Review of 'Comparison of temperature and wind between ground-based remote sensing observations and NWP model profiles in complex topography: the Meiringen campaign' by Bugnard et al.submitted to Atmospheric Measurement Techniques

March 27, 2024

The study investigates wind and temperature in the boundary layer of the Haslital in Switzerland using ground-based remote sensing and in situ instruments deployed during the Meiringen campaign from October 2021 to August 2022. It compares the observations of a microwave radiometer, Doppler lidar, and surface meteorological station to the COSMO-1E model analysis (KENDA-1). The valley is rather narrow with 1.5 km width and one of the sites is affected by the low altitude Brünigpass to the north. By means of monthly composites, nighttime inversion and thermally driven wind systems were detected. Large model errors were found for nighttime temperatures on the average as well as on individual days.

This review is of the revised submission of the manuscript and is provided in view of the interactive public discussion. In the revised version, many of the comments of previous reviewers are addressed. For example, the manuscript was shortened and the appendix was moved to a supplemental. The gathered data provide a great opportunity to investigate the boundary layer conditions in a mid-sized Alpine valley and to evaluate the performance of KENDA-1. In my opinion, the issues with partial data assimilation at MER is of particular interest and may be relevant for other locations as well, leading to potential forecast improvements in complex terrain. While I believe that the manuscript is improved compared to the original version, I still have some major concerns about some of the aspects of the study, some new, some already raised by the previous reviewers. For example, the authors strongly focus on thermally driven flows, however, they don't distinguish between thermally and synoptically driven conditions when computing the monthly composites, which, in my opinion, masks many of the relevant features. KENDA-1 has a 1.1 km horizontal resolution, while the Haslital is 1.5 km wide. The authors discuss the difference in valley floor height between reality and the model, but they do not discuss the terrain in the model at all, such as shape and depth of the valley, and to what extent the Brünigpass is resolved or how many grid points are available in the valley. I provide more details on these two aspects as well as many others in my comments below. Overall, I think the manuscripts includes too much description and not enough analysis in its present form. I suggest that the authors consider my comments before the manuscript can be accepted for publication.

1 General comments

1. The manuscripts contains a very detailed description of conditions in the results section, but hardly any investigation of the relevant processes. It first presents a lengthy description of the temperature

and wind field using monthly composites with lots of details on values in specific layers etc, followed by a description of a three-day case study when the flow from the Brünigpass affected the conditions in the Haslital and by a description of three foehn events. In my opinion, it currently is a mix between a campaign overview and some preliminary results. It is very descriptive without a clear story line. If the case studies are meant as teasers (for future manuscripts) they could be much shorter. No in-depth analysis of the case studies is presented and there are many open question, which arise to the reader. For example, why should a warm upvalley wind from the Sarneraatal descend into the Haslital? An in-depth analysis would probably be beyond the scope of the manuscript, but such contradictions to theory and open question should at least be mentioned. Also, I don't think that there is enough observational proof presented to conclude that the wind conditions in the Haslital are affected by the flow from the Brünigpass or that there is a vortex present. These findings should be formulated as hypotheses. A more detailed study, possibly using the 3D model output is needed to provide strong evidence for this and to investigate the physical processes.

2. The analysis focuses very much on thermally driven flows, but the monthly composites are not separated for days that are dominated by large-scale conditions (frontal passages, foehn) or are affected by clouds (unfavorable for thermally driven flows) or are dominated by fair weather conditions (favorable for thermally driven flows). Computing composites over all days masks the signal of thermally driven flows (which primarily occur under fair-weather conditions, as correctly mentioned in l. 48). The authors still interpret the weak signals in wind in the composites and conclude that e.g. no thermally driven winds were observed in December and January. However, they might still be there just not in the monthly composites. Furthermore, including everything in the averages makes the comparison to other studies (Discussion section), in which days where filtered for thermally driven flows, not very meaningful. For example, conclusions on thermally driven flows are drawn from the composites of along valley wind component in Figs. 7 and 8. There are several features that are not typical at all for thermally driven flows and differ from theory, like the persistence of upvalley flow near ridge height and the decrease in downvalley wind strength during the night. This could be a result from the projection of the flow on along valley wind direction or a result of sampling over all types of conditions. Either way this discrepancies need to be discussed and attempted to be explained.

