

Response to referee #3 (report #1)

General comments

I was not involved in the first round of review and have read the entire paper from scratch, after which I read the answer to reviewer comments. Overall, I feel the paper is definitely suitable for publication in HESS and is quite relevant for hydrological applications in alpine and glaciated catchments. However, the paper is too long, overly descriptive, and lacks a sound research question. Even though it's a technical note, it tries to address too many questions: (1) role of statistical calculations from dynamic downscaling models to improve model outputs, (2) method to remove drizzle effect from precipitation simulations, (3) using WRF simulations when lacking high-quality long-term meteorological measurements for statistical bias correction, (4) role of temperature lapse rates on humidity and its consequent impact on evapotranspiration and snowmelt simulations, (5) physics-based hydrological modeling, etc. In my understanding, a technical note should be a concise research article with one or two main points. Even in the first review round, this was pointed out, and somehow the length didn't undergo any significant change. Some points were removed, but then relative humidity calculations were added (which is good, because it results in slightly better hydrological simulations). However, the primary problem around conciseness was not addressed, and this should be accounted for in any subsequent draft.

Thank you for taking the time to evaluate the manuscript and response letters comprehensively. We are pleased to hear that the reviewer recognizes and supports the key aspects of this technical note and their significance for the hydrological community. We understand the fundamental criticism that the technical note needs to focus specifically on one or two key points and is currently too lengthy. Due to the complexity of the methods used, their evaluation, and the additional analyses requested by the reviewers in the first round, the manuscript has actually become even longer. In the revised version, we have taken this core concern into account and addressed it.

Major comments:

1. The title includes "land cover representation in WRF" as a key message. However, from my reading of the manuscript, the article calculates temperature lapse rates from different geopotential heights from WRF outputs, which doesn't even account for land cover representations at high resolutions. The paper doesn't systematically quantify how different land cover representations can influence calculations of temperature lapse rates. Instead, the article averages temperature lapse rates over different WRF land cover classes (as explained in P19 of your responses to reviewer 2 comments). The exact response is "We therefore calculate the mean over the entire research area, so that the eta levels are less dependent on the WRF land cover types". The article instead shows the hydrological skill gained due to an additional bias correction step using outputs from a dynamic downscaling model (WRF in this case). This should be clearly reflected in the title, as the paper doesn't carry out different variants of WRF simulations based on different land cover representations at high resolution. It uses static WRF outputs to run a hydrological model and doesn't quantify "land cover representation in WRF" on "hydrological modeling" outputs as promised in the Title.

We agree on this point and have therefore rephrased the title accordingly. The new title is:

"A physically based correction of WRF temperature using atmospheric lapse rates for high-resolution hydrological simulations in a glacierized catchment"

2. The paper has 11 figures, 13 supplementary figures, 3 tables, and 5 supplementary tables. The length of the main article is ~800 lines. In my conceptualization, a technical note is supposed to be a condensed article with a few figures revolving around 1 or 2 main points. This paper is overly long, descriptive, and doesn't read as a technical note. Additionally, the scientific questions being addressed here are not concisely mentioned in the Abstract.

We agree and have condensed the manuscript, focusing primarily on the temperature correction method and its benefits for hydrological simulations in snow- and glacier-dominated catchments. We moved the sections on the correction of precipitation and relative humidity to the supplementary material. We have shortened the main text of the manuscript by 95 lines. Together with the moving of tables and figures to the supplementary material, the revised manuscript is now six pages shorter.

We have comprehensively revised the Abstract, now focusing on the developed post-processing workflow that derives hourly temperature lapse rates directly from the WRF atmosphere using temperature and geopotential height fields from multiple eta levels.

