Final response to RC1

General comments

Hofmeister et al. applied WRF 2 km data to a physically based hydrological model with a resolution of 25 m resolution in a small (62 km2) highly glaciated catchment in Austria. They provided two different types of air temperature to the hydrological model, one is the temperature 2 m above the surface, and one the so-called "corrected" temperature, which was derived from WRF temperature on the first and 20th level (~3 km above ground). They also corrected a drizzle effect in WRF by setting all values to 0 below a monthly determined threshold based on a station nearby the study area.

Comparison with four stations within the catchment showed an improvement for the corrected temperature, especially for the highest station which was presumably wrongly classified as glaciated in WRF. With the corrected temperature also the hydrological model performance improved, e.g. comparison with MODIS determined snow covered fraction improved, the sign of glacier mass balance turned from obviously wrong positive to realistic negative and comparison with observations of streamflow improved.

However, the authors promised in the title, in the abstract and the introduction with the stated research questions much more than this synthesis, which was later not presented in the manuscript. I will provide examples in the next section. Oppositely, the manuscript contains significant parts which do not relate to the stated research questions and rather show interesting applications of such a model chain or a discussion/review on other studies. It is a long manuscript and there is a clear need to focus.

While reading the manuscript it was often not clear to me what motivation was or how the methods were applied. Reading the conclusions at the very end was a (belated) eye-opener for me, and yet questions remain (see next section for details).

Another important issue is the limited availability of data to ensure that the hydrological model used to evaluate the altered temperature inputs does not simply improve the results due to error compensation (see details in also in next section).

I acknowledge that preparing WRF model results and applying a high-resolution, physically based hydrological model as WaSIM is an extensive task (WRF results were probably taken from Altmann et al. (2024) and it is not clear if they were independently generated, but this is not the relevant point here). I suggest that the authors completely revise their manuscript and clearly focus on what they can provide with the presented analysis. However, I have the feeling that a distilled manuscript is not sufficient for an own publication and that the authors should rather proceed with their obvious plans of analyzing model output from long-term runs since 1850, where this here presented part of adjusting air temperatures is a small paragraph.

We are grateful for the comprehensive feedback on our manuscript. We agree that the manuscript is relatively long and extensive. This originates from the fact that the application of such a model cascade from an RCM to a distributed hydrological model is rather complex when cryo-hydrological processes are modeled with physics-based model approaches. Consequently, the method and discussion sections of the manuscript are relatively detailed, providing descriptions of the correction of WRF temperature and precipitation data, as well as their respective limitations, which are discussed in the discussion section. We apologize for the confusion, as the manuscript was intended to be submitted as a technical note rather than a research article. Nevertheless, we agree that a revised manuscript should clearly link the title, abstract, introduction, and synthesis to enhance comprehensibility. Additionally, we will condense the manuscript to its core aspects, remove unnecessary content, or shift it to the supplement when needed as additional information.

Although the intention of the developed workflow is to model long-term hydrological changes from 1850, we focus on a more recent period (1970-2015) in the presented manuscript to evaluate the model results against cryo-hydrological observations. As the manuscript is already extensive, we have refrained from including the long-simulation results that start in 1850. To avoid any confusion regarding the periods of interest, we will focus solely on the evaluation period (1970-2015) in the revised manuscript.

As the developed methodology for correcting and evaluating the WRF temperature is rather complex and, from our perspective, novel in the field of meteohydrological model cascades, it is not possible to briefly describe the applied workflow in a single subsection. However, it would be conceivable to integrate a large part of the technical descriptions and evaluations relating to the WRF corrections into the Supplement.

Detailed comments

Stated but not done (as it first appears)

Small-scale surface representations in WRF

The title and the first research question (line 96) suggest that small-scale surface representations in WRF and their impact on subsequent hydrological models have been studied. The authors want to present the effects of an alternative derivation of near surface air temperature, which should consider elevation and land cover classes (line 100). However, this correction is not based on any small-scale surface properties; the opposite is true: It is removing the influence of the probably too coarse WRF land cover information by simply averaging temperatures over the whole domain at elevation levels where they are relatively insensitive to land surface properties (which is at the end also stated in the conclusion, lines 684-687).

We agree on the misleading terminology and the conflict between the title/abstract and the conclusion. Our motivation is to first demonstrate the impact of WRF land cover classes on WRF near-surface (i.e., 2 m) temperature and secondly to provide a workflow for adjusting the WRF temperature, which is independent of the land cover. Since the spatial WRF land cover resolution cannot resolve the small-scale surface characteristics of a meteorological station, which is typically a few square meters in size, a robust and consistent workflow for adjusting WRF temperature is highly needed, especially in the complex terrain of high-elevation catchments. Our study demonstrates that spatial averaging of WRF temperature from different eta (i.e., pressure) levels still yields more reliable temperature predictions at several meteorological sites located at different elevations compared to the initial 2 m WRF temperature.

