Response to Reviewer

RC1 - https://doi.org/10.5194/egusphere-2026-3922-RC1

Dear Reviewer, Thank you very much for the detailed and constructive feedback. It is much appreciated! Please see our responses to each comment below.

Major points
1) LES averaging: Please correct me if I missed something, but there seems to be a major flaw in the model output. Do you really write only one file per hour for your LES? And, as a follow-up question, is this output instantaneous (i.e., non-averaged)? If you really employ an LES at a horizontal grid spacing of 50m, your flow structure is highly turbulent and one value every hour is very likely not representative for your model. Furthermore, if you want to grasp information on the characteristics of turbulent flow, you would have to calculate a mean and variance for turbulent variables (e.g., wind components, temperature, etc) to characterize and analyze its properties correctly. If you did not average your turbulent flow properties from your simulation results, and only write one file very hour (even 10-minute or 5-minute values would be more beneficial), it is also no surprise that your model does not capture transient regimes, as discussed in the manuscript, because these phenomena occur on timescales of an hour or less. I am afraid that the "one file hourly" per LES simulation renders your numerical data useless. However, there are options in ICON to write high-frequency turbulent output, this function is called "meteogram" and can be set in the model's namelist. With meteogram, you can write output for each model time step (i.e., 0.25 sec for your inntermost domain) for selected points on your model domain. A good choice would be the locations of your HEFEX stations, since you mostly compare your simulation output to point observations. A second option would be online averagring of the 3D fields, like in Umek et al (2020) and Goger et al (2022), but this needs a large chunk of coding work, because, to my current knowledge, this feature is not implemented in ICON for turbulent quantities. As a third option, I think ICON offers the calculation of time-averaged quantites, and I assume the authors used this function for the wind fields. However, 30 minutes might be still a too large time window to grasp information on the flow structure in the mean - as discussed abive, you still do not have the variance of the flow by this function.

We understand the concern and agree that this part is poorly described in the manuscript.
 	
Our experiments created various model outputs:
1. Full grid: Hourly instantaneous fields for all model levels and selected pressure levels
2. Full grid: 30-min averaged wind, temperature, and humidity fields (ICON internal calculation)
3. Full grid: 30-min variances and covariances (u', v', w', T', q') based on the defined time step (= 4 Hz for our innermost domain); implemented by using ICON's Community Interface (Com-In); only for selected time periods
4. Meteogram output for selected coordinates from HEFEX campaigns at 15 min time steps
 	
Since turbulent fluxes and covariances have not been analyzed in this study, high resolution output is not necessarily required. Instead, this validation uses mostly the 30-min averages calculated by ICON (2) and the meteogram outputs (4). We will update the manuscript accordingly and better highlight the calculation basis of the data used for the analysis. Furthermore, we'll update the dataset description in the Zenodo record for full transparency.
 	
As part of this response, we compared and analyzed the different temperature outputs for station T275 for a clear-sky day from 00:00 until 18:00 (see Fig. 1 in the Appendix). Our main findings are:
1. As expected, the high frequency meteogram output shows temperature fluctuations which are not captured in the aggregated data
2. Also as expected, the hourly instantaneous data deviates slightly from the 30-min averages
 	
Furthermore, based on all your suggestions, we have decided to rerun the simulation for the HEFEX II and HEFEX III periods, with the following configuration changes:
- Output settings:
    - Full grid: 30-min averaged temperature, humidity and wind fields
    - 4 Hz meteogram outputs for all GPS coordinates where measurements have been taken
- ExtPar/Namelist settings
    - CORINE instead of GLOBCOVER (but still with updated glacier extents and albedo values)
    - slope correction and topographic shading are turned on
    - ecRad radiation scheme
 	
The updated version of the manuscript will incorporate the new simulations. This will lead to changes in the 'Methods' section and the figures and metric tables will most likely be updated. The discussion part may also be updated if the new outcomes lead to different findings compared to the current version.