   Monthly plots for wind direction are separated using a wind speed threshold which seems a bit arbitrary and needs more justification. It is currently not clear at what height this threshold must be met and during what time period. Also, why do the authors not distinguish days for the other composites (along-valley and across-valley wind speed, temperature) as well? Since the focus is on thermally driven flows, the analysis could also be restricted to composites of days with thermally driven flows. This would reduce the number of panels and streamline the manuscript. A better and more physical way to distinguish days with thermally driven flows from days dominated by synoptic flows could be by looking for a wind direction reversal twice per day and/or by considering cloudiness.

3. Composite plots for wind (Figs. 6, 8, 9) are presented with 10 panels per row, which makes it impossible to see any details on the time axis and to follow the detailed description for specific hours. I highly suggest to think about another way of presenting these composites plots. I understand the desire to reduce the number of figures, but this should not be done at the cost of visibility.

4. A discussion of the model terrain is needed, which goes beyond the difference in valley floor height. For the capability of the model to simulate terrain induced features, the shape and depth of the valley and the numbers of grid points is most relevant. This is not discussed at all. The Haslital is 1.5 km wide and KENDA-1 has a 1.1 km resolution. What about the Sarnaraatal? Is the Brünigpass even resolved in the model? How are the lakes resolved? How much terrain smoothing is done? In l. 74, it is stated that the grid resolution should be about 10 to 20 times higher than the relevant topographic scale to fully capture the different exchange processes. This aspect is very important when interpreting

KENDA-1 results, but is currently not considered at all.

5. Some sentences are not very clear and perhaps the authors should consider using a professional editing service to remove these language issues. Examples include, l. 202-203: Layer with higher T develops gradually from sunset to sunrise to reach monthly-related maximal T and height. l. 250-251: Globally, the measured MWR/MEE first level T are closer to the SMN/MER T than the modeled T. l. 256-257: The analysis of the negative ground T difference between MER at 590 m and BRU at 998 m (horizontal distance = 3.7 km) shows that near ground T inversions are common during the night for all months in the study. l. 269-271: The missed T inversions by KENDA-1/MEE leads to both its important overestimation of the T at ground level (Fig. 4) and its slight T underestimation between 850-1200 m (Fig. S5 for detailed examples). l. 296-297: ..., whereas flows from W-NW are measured in the rest of the profile concerned by up valley winds (see further explanation in sect. 3.3). l. 303-305: The good KENDA-1/MEE performances comprise first the influence of the foehn up to 2500 m (ws >20 km/h) as well as the presence of valley wind below 1200 m (ws <20 km/h) in March. l. 345-347: Thermally induced wind height increases with temperature, reaching 1000 m in February, 1800 m in May and up to 2000-2200 m in July and August. l: 371-373: This suggests a circular motion with North updraft winds (median vertical velocity of 1 km/h) that cross the valley at a low altitude, rise against the north facing slope and come back at higher altitude with a South downdraft component l: 599-600: This is especially the case at the end of March, when enhanced night time radiative cooling and important global solar radiation form strong inversions.

2 Specific comments

1. l. 64: The classic work of Whiteman and Doran (1993) could be referenced here.
2. l. 73-74: The classic work of Skamarock (2004) Skamarock and Klemp (2008) could be referenced here.
3. l. 84-87: Very complicated sentence. Please rephrase.
4. l. 88: Is the "first objective of the campaign to study the seasonal and diurnal cycles" or is this the first objective of **this study**?
5. l. 93: I think a short introduction on what KENDA-1 is and why its evaluation is important needs to be added to the introduction (possibly where NWP models are mentioned).
6. l. 94: I find the acronyms MER for Meiringen and MEE for Unterbach a bit unfortunate, since they are very similar and easy to mix. Maybe just a matter of taste.
7. l. 104: Why are the times not given in local time? This is advantageous for a study that focuses on thermally driven diurnal wind systems. At a minimum the time difference between local time and UTC needs to be given.
8. l. 105-106: The temporal resolution of the observations is higher than 1 hour. How is that considered? Are these values averaged to 1h values before computing monthly composites or are instantaneous hourly values used? When computing the composites how is data availability considered? Are model data only plotted for times and heights where observational data are available?
9. l. 110-112: The Gadmertal and Rychenbachtal are not shown in the map and not relevant for the study. Please remove or modify the map in Fig. 1 to include them. This was already a comment of a previous reviewer.