3. The discussions are overly focused on statistical vs dynamical downscaling approaches. It seems (from Discussion) that the primary objective of this manuscript was to characterize the skills gained via a dynamic downscaling technique as statistical downscaling requires long and high-quality meteorological timeseries, this was not at all clear to me from the Abstract, which reads more as a study to improve WRF model simulations. The main objective of this paper is defined as (L74-75): "The objective of this technical note is to demonstrate the variability and efficient correction of WRF near-surface temperature by spatially averaging hourly temperature and geopotential height at each WRF pressure level". The main research question reads more as a "Method", rather than a question. The authors should reframe the question as reducing biases in meteorological forcings by using the internal elevational gradients in air temperature from convection-permitting atmosphere-land surface models and its evaluation over a high-elevation glacier-dominated alpine catchment.

We rephrased the Abstract to avoid the misleading impression that we are improving the WRF simulations. We highlight that the developed workflow for correcting the WRF temperature is a post-processing step. The rephrased sentences in the Abstract (lines 8ff) are as follows:

"To address this limitation, we developed a novel post-processing workflow that derives hourly temperature lapse rates directly from the WRF atmosphere using temperature and geopotential height fields from multiple eta levels. These lapse rates are subsequently used to reconstruct near-surface temperatures from the first atmospheric model level, thereby reducing the influence of surface-related biases."

4. Please specify in the Introduction why only humidity was corrected using temperature lapse rates, and not shortwave incoming radiation and wind speed. Were there no lapse rates in solar radiation? From the point of physics-based snow modeling, radiation-induced melt on different slopes of a mountain and wind-based snow redistribution would be equally important (see L306-309) for snow ablation dynamics in steep alpine catchments. I am not suggesting doing a multivariate bias correction, as that would be beyond the scope of the article and make the paper even longer, but to at least justify the reason for only choosing humidity for additional correction (and not radiation / wind). Additionally, in L327 and L625-626, it's briefly mentioned that shortwave incoming radiation was adjusted according to topographic shading, what were the adjustments made? This should be described in details.

Yes, we agree on the general importance of shortwave radiation in computing the energy balance of a snowpack. We have added this aspect to the introduction, explaining why we have not considered it for post-processing correction:

“The objective of this technical note is to develop and evaluate a physically based post-processing workflow for reducing biases in coarse-resolution regional climate model output used to force high-resolution hydrological models in high-elevation, glacierized catchments. We focus primarily on near-surface air temperature, as snow and glacier processes are highly sensitive to its spatiotemporal variability. Rather than relying on empirical bias-correction techniques, we derive hourly temperature lapse rates directly from the free WRF atmosphere and use them to reconstruct near-surface temperatures from the first atmospheric model level. This approach aims to reduce biases arising from the coarse representation of snow- and ice-covered surfaces and associated land–atmosphere interactions in convection-permitting WRF simulations.

In addition, we apply a sequential, thermodynamically consistent correction of WRF relative humidity based on the corrected temperature fields. This step is particularly relevant for energy-balance-based snow and glacier melt modeling, where latent heat fluxes contribute substantially to the surface energy balance. In contrast, no post-processing corrections are applied to shortwave radiation or wind speed. Physics-based hydrological models such as SNOWPACK (Lehning et al., 2002), openAmundsen (Strasser et al., 2024), and WaSiM (Schulla, 2021) already account for scale-dependent topographic effects on radiation and, where applicable, include representations of wind-driven snow redistribution “

We extended the description of the topographic correction approach for shortwave radiation in WaSiM as follows (lines 314ff):

“Incoming shortwave radiation was adjusted according to slope, aspect, and terrain shading through solar-geometry calculations, and air temperature was subsequently modified using an empirical relationship linking local temperature deviations to the radiation correction under prevailing sunshine conditions.”

5. Section 4.3: There is negligible connection of this section with what was actually done in this paper. The first paragraph (L667-676) describes the Tomasi et al. (2017) paper. There is no actual discussion, which connects results from this study to the results obtained by Tomasi et al. (2017). This section can be easily removed!

We agree and removed this section from the Discussion.