2) Comparison observations -- model output. There is another concern on the comparison between observations and simulations. I assume - at least in the case of the sonic anemmeters - the observed variables are averaged over a certain period of time. How were they processed, and with which averaging persiod and algorithm? Furhtermore, and this is a follow-up to the model data - comparing averaged and instantaneous model output data also leads to inconsistencies int he interpretation of your results. Please clarify.
Since observational data is coming from different measurement systems (e.g. AWS, Lidar, UAV) and from multiple field campaigns (HEFEX II and III) and collected from various institutions, each dataset has its own processing routine/algorithm. For the comparison of observations with model outputs, the

same averaging periods have been used (e.g. hourly averages where assigned timestamp denotes end of averaging interval). For HEFEX II, a dedicated campaign description article has been published already (Nicholson et al., 2025). A manuscript for HEFEX III is already in process and the manuscript will be submitted September 2026 (DOI: tba). However, we understand it’s cumbersome to collect the relevant information from referenced articles. We will append a table that lists the used sensors, measurements heights, averaging periods and used algorithms (especially for the ultrasonic anemometers). Furthermore, we’ll also include this information in the Zenodo records for the sake of completeness.
Minor comments
line 20: "unique microclimates" - What exactly makes them unique? You might also want to define the term "microclimate".
We agree. The term will be defined more specifically.
line 21: "short distances": please provide numbers
Accepted. Numbers will be added.
line 22: MBL: PLease use the abbrevation "MoBL" as defined by Lehner and Rotach (2018).
We will update this in the manuscript.
line 22: "The MBL is the lowest part of the troposphere, directly influenced by mountainous terrain, where local topography and thermal effects strongly affect atmospheric flow and turbulence." - There are two relevant points missing: the influence directly by the surface and the mentioning of timescales of a few hours or less. I would recommend that you use the direct definition from Lehner and Rotach (2018) before inventing something different.
We agree. The already established definition will be used.
line 25: "katabtic flows": They are also thermally-driven, right? you might mention what makes them special over glaciated surfaces compared to other anabatic/katabatic flows in complex terrain (Farina and Zardi, 2023).
We will rephrase the sentence to make the distinctions more clear.
line 26: "interact with larger-scale pressure gradients in ways that remain poorly understood": A flow is the results of a pressure gradient, so what do you mean with "interactions"??
This refers to two-way interactions, meaning that not only will large-scale pressure gradients influence the local (valley) climate, but valley-scale processes will also propagate to larger scales. These feedback mechanisms and interactions are not well understood, but NWP experiments and sensitivity studies can help clarify them. We will be more precise here to avoid misunderstanding.
line 26: "remain poorly understood". This is a very bold statement. The interaction between thermally induced flows is a major body of research in mountain meteorology and is described in numerous oublications and textbooks, e.g. by Whiteman and Durran (1993), in Whiteman (2000), and Steyn et al. (2013)
See previous comment.

line 46: Since the manuscript is supposed to be part of the TEAMx collection, You might mention the aim and purpose of the TEAMx campaign here, trying to solve these issues in part. e.g., Rotach et al (2022)
We'll update this in the manuscript and add more information about the TEAMx program as well as reference the campaign description paper from Lehner et al., 2026.
line 50: It might be fair to mention Goger et al (2025) here as well on changing glacier surfaces.
Fixed.
line 55: I would not call the Inn Valley "less complex terrain".
We agree. The sentence will be updated.
line 55 onwards: You might state your motivation for this manuscript more clearly, by honestly mentioning that there were simulations done at this very location, but with the WRF model. You can directly start mentioning ICON's potential advantages compared to the WRF model (there are plenty), and then state that a detailed validation of ICON for this particular glacier is of general interest for the scientific communtiy. Otherwise, to be provocative, it is really not clear why we need "yet another Hintereisferner LES".
We agree. The motivation for this paper will be restated more clearly. Why ICON? To our knowledge, no comparable validation has been conducted for ICON-LES, especially regarding the combination of complex terrain, high resolution, and glacier surfaces. The latter is of particular interest to us. For our research, it is important to ensure that ICON can reproduce atmospheric conditions for use in climatological sensitivity studies with factorial experiments under retreating glaciers. Furthermore, ICON's triangular grid is well-suited for steep terrain. Unlike WRF, ICON is a European model with the important advantage of a growing community actively developing it further. Why Hintereisferner? We decided to validate ICON for the Hintereisferner domain because the HEFEX II and HEFEX III field campaigns provided us with a very rich dataset, which is necessary for an in-depth validation.
line 58: Which gap?
This refers to the knowledge gap that we actually do not know how well ICON performs in these environments.
line 60: I think at this point the Rotach et al (2022) papr is more suitable.
We'll consider this.
line 60: Is LES not inherently "high-resolution"?
This is correct. We will update this in the manuscript.
line 60: and what are high-resolution observations?
We will state more clearly that we mainly refer to high **spatial** resolution (horizontally and vertically). However, some observations from AWS (EC stations) and Lidars also provide high **temporal** resolution but this is less relevant for this study since data is averaged before comparison.
line 61: Please be aware that not all your potential readers are experts in mountain meteorology. What is the glacier-valley circulation?