10. Fig. 1: Fig. 1b is impossible to see. Please increase line width, axis labels, and legend to make it readable. I suggest flipping the x-axis, so that the location of the sites is consistent with 1a and 1c. Why are the dots in Fig. 1b not at the valley floor? Are the station heights in Fig. 1b from the model? In Fig. 1c, labels along VW and across VW are very hard to read.

11. Section 2.2: This section sounds like COSMO-1E is used: "The NWP model used in the study is the limited-area non-hydrostatic atmospheric model from the Consortium for Small-Scale Modeling Model (COSMO)". I think it would be better to make clear from the beginning that the analysis KENDA-1 is used, which uses 1-h forecasts from COSMO-1D and observations. It would also be helpful to mention here how close to the investigation area the different observations are assimilated. SMN is assimilated at MER (this should already be mentioned here). But what about the sondes and profilers? How far away are they from the investigation area?

12. l. 141ff: Where are the MER and MEE model grid cells located in the model terrain? Are they at the valley floor in the model terrain? This is more meaningful than comparing the grid cells to the real terrain, which the model does not know.

13. l. 150ff: I think the fact that MER data are assimilated in KENDA-1 is critical for the results and needs more attention. If the observations are assimilated during certains conditions (daytime) and not during others (nighttime inversions) this may affect the error distribution. In the response to reviewer 2, the authors included an example demonstrating how often MER data are rejected and argue that 'One of the reasons to compare observed and modeled data at MER is that the MER ground observations are only assimilated if the difference with the modeled data is inside a given threshold." This information as well as the threshold would be helpful to include in the manuscript. Furthermore, I think including statistics on how often MER data are assimilated during the whole campaign and at what times would be most helpful and interesting.

14. Section 2.3.2: Was the RPG neural network retrieval used? As far as I know, HATPRO-G5 comes with a surface met sensor. Are this information used in the retrieval? If not, this surface met measurements at MEE could be an additional observational source to evaluate KENDA-1. Why was this not used? What does 'line of sight of about 10 km' mean? Is this the line of sight in the direction of the low elevation scans? In which direction was this performed (up or down the valley)? Temperature biases in microwave radiometer retrieval are highly instrument dependent, arising from spectral biases or liquid nitrogen calibration. The cold bias found by Hervo et al. (2021) is not necessarily transferable. 'The instrument at MER ..' isn't the MWR installed at MEE?

15. Section 2.3.3: The scan strategy of the Doppler lidar is not clear. It sounds like the lidar did a combianation of vertical stare (every 10 min for two minutes) and DBS (every 5 min). What did it do in between? The scan in Fig. S9 looks like it is an RHI scan, this needs to be mentioned here. According to the Vaisala data sheet the maximum range for Wincube 100S is 3 km. Was this system modified to reach 12 km maximum?

16. l. 195: Heavy **liquid** precipiation ?

17. l. 189ff: None of the information on the general conditions during the campaign are considered when presenting the results on temperature and wind. How do the boundary layer conditions differ over snow covered vs. snow free ground? Does KENDA-1 performance depend on snow cover? Because of the monthly composite and snow cover lasting until mid-December this is all mixed together. How do the heat waves reflect in boundary layer conditions and KENDA-1 performance? If the information on general conditions is included, it should be considered when desribing the results.

18. l. 202-209: The description of temperature changes should be related to the boundary layer evolution. 'Layer with higher T' is the daytime boundary layer. Perhaps the authors could compute convective

boundary layer height using the parcel method. I do not see the value of presenting temporal T gradients (Fig. S4), especially not for monthly composites. It is not surprising that T increases during the day and decreases in the evening. I suggest removing this to streamline the manuscript. In Fig. S4 temporal gradients are shown, in the text vertical T gradients are discussed (C/km).

