District in Hebei Province, China, Journal of Geophysical Research: Atmospheres, 127, e2021JD035 542, https://doi.org/10.1029/2021JD035542, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021JD035542, 2022.