6. L216-217: “On the contrary, eta level 20 does not show any influence of the surface properties (Fig. 2d) throughout the season (Fig. S3d)”. I don’t get what a reader should be looking at in Fig 2d that should tell them that eta level 20 does not show influence of surface properties as the temperature range is still 3.5-4C. Based on this figure, one could argue that Fig. 2e, which probably corresponds to level 30, definitely does not show influence of surface topography as temperature range is <0.5C. Also, looking at Figure 4, one could argue that lapse rate calculations might have changed significantly if a different level (e.g., Level 27) was chosen. Please justify the choice of choosing level 20 for calculating lapse rates. In L415-417, levels 1-20 have been shown as those resulting in the lowest errors, but a brief glance at Table S2 rejects this. Errors are even zero at ETA 25 (why?). This was a point also raised by reviewer #2 in the last round, and has clearly not been addressed. The choice of levels 1-20 seem more rudimentary than with any proper statistical backing.

Fig. 2d illustrates a very strong linear relationship between elevation and temperature at eta level 20. At this height above ground (3278 m), we observe a horizontally uniform lapse rate structure, as temperature is controlled almost entirely by altitude rather than by land and snow cover. The key aspect is the scattering around the regression line that is caused by surface properties (land cover and snow cover). We rephrased this sentence (lines 240ff).

“On the contrary, eta level 20 does not show any scattering around the regression line (Fig. 2d) or any season dependencies (Fig. S3d) that is caused by surface properties (i.e., land cover and snow cover), which indicates a horizontally uniform lapse rate structure in the domain.”

We agree on the misleading formulation and have revised the sentences in a more robust and scientifically sound argumentation (lines 400ff):

“From the evaluation of goodness-of-fit criteria (Tab. S3), we conclude that the lowest atmospheric levels (1-2) perform poorly, suggesting a strong influence of local surface and boundary-layer processes. Although individual eta level combinations yielded slightly better performance at specific stations, the optimal eta levels varied substantially among sites, indicating limited spatial transferability. Therefore, a bulk lapse rate derived from eta levels 1–20 was adopted because it represents the lower-tropospheric thermal structure in a physically consistent manner, minimizes dependence on station-specific calibration, and provides a transferable framework for temperature extrapolation across catchments with differing topographic and climatic characteristics.”

Since the calculation of the goodness-of-fit criteria in Tab.S3 was always performed for two adjacent eta levels up to level 25, it was not possible to calculate the values for level pair 25 to 26. We have therefore removed level 25 from the table, as the values of 0 are indeed confusing.

7. There are too many abbreviations in the paper. An example is HM standing for hydrological model, I don't think there is much to be gained from such abbreviations (one less word in this case). I have pointed some examples below:

- L11: “... TLRs and the HM results are significantly dependent on the coarse surface properties of WRF.”
- L14: “.. when the HM is forced by TLRs from the WRF free atmosphere ..”

We followed the suggestion and removed the abbreviation HM in the revised manuscript.

- L123: “SH” is the abbreviation for snow depth, sorry this is unacceptable. Call it snow height and don't use abbreviations for two words.

We removed this incorrect abbreviation in the revised manuscript.

- L526: “ETR” as an abbreviation for evapotranspiration makes little sense. Its still only one word! Even if you wanted to use an abbreviation, it should be ET which is more typically used in hydrological literature

We agree and removed the abbreviation (etr) for evapotranspiration in the revised manuscript. Indeed, this abbreviation originates from the WaSiM output for real evapotranspiration, as it differentiates between potential and real evapotranspiration.

Other points

L4: explicitly mention from which temporal and spatial resolution 20CRv3 was downscaled from

We added the spatial and temporal resolution of 20CRv3 (i.e., $1^\circ \times 1^\circ$ and 3 h).