19. l. 213ff: Fig. 2b show the bias. Did the authors also investigate mean absolute error or root mean square error? Is mean absolute error also small when biases are small or are the small biases related to averaging artefacts (days with very large positive biases vs days with very large negative biases)? How are observed and modeled profiles compared? Are modeled heights interpolated to observed heights or vice versa? Since the temporal resolution of the observations is higher than KENDA-1 are they averaged before comparing to the hourly model output? This information should probably be included in Sect. 2. I assume the authors refer to the study of Hervo et al. (2021) when saying 'The cold bias between the MWR and the radio sounding could however suggests a larger error of KENDA-1.' As mentioned before, biases are instrument specific and I do not think this is a valid conclusion.

20. l. 223: Please explain why observations at MEE are compared to MER (here and later in the manuscript). These are different sites affected by different physical processes.

21. Fig. 2: Please add ridge height to panel a). To better see stablity it would be helpful to add potential temperature isentropes. For example, observed potential temperature isentropes could be added to a) and simulated potential temperature isentropes could be added to b).

22. Fig. 3: Please enhance label and legend size.

23. l. 234-239: What is the justification to use a mean environmental lapse rate of -6.5°C/km during nighttime temperature inversions? The height difference between the model grid point and real world is only one challenge when comparing observations to model output. The terrain in the model needs to be considered as well, i.e. is the grid point at the model valley floor or on the slope? What is the valley depth, etc. (also see my general comment on this).

24. Fig. 4: Are the number of samples the same for all time series, that is, are only time stamps used when all observations and model output was available?

25. l. 252ff: Using ground stations as pseudo-profiles can be affected by local impacts (e.g. solar heating during daytime, slope winds), even more since BRU is located at a pass. A brief discussion on potential error sources should be included (Whiteman and Hoch, 2014). How are temperature inversion determined? Just by computing the temperature difference between the upper and lower height and detecting negative difference? Was a minimum absolute value required? The authors compare temperature difference over layers which are not exactly the same depth and which introduce additional uncertainties. Why not compute gradients per fixed height interval? It is interesting that even though March was dominated by foehn events there is a very clear diurnal cycle in inversion frequency and amplitude. This should be discussed. In Fig. S5, there is absolutely no inversion visible in the KENDA-1 profiles. This is very strange and deserves further discussion, in my opinion. Furthermore, it has to be considered that the MWR profiles also have uncertainties and the smoothed shape of the profile may lead to an overestimation of inversion amplitude.

26. Fig. 5: I find the amplitude plots confusing. I assume that sample size is not constant during the day and some of the spikes during summer daytime in the MWR data are probably caused by averaging over very few days. Perhaps a minimum sample size should be required for showing the amplitudes. The grid lines do not fit to the tick labels on the x-axis. What is meant by '10m spaced vectors'?

27. l. 272ff: As mentioned before, I find this issue with assimilation very interesting. Do the large differences between the observations and the 1h forecasts mean that there is no inversion in the model? What about later forecast hours? Is this maybe related to spin up time?

28. l. 288-289: Wind speed values at which height, time, and station are used to distinguish between days dominated by thermally driven and synoptically driven flows? As mentioned in my general comment, it would be more physical to inspect days for the typical reversal of wind direction twice a day instead of looking at thresholds.

29. l. 295-297: Do the authors have an explanation why the upvalley wind shifts with height? Could this be a result of averaging over a variety of conditions?

30. l. 297ff: What is the purpose of providing this amount of detail on synoptic flows? Consider removing or shortening to streamline the manuscript.

31. Fig. 6: Please add ridge height to the plots. It is really hard to see anything in the panels. Consider limiting y-axis to 2500 m (like for T plots) and rearranging the figures (maybe remove synoptically driven days). What is the temporal resolution of the DWL plots? The same as for the model? Were the winds averaged to 1 hour before computing the monthly composites? What is the point of showing panel c? It is not discussed here.

32. l. 307ff: I don't see a good agreement in November. Also there are quite large difference between the observations and the model in summer/spring. The features described here (e.g. N flows from Brünigpass) are hardly visible in Fig. 6b in its current presentation. The conclusion 'This feature is mostly caused by the KENDA-1/MEE cell overlapping the slope towards the Brünig Pass so that winds at the junction between Haslital and Sarneraatal can influence the median modeled wind compounds.' is not valid in my opinion without a detailed analysis of the model terrain.