Bridging the scale from WRF to WaSIM

In the abstract the authors stated that the main challenge is to transfer forcing data to the much finer resolution of the hydrological model. The authors wanted to present a workflow for bridging this scale. This relates to the second research question in the introduction (lines 97-98). However, I can only see the application of rather simple WaSIM integrated algorithms (lines 300 to 309) which cannot be meant by a newly presented workflow.

Unfortunately, the wording in the abstract is too vague, causing confusion. It is true that the applied algorithms for inter- and extrapolation of meteorological variables are already implemented in WaSiM and were not developed in this study. We will revise the sections accordingly.

Consistency of bias-corrected RCM data over longer periods

This relates to the last and third research question (lines 98-100). However, the authors have not applied bias-correcting method for temperature using observations. The temperature adjustment presented is not using observations (but still removed biases). The consistency was tested in comparison to a nearby valley. However, this was not done in relation to non-corrected data (and probably only for one reference grid point), so one cannot state an improvement.

Yes, that is true. We will also include the uncorrected WRF temperature in this comparison for consistency. Yes, it only represents one meteorological station and the respective grid cell. However, we can perform further evaluations for other stations with long temperature time series from other valleys and add them to the supplement.

For precipitation they removed a drizzling bias using long-term observations which is a classic bias-correction case. However, no consistency results were presented here.

We only present evaluations of the monthly precipitation days in the Supplement (Figure S8). We will extend the evaluation and add some consistency results to the Supplement.

Modelling period

In the abstract it was stated that the modelling period is from 1850 to 2015, however, only data from the 1973 to 2015 are analysed in this manuscript.

As previously mentioned, we will focus solely on the period from 1973 to 2015 in the revised manuscript.

Topics outside of the focus

The manuscript can substantially be reduced to topics which are part of the research questions. Example of topics, which are not part of an evaluation of presented temperature and precipitation correction may be the determination of significant change points (unless it is not evaluating the corrected input).

Another example is the long discussion as it does not relate own results with other studies. It can be shortened and put in the introduction or placed in a review paper. The implications of own results are hardly discussed in the discussion.

We will condense the manuscript to its core aspects, remove unnecessary content, or shift it to the supplement when needed as additional information. We will set the main focus of the discussion on our own results.

Clarity

In equation 5 and 6 it is stated how the corrected and uncorrected temperature at a certain station location is calculated. However, it is unclear to me how this is done for WaSIM grid points: In lines 302-304 I can read some information, but how the variable lapse rate is determined is unclear, and whether this is done for both variants. This is particularly relevant as Eq. 6 contains a fixed lapse rate to account for elevation differences.

WaSiM computes a time-step-dependent (i.e., hourly in our study) elevation-temperature gradient based on the constant elevation of the WRF grid cells and the respective WRF temperature and assigns each WaSiM grid cell a temperature according to its respective elevation. For consistency, this approach is identical for uncorrected and corrected WRF temperature. We will add another sentence after line 304 to clarify this. The fixed lapse rate from Eq. 6 is only used for comparing corrected and uncorrected WRF temperatures at station locations, but it is not considered in the WaSiM simulations. As already mentioned, the temperature lapse rate is not fixed for both WaSiM simulations (i.e., corrected and uncorrected).

Equation 5 and 6 differ in two ways: First the mentioned lapse rate is fixed and second the 2 m temperature is used in Eq. 6 for the uncorrected version, while Eq. 5 used an hourly changing lapse rate and the first level temperature. It is unclear to me, which part is the relevant one for improvements. As this is the core part of the paper, I would suggest some additional analysis.

As we are primarily evaluating and comparing the ECDFs of uncorrected and corrected WRF temperatures with station observations (see Fig. 4), the expected difference should not be significant, regardless of whether it involves a fixed lapse rate. Nevertheless, we will add some additional evaluations regarding the sensitivity of the fixed lapse rate on the uncorrected WRF temperature at station locations. From our understanding, the main difference between WRF temperature computed with Eqs. 5 and 6 is due to the different interaction with the WRF land use classes (see Fig. 4). As mentioned before, the fixed temperature lapse rate is not considered in the WaSiM simulations.

One main argument for using the corrected temperature approach was that the land cover classification too coarse in WRF (conclusions lines 684 – 686). The authors missed to show a map of this classification.