L5: in the previous sentence, 2 km resolution was mentioned, now 25 m resolution is mentioned. It's confusing which spatial resolution was downscaling performed to (25 m or 2km). Additionally, actual dynamic downscaling was only done to 2km, and the 25m model forcings was resampling, the details of this resampling are also not clear in the manuscript (e.g., topographic shading for incoming shortwave radiation is mentioned briefly in L327 and L625-626, but doesn't tell how this was done)

We agree on the confusion and have removed this sentence. The target resolution of the dynamical downscaling with WRF is 2 km. The resampling (i.e., interpolation) of the meteorological variables to the 25 m resolution of the hydrological model WaSiM is based on a bilinear approach for relative humidity, precipitation, wind speed, and shortwave incoming radiation. Only temperature is interpolated and extrapolated using a linear regression model. Shortwave radiation and temperature are also corrected for topography (shading effects) following Oke (2002), a relatively simple terrain-correction scheme widely implemented in hydrological models. The idea is that a sloped surface does not receive the same amount of direct solar radiation as a horizontal surface. Oke's scheme assumes that differences in slope temperature are mainly driven by differences in the amount of received solar energy. In mountainous catchments, this captures the first-order effect that sunny south-facing slopes are warmer and receive more energy for snowmelt and evapotranspiration than shaded north-facing slopes. We have added the following sentence describing this method (lines 314ff):

"A topography-dependent correction following the approach of Oke (2002) was applied to account for terrain effects on solar radiation and air temperature. Incoming solar radiation was adjusted according to slope, aspect, and terrain shading through solar-geometry calculations, and air temperature was subsequently modified using an empirical relationship linking local temperature deviations to the radiation correction under prevailing sunshine conditions."

L8: "2 m WRF relative humidity" can be interpreted as relative humidity was available at 2m spatial resolution. I am presuming authors mean relative humidity at 2m. I would suggest using "relative humidity" instead of "2 m WRF relative humidity".

Yes, the "2 m" means above ground. We followed this suggestion and removed the 2 m.

L9: "... the impact of TLRs computed from different WRF layers (i.e., 2 m and free atmosphere)". Mentioning 2m and free atmosphere layer is confusing to readers (and can be skipped) as they don't have any perspective of how many layers, its discretization, etc.

Yes, we agree and have removed this information.

L11: "... TLRs and the HM results are significantly dependent on the coarse surface properties of WRF." Three points: (1) with so many abbreviations, its hard to read this sentence; (2) which "surface properties" of WRF are being referred to?; (3) most importantly, the relationship of TLRs and HM have not been explicitly linked with surface properties, as highlighted in this text. For linking with surface properties, one would need some kind of correlation/classification analysis as to how TLRs were different in different land uses, or change land use in WRF (e.g., spatial resolution of landuse) and see if this results in different TLRs.

We totally agree on this point. We reduced the use of abbreviations in all parts of the manuscript and rephrased the entire abstract.

L15: “bivariate correction” can be confused with bivariate bias correction, which often refers to statistical downscaling using two variables. In this article, only a single dynamic downscaling was done, therefore this sentence could be simplified to something like: “the effects of temperature lapse rates were used to further improve relative humidity calculations ...”

Thank you for pointing out this possible misunderstanding. We agree on this point and have rephrased the descriptions as follows (lines 87f):

“In addition, we apply a sequential, thermodynamically consistent correction of WRF relative humidity based on the corrected temperature fields.”

We rephrased the term “bivariate” correction throughout the manuscript.

L38: I think “preserved” would be more apt than “resolved”

We reversed the word accordingly.

Figure 1: Typo: should be “delineation”

Checked.

As glacier cover was 27% in 2020 (see L100), please mention this information in L92 to give readers a perspective of the extent of glacier loss since 2015.

The relative glacier cover of 2015 is already available in line 111 (30 %). Accordingly, the relative glacier cover declined by 2 % from 2015 to 2020. We added this information (line 115) and restructured this section.

Table 1: “Amount of NA” is not useful information, especially when “Share of NAs” is provided

We removed the column “Amount of NA” from Table 1.

L142: please mention the resampling method used to convert 250 m MODIS grids to 25m resolution

We applied a bilinear resampling method. We added this information to line 159.