We’ll update this line to make it more accessible for a broader audience.
line 83: This list of phenomena needs a reference ... there are plenty availabel about the HEF region.
We all add relevant references, e.g. Obleitner (1994), Mott et al. (2020), Goger et al. (2022), Nicholson et al. (2025)
line 83: The "previous field campaign" is called HEFEX-I.
We will add this.
line 93,94: At which measurement height were the quantities recorded?
This is all described in detail in Nicholson et al. (2025). However, we’ll also add details in the appendix (see second comment from “Major points” section).
line 95: Do we see on the figure which stations are ultrasonic anemomenters?
For this study, we mostly use low frequency measurements (except for wind data of HEFEX II stations). We will show this in the map of the study area (Figure 1). Curious readers will find detailed information in the corresponding Zenodo upload or the referenced campaign papers, respectively.
line 98: These phenomena can ba also important during accumulation, for phenomena like snow drift, see Voordendag et al (2024).
Good recommendation. We will add this.
line 143: "competing forcings": What do you mean by that exactly?
We can see from the data that the valley atmosphere is forced by large scale flows (from ENE) which competes with the locally produced thermal forcing resulting in katabatic flows (down-glacier toward NE). We will try to rephrase it to avoid misunderstanding.
line 124: Which "glacier valley exchange processes" do you wish to investigate? You often mention them in a diffuse way, but do not define them anywhere.
We will rephrase the sentence to be more precise and avoid misunderstanding. We refer to the exchange of heat, moisture, momentum and trace gasses/mass.
line 160: "nonhydrostatic dynamical core, unstructured triangular grid, and terrain-following vertical coordinates" - every state-of-the-art NWP model does have these features, this does not make ICON unique.
We will be more precise on the features that make ICON unique and advantageous over other NWPs.
line 163: "explicitly" [...] "resolves a wide range of turbulent eddies": LES models are designed to resolve the largest eddies (hence the name) on their grid, but not a "wide range". What do you mean by that anyway?
We will rephrase it in the manuscript.
line 164: "reducing the reliance on paramterizations": So which paramterizations are not neessary anymore? about which grid spacings are we talking now?
We will be more precise in the manuscript.

line 165: misplaced reference, Dipankar et al (1015) Did not really discuss mountainous terrain the publication on idealized simulations. Perhaps a referecne on 3D turbulence over complaex terrain might be useful, like Goger et al (2018) or Rohanizagedan et al (2015)

We will check this and update accordingly.

line 166: Every other NWP model works with limited-area grids. The relevant point is the suffiecient horizontal and vertical grid spacing. Think about the scale of the phenomena you are discussing

We agree. This will be rephrased.

Table1: Please do not skip information on the intermediate grids. You submitted your manuscript to Geoscientific Model Development, your results need to be reproducible!

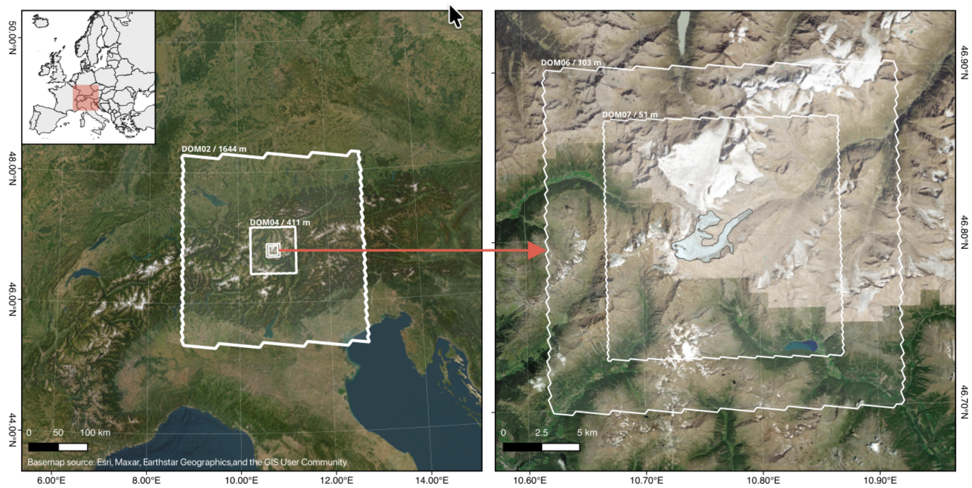
This might be a misunderstanding. The domains are not listed in the table because they have not been used in the experiments (see line 175 for explanation). DOM04 was forced with outputs from DOM02 (not DOM03). We will rename the domains in a sequential order to make it more convenient for the reader, although the technical names are different.

line 171: "resolution": I guess you mean horizontal grid spacing.