33. l. 324-325: Typically, downvalley winds gain in strength throughout the night since the driving horizontal temperature and pressure gradients strengthen. Do the authors have any explanations why this is different here? Could this be a result of clouds forming during the night or a sampling issue? Please discuss this contradiction to theory.

34. l. 328-334: The described upvalley wind characteristics are not typical. For example, why should the upvalley wind near the surface be stronger and more regular than at 200 m above the ground. This should be discussed. Could this be related to sampling or an artifact of the projection to along valley direction? Would the valley wind system show more typical characteristics when synoptically dominated winds were excluded from the composites? The difference in onset time of downvalley wind at MER and MEE could be related to the Brünigpass. However, this needs to be formulated as hypothesis without sound observational evidence (which is not given in Sect. 3.3).

35. l. 335-336: The statement that turbulence is leading to daytime varying wind direction is not obvious and needs to be supported by observational evidence or removed.

36. Fig. 7: Please enlarge axis labels. Replace MWR/MEE in caption with DWR/MEE. What values are plotted? Are these hourly values or monthly values? If monthly why not plot hourly values to show day to day variability?

37. l. 338: 7b instead of 7c.

38. l. 340-343: This atypical behavior of the valley wind (upvalley wind during the night, upvalley wind during all day in winter) in KENDA-1 is possibly a result of averaging over all types of conditions. It is hence not clear if KENDA-1 struggles with thermally driven flows or channeling events or on clear or cloudy days. Filtering and focusing on specific conditions would be beneficial to learn more about model deficiencies.

39. l. 344-354: In my opinion, it is not valid to draw conclusions for up- and downvalley winds and on impacts of synoptic flows from the monthly composites, since the composites are most likely strongly

affected by synoptic winds and clouds. To draw meaningful conclusions on up- and downvalley winds, the days need to be filtered for conditions favorable for thermally driven winds. Also, some of the aspects (decrease of downvalley wind speed with height) could be a result of projecting on the valley axis and should be investigated. Without investigating other factors (clouds, synoptics, sampling size, etc) it is not valid to attribute varying wind direction during daytime to turbulence.

40. Fig. 8: How are the composites computed? Are KENDA-1 data only used when observations are available and valid? If not, the comparison is not fair. Are white gaps in panel a) due to small wind speed or missing data? A different color should be chosen for missing data (e.g. grey).

41. l. 366: What does 'data are scarce' mean? Shouldn't the sample size for across valley wind be the same as for along valley wind?

42. l. 366-367: The northerly flows in January and February are not clearly visible.

43. l. 370: What are north-facing slope winds? Do the authors mean downslope winds from the slopes north of the valley (they are south-facing)? For clarification, it would be helpful to repeat the colors used in Fig. 1 to distinguish southerly and northerly cross valley flow.

44. l. 371-374: In my opinion there is not enough evidence for a cross valley vortex from the example RHI plot in Fig. S9. I can see downslope and upslope components, but no closed circulation.

45. l. 375: What color are the winds that descend from Brünigpass? Are they red? Please specify.

46. l. 380-383: Why is the along valley wind mentioned here? This section is about cross valley. It needs appropriate context and reference.

47. Fig. 9: How are the composites computed? Are KENDA-1 data only used when observations are available and valid? If not the comparison is not fair. Are white gaps in panel a) due to small wind speed or missing data? A different color should be chosen for missing data (e.g. grey). What is the sample size at each point (can some of the noise be explained by varying sample size)?

48. l. 394: On this fair weather day, wind speeds of 25-30km/h are reported. How does this fit to the filter of 20km/h to distinguish thermally from synoptically driven days?

49. l. 399-401: The outflow from Brünigpass cannot be thermally driven. Why should warm air during the day descend from the pass to MEE? If there is upvalley wind in Sarneraatal that reaches the pass and descends on the south side of Brünigpass there must be a dynamic effect driving this (wave, etc). Please explain.

50. l. 409-418: This whole paragraph is based on Fig. S8. If this figure is so important, it needs to be included in the manuscript.

51. l. 419ff: It is essential to include the model terrain in this analysis to understand how the model sees the Sarneraatal and the pass.

52. Fig. 10: Both panels should have the same y-axis range.

53. l. 439-441: Specify where clear weather conditions can be expected. Describe the foehn characteristics at MER (direction, over which ridge it is coming). What stations are used to compute the foehn index?