L210: Typo, should be “Eq.” instead of “Equ.”

Done.

Figure S8: why has glacier melt contribution been quantified in this? I guess to evaluate good-of-fit readers should only compare Q simulated by WRF with observed Q?

Simulated glacier melt was included solely to illustrate differences in its seasonal contributions (Fig. S9), as it shows the most pronounced variations depending on the temperature corrections applied. It is not taken into account in the calculation of the GOFs. The GOFs were calculated for daily runoff, not for the mean day of the year time series shown in Fig. S9.

L495: shouldn't it be lower "underestimation of streamflow peaks"? Q WRF in Figure S8a is lower than Qobs and more comparable in Figures S8b and S8c.

This sentence focuses on the comparison of the two WaSiM simulations forced with corrected WRF inputs (i.e., T corrected (Fig. S9b), and T and RH corrected (Fig. S9c)). In this context, the wording "overestimation" is correct (lines 472ff):

"The improvement in model performance is due to lower overestimation of streamflow peaks during the summer period (Fig. S9b and c), resulting from reduced glacier melt contribution."

L525-535: the role of temperature lapse rates is prevalent on relative humidity, with relative humidity values reducing significantly after applying this correction. Therefore, it makes sense that evapotranspiration increases as a drier atmosphere causes higher vapor pressure gradient. Post the temperature lapse rate correction, even sublimation should increase. However, this is not shown anywhere in the manuscript. Does the model not simulate sublimation?

Yes, we add this explanation for higher evapotranspiration rates with WRF T and RH corrected.

Although WaSiM simulates sublimation, the model version we used does not output mean catchment-wide sublimation rates, as it is only part of the evapotranspiration output. However, it does output the sublimation separately for glacierized grid cells. We additionally evaluated the sublimation rates on glaciated cells for the period 2009-2015 and added these findings to the revised manuscript (lines 508ff):

"We additionally evaluate the simulated sublimation rates on glaciated cells. The uncorrected WRF temperature yielded the lowest mean annual sublimation with 8.39 mm. In the case of corrected WRF temperature, 12.8 mm/a mean sublimation was simulated, and 20.5 mm/a sublimation when corrected WRF temperature and relative humidity were considered as model forcing."

L537: consider adding uncorrected here: "the near-surface uncorrected WRF temperature ..."

Done.

Figure 11b: why have you not incorporated GLMB obs for three years (as done in Figure S13)? This will provide readers some comparison with observations.

We haven't incorporated the observed GLMB in Fig. 8b, as it covers only the period 2000-2015. As the plotted time series in Fig. 8b are cumulative starting in 1973, we unfortunately can't add the observed GLMB to the extended period shown in this figure.

Response to referee #4 (report #2)

General comments

In this technical note, the authors detail a method for downscaling WRF output to a high resolution (25m) grid needed to run a physical hydrologic model. While the intent behind the study and the hydrologic modeling itself are interesting and well motivated, the experimental setup of the WRF simulations and use of the WRF output as forcing data suffer from an unclear methodology. If the methodology is clarified in text, and corrected in the analysis, if necessary, the publication may be technically sound as a result.

We would like to thank you for taking the time to review our manuscript. According to your comments, parts of the methodology were not clear. We deeply regret this, as the applied methodology is at the core of our work, and this lack of clarity makes it difficult to follow the results and discussions presented. We are confident that addressing the comments and revising the methodology section will improve clarity and that the revised manuscript will provide added value to the scientific community.

Major comments:

The crucial problem with the technical note is the lack of clarity surrounding what temperature data is used where. The conclusion mixes references to temperature output from WRF, and writes “corrected 2 m WRF temperatures”, where the equation for corrected temperatures (EQ 4) notes that the WRF temperatures from the first model level are used. 2m temperatures from WRF are diagnosed from the skin temperature and sensible heat flux or the exchange coefficient, both of which are heavily dependent on the surface land cover type. Thus, the fact that the correction reduces reliance on WRF surface types relative to the uncorrected temperature is likely more because one (corrected) uses temperature of the first model level as the starting point, where the other (uncorrected) uses diagnosed 2m temperature as the starting point. If I have misinterpreted the methodology here, I apologize, but I have tried my best to understand what exactly was done.