Correct. Will update this.

lines 170--177: Please add a plot of a map of central Europe where you show your nested domains: on the one hand, the reader knows the location of yout study, and furthermore, the domain placement is important for reproducibility.

We will add a map in the appendix. Also, all grid definitions have been uploaded to Zenodo already to ensure and support reproducibility (DOI: 10.5281/zenodo.19603818).



line 173: Domains are missing in the description.

We will rephrase the sentence in the manuscript.

line 174: "square root of...": Add the equation instead of the description with words.

We don’t see the need to add an equation here since grid properties are already better described in corresponding articles. We will add a reference for the curious readers.

line 175: So does this mean that you use offline nesting for all your domains? Did you consider spinup issues or do you start all your simulations at the same time?

Correct. We consider a spin-up period of one day (line 206).

line 176: Do you mean "a suffiecient numer of grid points"?
Correct. We will accept this proposal.
line 178: "consistent model parameter settings": What exactly do you mean by that? Some tuning paramters or paramterizations have to be adapted for decreasing horizontal grid spacing.
Correct, each domain has its own parametrization (namelist settings). What we mean by this is that the same parameter settings have been used for the HEFEX II and HEFEX III time frames (except for initial/boundary conditions of course). We will be more precise in the description.
line 181: "validated by Dipankar (2015)": they implemented the scheme into the ICON code - and then validated it.
We will note that it was implemented and validated by Dipanker at el., 2015.
lines 183,184: topography smoother: which smoother did you employ? I guess you still needed to adjust the smoother's paramters for the LES runs a bit.
We are unaware of the existence of different smoothers. ICON uses an iterative, selectively applied, second-order Laplacian diffusion followed by a monotonicity-limited, fourth-order (biharmonic) diffusion to smooth topography. We applied the standard ICON procedure as described in the tutorial (extpar_nml) with n_iter_smooth_topo=2 for the innermost domain. We will expand on our use of smoothing in the manuscript.
line 185: Out of curiosity, why RRTM and not the more modern ecRad scheme?
We used a template provided by MPI to create our name lists and did not adjust the radiation scheme. There was no particular reason to choose one over the other. For consistency, we used RRTM for all additional simulations and experiments. We also conducted model-configuration comparisons (see Fig. 2 in the Appendix) to determine whether ecRad produced different results than RRTM. However, the two models' outputs were nearly identical, at least for a clear day. We will repeat this comparison on a cloudy day. As previously mentioned, we will use ecRad to rerun our experiments since RRTM is deprecated and will be removed from future ICON versions.
line 191: "default preproseccing data sources": Which ones? Do they have names/sources? did you preprocess with extpar of icontools?
Correct. We used ExtPar and icontools for generation of grid, IC and LBC files. We will add this to the manuscript.
line 192: "FAO digital soil maps": reference is missing
We will add this.
line 193: What is the resolution of the satellite imagery?
We will add this.
line 195: Many previous studies with ICON LES (e.g., Goger & Dipankar 2024; Sakradzija et al. 2025, etc) use the CORINE dataset for land use. Is there a specific reason why the authors decided on the coarser, more outdated GLOBCOVER dataset?
When we created the grid files for our domains, the grid generation tool did not offer CORINE as a basis for land use fractions. Since the new Zonda provides this option, we will rerun our experiments based on CORINE, but with updated glacier outlines. We will also update the method section of the manuscript accordingly.