Yes, we will add a map showing the WRF and WaSiM land cover classes to the Supplement.

Improvements or error compensation?

In the conclusion the authors state that due to the lack of meteorological stations above 2700 m they evaluate the impacts of the corrected temperatures with a hydrological model. However, it is unclear to me if the authors have included not more issues with this attempt than less. The hydrological model certainly includes uncertainties in the modelling cascade, which are also discussed in the manuscript.

One issue is that the study area is not really suited for this evaluation. While I see the point that there is a switch from an unrealistic to a realistic sign in the glacier mass balance trend, there are no measurements of glacier mass balances available. There is also no spatial snow distribution data used in this study. So, the reader does not know whether the shown improvements in fractional snow-covered area, glacier mass balance and runoff comes not from error compensations elsewhere in the hydrological model.

Before forcing the hydrological model WaSiM with WRF data, we first ran the model using meteorological station data. We evaluated its performance using a process-based approach (similar to the one employed in this study). The calibration and validation period was from 2006 to 2021. This ensures that WaSiM delivers reliable results for the Kauner Valley. The results are not included in this manuscript and supplement, as they would increase the complexity of this study. We used the same parameterization and only changed the meteorological input data (i.e., regional climate data from WRF) to examine its plausibility. We should ensure that the hydrological model gives reasonable results with these new meteorological inputs. Previous studies have demonstrated (e.g., Hofmeister et al., 2023; Clerc-Schwarzenbach et al., 2024; Clerc-Schwarzenbach et al., 2025) that meteorological forcing data can be one of the primary source of uncertainty, leading to implausible hydrological model results. The evaluation with a hydrological model is an essential step to validate meteorological / climate products for use with a hydrological model, but is often omitted. The result of this evaluation says whether the product is compatible with the hydrological model, but it does not necessarily assess the "absolute" value of that product for other applications.

Our assumption is that high-quality forcing data should enhance the reliability of hydrological model results even if the hydrological model is uncalibrated. We refrained from a forcing-specific calibration of the hydrological model due to the risk of compensating for the actual differences between the meteorological / climate products. In addition, we performed a process-based evaluation (e.g., streamflow, SWE, snow cover) that enables the reduction of potential compensation in the model, allowing for a more comprehensive approach beyond focusing solely on streamflow.

Although the Kauner Valley lacks spatial snow and glacier observations, it remains a unique research site with continuous streamflow and meteorological observations dating back to the 1970s. Since the bias problem arises particularly for high-elevation catchments with complex topography, unfortunately, these catchments still contain glaciers.

The authors argue with Figure 7 that the improvements stem from air temperature corrections leading to an improved distribution of snow, to an improved snow and glacier melt and finally to an improved runoff. One indication for an error compensation is the largely biased snow water equivalent (SWE) as well as its insensitiveness to the temperature correction (Fig 7a) at one single station. I think the authors need to provide additional insights in the correctness of their argumentation. For this it may be relevant to evaluate the model chain in regard to a realistic SWE distribution. To my opinion a 500 m resolution MODIS product is not sufficient for this complex terrain, which was also stated by the first author in Hofmeister et al. (2022). Such a study can be done using Sentinel2 data as done by Hofmeister et al. (2022) but needs to be presented for WRF input and for this specific catchment but may be limited to a few winter seasons.

We completely agree with the limitations of the MODIS snow cover map for evaluating the performance of the WaSiM snow module. We would have also preferred to use snow cover data based on Sentinel 2, but unfortunately, there is not a sufficiently long overlap between the data, as the WaSiM simulation ends in 2015, the year Sentinel 2 was launched. Therefore, the comparison between MODIS and simulated snow cover is still better than no evaluation at all. We will add this constraint to the discussion.

References

Clerc-Schwarzenbach, F., Selleri, G., Neri, M., Toth, E., van Meerveld, I., and Seibert, J.: Large-sample hydrology – a few camels or a whole caravan?, Hydrol. Earth Syst. Sci., 28, 4219–4237, https://doi.org/10.5194/hess-28-4219-2024, 2024.

Clerc-Schwarzenbach, F. and do Nascimento, T. V. M.: Evaluating the quality of the E-OBS meteorological forcing data in EStreams for large-sample hydrology studies in Europe, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-3710, 2025.

Hofmeister, F., Graziano, F., Marcolini, G., Willems, W., Disse, M., and Chiogna, G.: Quality assessment of hydrometeorological observational data and their influence on hydrological model results in Alpine catchments, p. 02626667.2023.2172335, https://doi.org/10.1080/02626667.2023.2172335, 2023.