Although we mentioned the WRF skin temperature in the Method section (2.4, Fig. 2a), we do not consider it in our correction approach or evaluation in the following sections of the manuscript. We evaluate the uncorrected 2 m WRF temperature against 2 m temperature observations at four stations. Due to the discrepancy between the 2 m WRF temperature and observations at the highest station (2640 m) caused by the WRF snow and ice surface properties, we decided not to include the 2m WRF temperature in our correction approach, which uses the first model level (eta level 1 at approx. 17 m above surface, Fig. 2c) up to level 20 (Fig. 2d). Since we are computing hourly temperature lapse rates across eta level 1 to 20, we can compute the WRF temperature with Equation 4 at any altitude in our domain. The target elevations are the four station locations. We see the potential for confusion with the term “corrected 2 m WRF temperatures” in the conclusion. Thus, we remove the “2 m” in this sentence.

Section 2.2: What height above ground are the temperature observations from? Are they all the same height? (2m vs. 10m) This is an important detail of the study, mainly if the WRF temperature data is being compared against observations from the same height above the surface.

All temperature observations are measured at 2 m height above ground. We ensured the comparability with the WRF 2 m temperature.

Does your 2km WRF domain only cover the area shown by the green triangles in fig 1? This is a small horizontal extent and makes me doubt that the precipitation data at the 2km resolution is influenced by processes at this spatial resolution. If the target domain is 200 km², and the target resolution is 2km, this would mean ~ 7 pixels in each direction. This gives little time for the influence of orographic lifting resolved at the 2km scale to influence hydrometeors, and they may fall out from the atmosphere after already leaving the 2km domain. Some of the measurement stations also lie within the first grid cell from the border, where the 2km domain state is still adjusting to the boundary condition imposed by the coarser resolution nest. Some discussion of this experimental flaw should be included in the text, or at the least a justification for why this setup was chosen.

The dynamical downscaling with WRF covers a much larger domain of approximately 24,416 km² at the highest spatial resolution of 2 km (see Figure 1). In this technical note, we analyze WRF outputs for a much smaller area of about 200 km², as illustrated in Fig. 1 of the manuscript. As our area of interest (the Upper Kauner Valley) is located in the middle of the 2 km WRF domain, we should not experience any boundary effects.



Figure 1: Dynamical downscaling with WRF is performed in three nested domains, with a grid spacing of 18 km (domain 1), 6 km (domain 2), and 2 km (domain 3). The WRF domain with 2 km resolution covers an area of about 24,416 km².

The 2km WRF domain also features roughly 30 levels in the troposphere, and an additional 20 levels spanning into the stratosphere. More vertical levels in the lower atmosphere are typically used when modeling in complex terrain at the kilometer resolution to ensure that turbulent structures on scale with the target resolution are resolved within the PBL. Taken together with the nesting setup and horizontal domain extents, the WRF simulations are not well constructed for simulating 2km precipitation patterns. The obvious explanation is that computational limits informed the modeling configuration, but the resultant setup begs the question of what value this 2km nest adds.

The vertical WRF configuration (first level ≈ 17 m AGL, 12 levels below 1 km, 16 levels below 2 km) reasonably matches the 2 km horizontal resolution over complex terrain. Indeed, the smooth transition between temperature lapse rates (Fig. 4 in the manuscript), which remain nearly constant from levels

1–20 (roughly -0.5 to -0.7 °C/100 m), indicates that the WRF model is not suffering from obvious vertical under-resolution in the lower troposphere. The main limitations for representing mountain boundary-layer processes are more likely to arise from the 2 km horizontal grid and terrain smoothing than from the number of vertical levels in the lower atmosphere.