line 203: "ICON-based atmospheric fields": this is a very vague formulation. Do these fields stem from reanalysis, forecasts, or climate simulations?	
We use "initialized analysis" fields, which are downloaded from the PAMORE service. These meteorologically consistent 3D grids combine raw observations and a short-range forecast model to serve as the initial state and/or boundary conditions for ICON. We will also add this clarification to the manuscript.	
lines 203-205: Just to clarify again, are the simulations in online-or offline-nesting mode? In case of an offline-nestng approach, this has relevant implications for turbulence spinup in your inntermost domains (i.e., turbulence has to develop for each domain after intialization and is not "inherited" from the outer domain).	
We used offline-nesting with at least 1 day spin-up (line 206). We will be more precise regarding the nesting approach.	
line 220: What are "high resolution observations"? in time and/or space?	
The iMet sensors measure in 1 Hz, i.e. one data point every meter having an ascent rate of 1 m/s (see also comment on line 60 above). In this case, we have high spatial and temporal resolution observations but the high spatial resolution has to be mapped onto the coarser vertical ICON grid.	
line 241: "gradients suppress turbulent mixing": please be more precide. how does this "suppression" work?	
We will rephrase this in the manuscript and be more precise.	
line 245: You might mention how ICON's T2m diagnostics work. You might find them in the description of the model's surface transfer scheme, likely in Louis (1979).	
The 2 m air temperature was diagnosed separately for each surface tile by stability- and roughness-dependent interpolation between the surface and the lowest atmospheric model level using the resistance formulation of the Raschendorfer turbulent-transfer scheme (Raschendorfer, 2001), which incorporates Monin–Obukhov-type surface-layer scaling. We will be more precise on this in the updated manuscript.	
line 269: Did you check the model's surface temperature?	
Yes, the model's skin temperature (t_sk) is constantly 0 °C over glacierized surfaces.	
line 271: We do not see this missing shallow yet, would this discussion make more sense in the wind speed/direction section?	
We generally agree. However, in this particular case, we see the inability of the model to accurately reproduce surface inversion and low-level jet, due to coarse vertical resolution, as the main reason for the near-surface temperature biases. Therefore, we believe it is important to mention this point here. This is discussed separately in the wind section.	
line 275: Computational restraints are usually not the issue for lowest model levels. On the one hand, model stability is affected, and on the other hand, usually the model's surface layer formulation is violated.	
Correct. We implicitly mean the same thing. Higher vertical resolution would require higher horizontal resolution (e.g. 50 m -> 5 m) to ensure model stability and preventing the grid from departing too much from isotropy. At this point, this becomes a computational issue.	

line 280-284: Well, you could perform an evelvation correction for temperature, see Simnonet et al (2026).
We discussed this approach but considered it as not useful for various reasons: the true lapse rate of the valley atmosphere is unknown but is expected to change significantly between thermally stable nighttime conditions with a deep surface inversion and daytime conditions with a well-mixed dry adiabatic stratification above the very shallow (~2 m) surface inversion. Applying a standard atmospheric lapse rate would hence rather increase the error made. Using the lapse rate obtained from the model at every time step would potentially work, but this cannot be validated without continuous observations of the true lapse rate and could hence add to the uncertainty.
line 289: What is "surface damping"?
We will update the manuscript and explain what we mean by this.
line 292: "land surface paramterization": I guess you mean the surface transfer scheme.
Correct. We'll update this in the manuscript.
line 312: You might define a bit closer the term "stable synoptic regime"
We will change this to "anticyclonic synoptic situation, leading to stable fair-weather conditions" to avoid misunderstanding.
line 313: "and warm air masses over the glacier, favoring thermally stable stratification": this sounds counterintuitive, please reformulate
We will reformulate this to avoid confusion.
line 318: Now you say that the model represents the katabatic flow well, but a few sections before you mentioned that is not well-developed?
We apologize for the confusion and did not intend to say it is not well-developed. The point we want to make is that the observed jet maximum is at ~2 m and, thus, not sufficiently captured by the model due to the vertical resolution. This might also lead to the underestimation of the wind speed. We will reformulate the paragraph in the manuscript to avoid misunderstandings.
line 325: strong winds at glacier terminus: this might be also due to the change of ice to rock in general?
We agree and this is what we wanted to express in the lines 326-328. We might reformulate this to make it more clear.
line 346: "persistence of thermally driven downslope flow," I would take care with the formulations here - is this now katabataic forcing or something else?
We are referring to the katabatic wind and will reformulate this to avoid confusion and/or misunderstanding.
line 363: I would take care with the statement on mixing in other models - they likely treat turbulence as fully paramterized, while turbulence is partly resolved in your setup.
Thank you for the suggestion. We will be more precise in the manuscript and point out that the other models used in the referenced articles use a different turbulence scheme or parameterization, which does not necessarily translate to our study.
line 387: What is a "real model bias"? Are there imaginary ones as well?
We will reformulate this in the manuscript and be more precise what we mean by this.