54. l. 442: No June episode shown.

55. l. 445 and Fig. 11: Unless foehn starts always at the same time of the day, the composites should be shown relative to foehn onset and not for hour of the day.

56. l. 453ff: This whole paragraph is again based on Fig. S11 and S12 from the supplement. If this is discussed in so much detail, it needs to be included. However, related to my comment on adding more focus to the paper, I think the whole discussion on foehn should be much shortened or even removed. I also cannot see a T gradient in the types of plots in the supplement.

57. l. 465-466: I don't see the point in comparing different heights at MEE and MER.

58. l. 467-484: I think this description of wind is way too detailed and distracting and could be much shortened or removed.

59. l. 485-494: Given that KENDA-1 provides 3D output, the foehn cases could instead be investigated in KENDA-1 to understand the spatial differences and the model errors during foehn events.

60. Discussion: The discussion section is too long in my opinion and should be more focused. For example, a lengthy comparison to inversions and thermally driven flows in other studies is shown, but given that the composites in the present study are not filtered for thermally driven flows this comparison is not very meaningful. The evaluation of KENDA-1 was done visually based on time-height sections and not by computing model skills. This would have been more meaningful, instead of the descriptive comparison of the model to the observations. I think saying that 'KENDA-1 proposes good monthly median values' (l. 662) and 'Despite the complex topography around MER and the induced elevation bias, the modeled climatology of ground T is comparable to standard verification results' (l. 595-597) is hence not sufficiently supported by the analysis.

61. l. 574-584: The vortex in the Inn Valley was caused by the valley curvature (Babić et al., 2021). This means that the mechanisms here are likely not comparable and caution is advised.

62. l. 585-587: No skills are computed for KENDA-1 and an objective verification of model skills was not done. This conclusion is hence not supported by the presented results.

63. l. 595-597: The reference to standard verification results is confusing and needs to be explained and justified by results.

64. l. 599-600: Since monthly composites are shown this statement is not supported by results.

65. l. 605-617: RH depends on T. Thus, a warm bias leads to a dry RH bias. The statement on RH does not provide additional information. Bias in terms of specific humidity would be more meaningful. What does 'artifacts from the NWP can be expected under conditions favorable to surface T-inversion' mean? The statement 'Finally, the differences with observations can also originate from a modeled ongoing turbulent mixing whereas in reality a cold pool with a full or partial decoupling from the above flow is present in the valley.' is not supported by results.

66. l. 618ff: MWR liquid nitrogen calibration plays a role in MWR profiles biases and should be discussed. Average differences are discussed, but what about individual profiles?

67. Conclusions: Since many readers only read the Summary some basic information on sites and data should be repeated. I don't think that all conclusions are sufficiently supported by the results (e.g. l. 688-690, l. 693-695). I furthermore do not think it is fair to say that the study 'deepens our consensual knowledge about atmospheric phenomena in complex topography'. It is mostly a description of conditions without any in depth investigation of processes.

References

Babić, N., Adler, B., Gohm, A., Kalthoff, N., Haid, M., Lehner, M., Ladstätter, P., and Rotach, M. W.: Cross-valley vortices in the Inn valley, Austria: Structure, evolution and governing force imbalances, Quarterly Journal of the Royal Meteorological Society, 147, 3835–3861, 2021.

Skamarock, W. C.: Evaluating mesoscale NWP models using kinetic energy spectra, Mon. Wea. Rev., 132, 3019–3032, doi:10.1175/MWR2830.1, URL http://journals.ametsoc.org/doi/10.1175/MWR2830.1, 2004.

Skamarock, W. C. and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, J. Comput. Phys., 227, 3465–3485, doi:10.1016/j.jcp.2007.01.037, URL https://linkinghub.elsevier.com/retrieve/pii/S0021999107000459, 2008.

Whiteman, C. D. and Doran, J. C.: The relationship between overlying synoptic-scale flows and winds within a valley, J. Appl. Meteor., 32, 1669–1682, 1993.

Whiteman, C. D. and Hoch, S. W.: Pseudovertical temperature profiles in a broad valley from lines of temperature sensors on sidewalls, J. Appl. Meteor. Climatol., 53, 2430–2437, doi:10.1175/JAMC-D-14-0177.1, 2014.