Line-level comments:

L40: The temporal resolution of your output does not influence the runtime of an atmospheric model – just list the dependency on the horizontal resolution here.

We agree and removed the temporal resolution of the WRF outputs from this sentence (line 47).

L181: There is not a surface cooling “up to” a 2m height – T2 is diagnosed in WRF from the exchange coefficient, surface temperature, and lowest atmospheric temperature. Since this is a diagnosed quantity, describing a cooling effect up to a particular height is misleading.

We see the misleading formulation and rephrased the respective sentences as follows (lines 197ff):

“The TSK and T2–elevation relationships (Fig. 2a,b) show substantially lower temperatures over snow- and ice-covered grid cells, together with considerable scatter. This suggests that the simulated temperatures are strongly influenced by surface properties, with the cooling effect of snow and ice extending from the surface to the near-surface air represented by T2.”

L188: I disagree with most of this paragraph. There is almost no scatter at level 49 – you just zoomed in the y-axis to amplify the apparent scatter. It is also unclear what the authors mean here by “lapse rates”. The lapse rate is defined as the change in temperature as a function of height in the atmosphere. The lapse rates discussed here, however, seem to be the spread of atmospheric temperatures for a given WRF level as a function of the underlying surface height. The later definition is not a lapse rate, but an artifact of the model grid.

We revised the paragraph accordingly (lines 207ff):

“From eta level 1 to eta level 20 (Fig. 2c,d), WRF temperature exhibits an almost perfectly linear relationship with surface elevation and very little scatter. This indicates that temperatures at these levels are predominantly governed by a nearly uniform domain-wide lapse-rate structure, with only minor influence from land-cover or snow-cover differences. Above eta level 20, the scatter progressively increases, and by eta level 49, no clear relationship with surface elevation remains. This behavior reflects the decreasing influence of terrain with height: the lower eta levels closely follow the terrain, whereas higher eta levels become increasingly horizontal and sample the free atmosphere, where temperature variations are controlled more by large-scale atmospheric structure than by the underlying surface elevation.”

Fig 2: The scale of the y-axis should be fixed between these panels. At the moment it is misleading to show a linear trend with surface elevation, when in reality the temperature of the given levels are simply becoming more homogenous as we approach isobaric coordinates.

We have chosen independent y-axis scales in Fig. 2 to show how the relationship evolves and how scatter changes with altitude. However, we understand the issue of independent scaling in interpreting the linear gradients and have explicitly stated in the caption that the y-axis ranges differ between panels.

L238: The derivation and utility of Equation 6 is unclear. Is HWRf The geopotential height of the first WRF level for the grid cell where the station is located? Why is the sign of the geopotential reversed relative to equation 4? Why is WRF's 2m temperature now used? No, HWRf represents the altitude of the WRF grid cell at the station location. We clarified this in the manuscript.

We see the inconsistency between Equations 4 and 6. We homogenized the two equations so that both perform the same operation.

Yes, we are extrapolating from two different vertical representations (eta level 1 in Equation 4 and WRF 2m temperature in Equation 6), which is not strictly physically consistent. The objective is to compare these to different WRF near-surface representations against station observations, after elevation normalization, to decompose the temperature bias and evaluate complex-terrain temperature behavior. We have added this additional information to the respective section and the Discussion (4.1).

L320: Why is no correction to longwave applied? For the differences in elevation reported over the domain (~500m in some locations), a bias in longwave of several Watts can be expected due to increased scattering through the atmosphere.

As the main focus of this technical note is correcting WRF temperature, we refrained from correcting or calibrating longwave radiation for the hydrological model study. However, we have added another sentence to the discussion (section 4.3; lines 634ff):

“Although we observed elevation differences between WRF and station of more than 500 m in some locations of the domain (e.g., Dammfuss), which probably causes a bias in longwave radiation due to increased scattering through the atmosphere, we have not calibrated the incoming and outgoing longwave radiation of WaSiM, as this would be outside the scope of this technical note.”