line 398-400: An analysis of vertical profiles can give you the answer....
Correct and thank you for pointing this out. We will be more precise and add a reference to the Figure showing the vertical profiles.
line 424: You have 3D model data. You can answer your questions with them.
We will reformulate the sentence to state more clearly, that our model data indeed indicate the presence of the mentioned phenomena (see Fig. 3 in the Appendix as an example).
line 438: thermal forcing is also orographzy-driven, due to differential heating. Please reformulate.
We agree. We will reformulate the sentence.
line 462. It might be worth chcking if this underestimation is systematic over the entire domain, and not only over the glacier. This might be a hint that ICON's steep orography has an impact on flow - because in other models, wind speeds over complex terrain are often overestimated because of the overly smoothed topography.
Thank you for the tip. We will perform a further analysis and compare the model outputs with those from stations in our domain with less steep terrain to see if the results are similar. Depending on the outcome, we will update the manuscript as needed.
line 497: You might mention that high-resolution (meteorological) input data is also a challenge for hi-res simulations.
Thank you again for the suggestion. This is a valid point, and we will add to the manuscript that the quality of the output depends heavily on the initial and boundary conditions used.
line 504: "ICON does not account for topograohic shading": This is incorrect, there is an implementation by MeteoSwiss on topographic shading in more recent versions based on Steger et al (2023) algorithm.
Thanks for the correction. It was indeed available in the ICON version that we used, but we did not activate it. We ran a simulation for a sunny day, activating topographic shading and slope correction. This results in temperature and wind differences throughout the diurnal cycle (see Fig. 2 in the Appendix). For this reason, we will also turn it on when we rerun both simulations, as previously described.
line 508: Simulating stable boundary layers is not an ICON-specific problem - there is a great overview paper by Holtslag et al (2013) and Singh et al (2025) on ICON.
We will check the recommended references and incorporate this into the manuscript.
line 513: "SEB" Please define.
We defined SEB as Surface Energy Balance (see line 504).
line 517: "resolve complex alpine terrain": In this case, not the terrain itself, but the land-use spatial variability needs to be resolved.
Correct. We'll update this in the manuscript.

Appendix

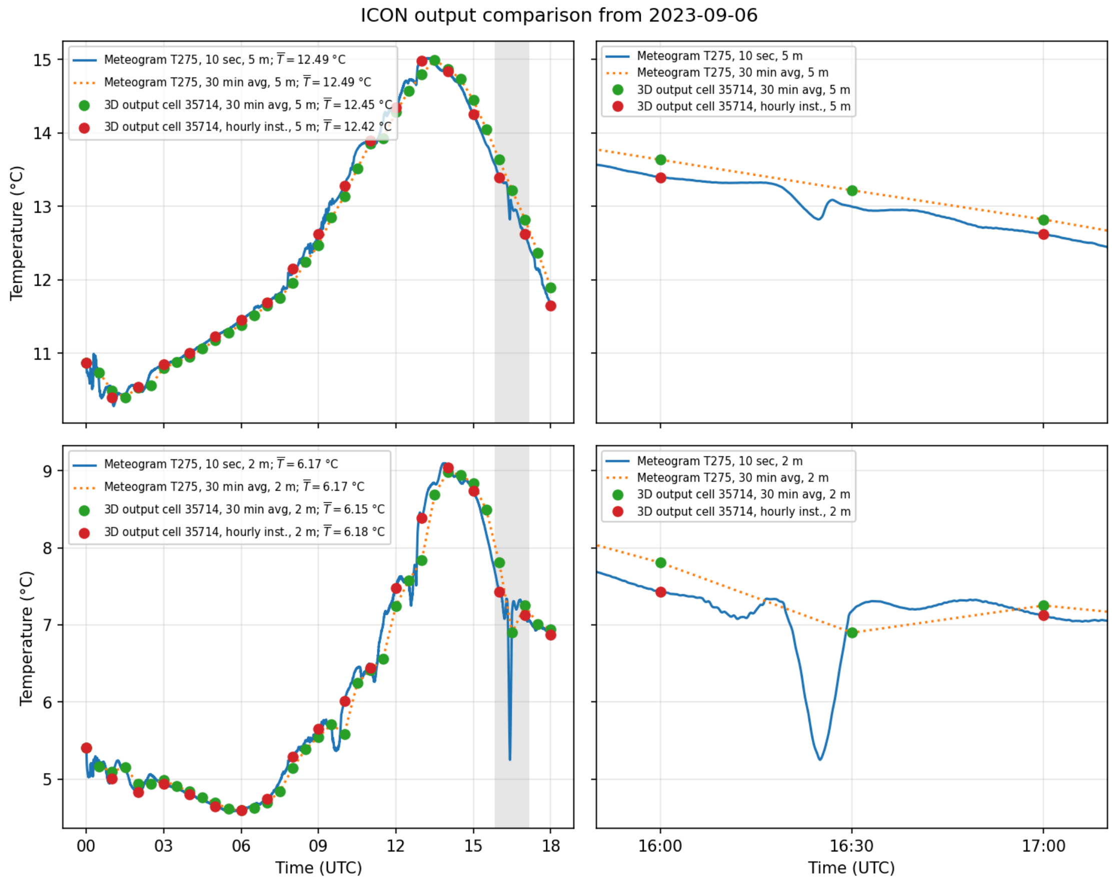


Fig. 1: Comparison of different ICON outputs for station T275. Left panels show air temperature at 5 m or 2 m, respectively. The right panels show selected time period (grey shading in left panels) with higher resolution. Each panel shows 10 sec meteogram output, 30-min resampling of meteogram output, 30-min averaged direct ICON output, and hourly instantaneous ICON output.

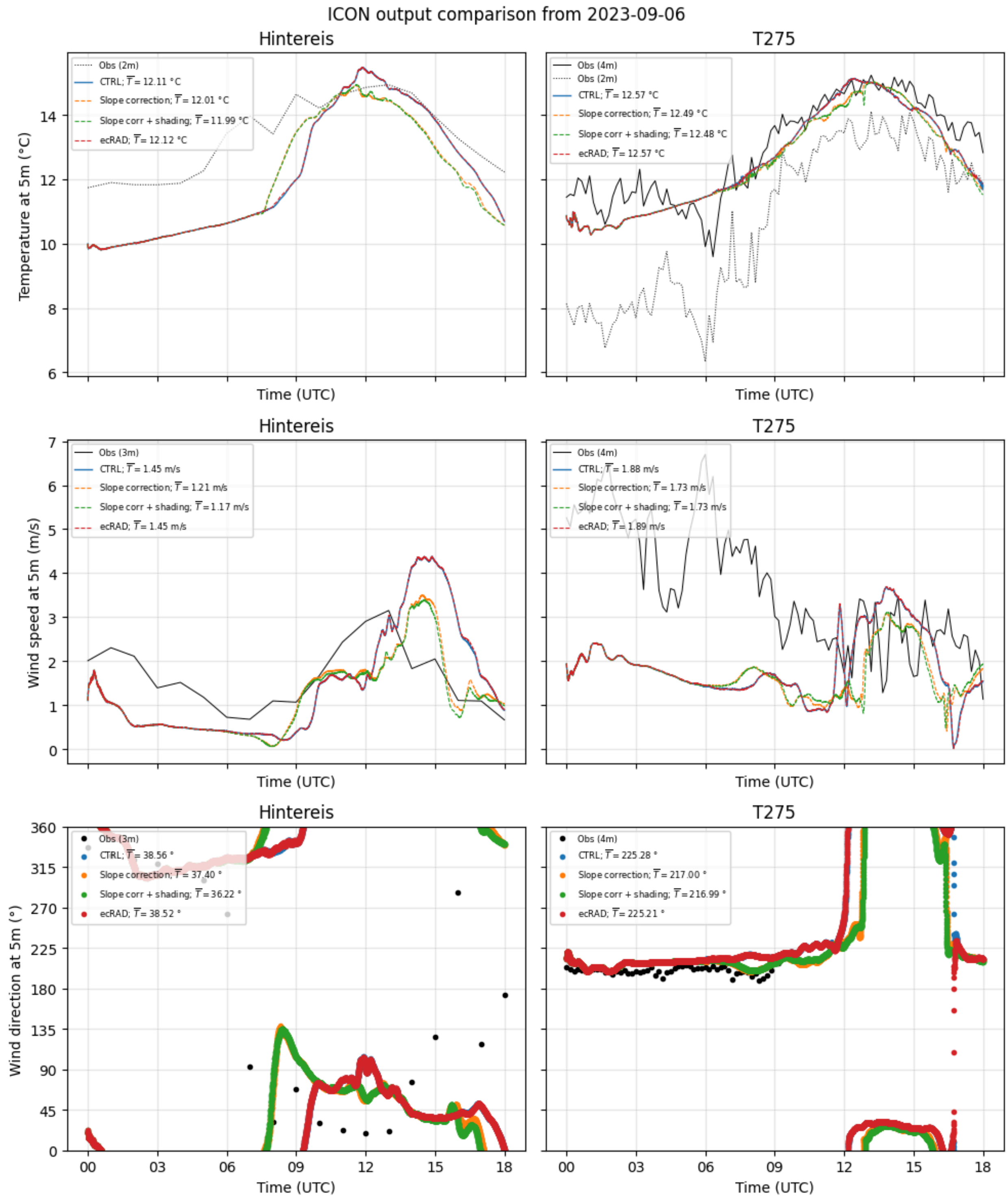


Fig. 2: Comparison of different namelist configurations for 06 September 2023, a clear-sky day. The panels show the diurnal cycle of temperature (first row), wind speed (second row), and wind direction (third row) at location Hintereis (first column) and tower T275 (second column).