L333: Preferential deposition is not simulated using the techniques presented here. The spatial scale of the WRF simulations and downscaling techniques used would not capture this process. Snow may be redistributed within WaSiM, but “preferential erosion and deposition” should not be mentioned here.

Yes, the horizontal WRF resolution of 2 km is too coarse for resolving snow redistribution processes. However, in Section 2.6 of the Methodology, the configuration of the hydrological model WaSiM is explained, which is able to simulate gravitational and wind-induced snow redistribution. Both processes are considered in this technical note.

L366: It is unclear what “The uncorrected WRF temperature corresponds to the WRF hourly temperature “ refers to. The 2m temperature output from WRF is not the same as the WRF temperature of the first vertical level. If the uncorrected WRF temperature referenced in the WaSiM runs throughout Section 3 refer to runs forced using WRF 2m temperature, and the “corrected” runs use the WRF temperature field from the 1st model level with a lapse rate applied, the comparison is ultimately apples to oranges, as fig 2. Shows. The diagnosed 2m temperature, especially over ice surfaces, is lower than the temperature of the first model level. What height does WaSiM expect the meteorologic forcing to be at? This would explain your result of growing glaciers in fig 9d.

We have rephrased the respective paragraph to point out the main objective of the comparison of WRF 2-m air temperature and extrapolated WRF temperature from eta level 1 (lines 352ff):

“As a first evaluation, we compare WRF temperature with station observations from the upper Kauner Valley and the surrounding area (Fig. 1 and Table 1). The uncorrected WRF temperature corresponds to the hourly WRF 2-m air temperature (T2) of the respective 2-km grid cell and is adjusted for the height difference between WRF and station elevation (Eq. 6). T2 is a diagnostic near-surface variable that depends on the model’s surface-layer parameterization and underlying surface properties. The corrected WRF temperature is derived from the temperature at eta level 1 and extrapolated to station elevation using the hourly temperature lapse rate (Eq. 4). The comparison, therefore, evaluates whether an elevation-adjusted temperature derived from the first atmospheric model level provides a more accurate representation of observed near-surface temperatures than the standard WRF T2 product.”

WaSiM expects the station elevation for computing the elevation lapse rates.

Yes, the diagnosed 2m WRF temperature is lower than the first model level, especially over the highest catchment areas covered by snow and ice. This explains the hydrological model result of growing glaciers in Fig. 7d.

L435: The uncorrected temperature time series should be included in fig 7. Excluding it removes a key contextualizing result: your temperature correction method made no difference here.

To make Fig. 6 easier to read, we have not included the uncorrected WRF temperature, as it is identical at this location and the lines overlap. However, we have added this information to the figure caption (lines 423f):

“Note that the uncorrected WRF temperature is not included as it is identical to the corrected WRF temperature for this site, and the lines would overlap.”

Fig 9d. The figure shows that uncorrected WRF data, from a WRF simulation forced by GCM data, predicts glacial expansion in the European alps over a 10 year period on par with the predicted glacial ablation from the corrected data. This is a surprising result considering the observed glacial retreat over the Alps during this time period, and makes me further suspicious of a methodological error in this technical note.

The hydrological model run forced with the WRF 2m temperature serves as a negative example of how significantly wrong the model performs, especially for the glacierized parts of the catchment (Fig. 7d). This is a good example of the complex feedback among cryo-hydrological processes, which are particularly temperature-sensitive. Our results suggest that the diagnosed WRF 2-m temperature is strongly influenced by surface properties and parameterization choices in snow- and ice-covered terrain. In contrast, temperatures derived from the first atmospheric model level and extrapolated using the hourly lapse rate provided a more robust representation of near-surface air temperature and led to improved hydrological model results in the study catchment.

L586: You mean fig 11c instead of 10c

Yes, we corrected this accordingly.