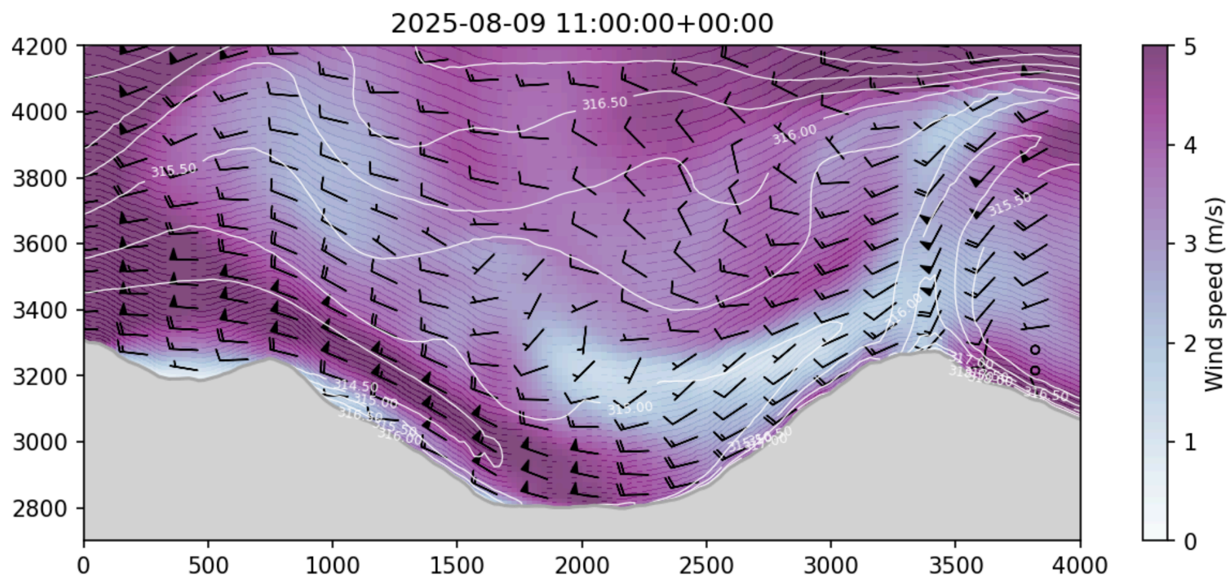


Fig. 3: ICON Transect corresponding to the 115° lidar RHI scan line from 09 August 2025 showing wind profile along the transect for the valley and neighbouring ridge lines. The isentropes are shown in white.

References

Goger, Brigitta, Ivana Stiperski, Lindsey Nicholson, and Tobias Sauter. "Large-eddy Simulations of the Atmospheric Boundary Layer over an Alpine Glacier: Impact of Synoptic Flow Direction and Governing Processes." Wagner. *Quarterly Journal of the Royal Meteorological Society* 148, no. 744 (2022): 1319–43. https://doi.org/10.1002/qj.4263.

Lehner, Manuela, Claudia Acquistapace, Marco Arpagaus, et al. "The TEAMx Observational Campaign." Version 1. Preprint, arXiv, 2026. https://doi.org/10.48550/ARXIV.2606.12152.

Mott, Rebecca, Ivana Stiperski, and Lindsey Nicholson. "Spatio-Temporal Flow Variations Driving Heat Exchange Processes at a Mountain Glacier." *The*

Cryosphere 14, no. 12 (2020): 4699–718. https://doi.org/10.5194/tc-14-4699-2020.

Nicholson, Lindsey, Ivana Stiperski, Giordano Nitti, et al. "The Second HinterEisFerner EXperiment (HEFEX II): Initial Insights into Boundary Layer Structure and Surface-Atmosphere Exchange Processes from Intensive Observations at a Valley Glacier." *Bulletin of the American Meteorological Society*, August 22, 2025, BAMS-D-24-0010.1. https://doi.org/10.1175/BAMS-D-24-0010.1.

Obleitner, F. "Climatological Features of Glacier and Valley Winds at the Hintereisferner (Ötztal Alps, Austria)." *Theoretical and Applied Climatology* 49, no. 4 (1994): 225–39. https://doi.org/10.1007/BF00867462.

Raschendorfer, M. *The New Turbulence Parameterization of LM*. Newsletter No. 1. COSMO News Letter. Consortium for Small-Scale Modelling, 2001.