Response to Reviewer #1 and Reviewer #2

of "A spatial evaluation of high-resolution wind fields from empirical and dynamical modeling in hilly and mountainous terrain"
by C Schlager, G Kirchengast, J Fuchsberger, Alexander Kann, and Heimo Truhetz.
Submitted to GMD, October 2018;

We thank the Reviewers again very much for the valuable and quite detailed feedback to our manuscript. We carefully considered all comments and made due effort to account for the concerns expressed; and we think it really helped improving the comprehensibility and quality of the text and how we convey the findings. We also would like to thank the Reviewers for the care also related to remaining typos and spelling mistakes. We corrected in line with all of these suggestions.
Comments by the Reviewer are black upright*, our responses blue italic. (Page and line numbers used in our responses below refer to the revised manuscript; to make this clear they are quoted like "now p10 L20-25")*

Response to Reviewer # 1 from interactive discussion

Answerers to your Major comments:

1) The term "dynamical modelling" is repeated through the manuscript, and even in the title. I think this expression it is not very common in the Regional Climate Modelling literature. This term seems to combine two more common expressions: "regional climate modelling" and "dynamical downscaling". Both are used in the literature more or less interchangeably, but I think "dynamical modelling" is not generally used. The reason for this is that, technically, a Global Circulation Model is also dynamical modelling, but I'm sure the authors do not mean this type of model. Therefore, I would advise to stick to one of the two aforementioned alternatives.
Answer: Thank you for this hint, we carefully rechecked our usage of the general term "dynamical modeling" (in the sense of empirical modeling vs. dynamical modeling) and replaced it by a more specific term where we see needed, such as "regional climatemodeling" or "dynamical high-resolution climate modeling" and so (now p1 L3, p1 L18, p2 L12, p2 L28, p2 L31, p10 L22).

2) The authors refer to two former publications (Schlanger et al. 2017, 2018) wherethe WPG seems to be further described. I acknowledge that I didn't read these publications, but it is not clear to me what this article improves or how it complements the formers. I think putting emphasis somewhere in the introduction on what new is-sues/questions this new article tries to address, compared to the formers, would help to frame this work and to better justify why it is necessary.
Answer: OK, we agree that the introduction about the ongoing work described in this article in relation to the two former articles gains from more context. Therefore, we have included a relevant paragraph in the introduction to clarify how this article complements the formers (now p2 L23-28).

3) The INCA dataset assimilates observations. Then this dataset is compared/validated with respect to the WPG, which are also observations. Are they the same? Are WPG observations assimilated to produce INCA? I assume not, as otherwise there would be an important circularity issue:
Answer: No, INCA does not assimilated WegenerNet data; and indeed we intentionally keep them independent just to avoid such circularity issues, yes. Having said this and based on rechecking our related description we agree, though, that the description of which station measurements are used in which model is a bit vague. We therefore improved this a bit to make clear that observations from ZAMG stations (the ones used in the INCA analyses) are not used as model input for the WPG and vice versa that the INCA just uses observations from ZAMG stations, but not from WegenerNet stations as model input (now p4 L29-31, p6 L13-16).

4) I'm not sure what is meant by a "wind event". I understand that the criteria in Table 2 is applied on an hourly basis, right? Are then the events hourly-based, i.e. a given hour might be included as a calm event, while the next one might be included as strong? Or do the authors select for instance the whole day when at least a single hour within the day meet the criteria? Another way of posing this question is, are there as many events as hours within each period.
Answer: Thanks for this comment which led us to notice that some further information regarding data selection would be helpful. To implement this, we modified some text passages in section 3.1 (Events for wind field evaluation). For example, the number of events shown in Table 2 corresponds to hourly events, but the data selection method for thermally induced wind events differs from the method for selecting strong wind events. In general, the data selection for thermally induced wind events is based on daytime and nighttime mean values as indicated by subscripts (dm, nm) in Table 2, which are also explained in the footnotes of this table. If a day was selected as autochthonous day, all 24 hours from this day were used for evaluating thermally induced wind events, i.e., such a day contributions 24 hourly wind events. In case of strong wind events, we compared the hourly mean values from the datasets with the hourly thresholds defined in Table 2 (hm subscripts). If the hourly mean value is larger than the defined threshold, this event is used for evaluating strong wind speeds. We explicitly included a further line of footnote to Table 2, making clear that the "Number of events" column denotes hourly wind events as the basis for the statistical analysis, and modified the text (now p8 L2-4, p8 L17-22).

5) Another detail I could not understand is how the WFSS is calculated for different spatial scales. Is the data interpolated onto successive grids with coarser resolution?
Answer: The calculation for different spatial scales is performed for defined neighbor-hood sizes, which have to be odd integers. A neighborhood size (n) defines the side length of a square, which is moved as sliding window over the dataset (e.g., n=5 corresponds to a neighborhoods size of 500 m at a spatial resolution of 100 m, and the square hence contains 25 grid points). We calculated the WFSS values for neighbor-hood sizes from n=1 to n=2N - 1, where N is the number of grid points of the largest domain size from the WegenerNet FBR or the WegenerNet JBT. The maximum domains size of 2N - 1 was used to ensure that the sliding window is large enough to always encompass the whole domain at every position - as a consequence, the fractions inside the domain are guaranteed to be the same at all locations within the domain and further enlarging the neighborhood will not change the WFSS value. We added additional text to section 3.2 (Statistical evaluation methods) to explain the calculation for different neighborhood sizes (now p8 L33-34, p9 L1-3).

Answerers to your Minor comments:

1) The abstract is in my opinion longer than necessary. For instance, between lines 5 and 10 a great amount of details are given about the datasets. This level of detail is overwhelming at this early point of the paper, and distracts the reader from the main conclusions of the manuscript
Answer: Ok, we agree that the abstract gives too detailed information. We therefore reduced the level of detail regarding the explanation of the two meteorological station networks at the beginning and also somewhat the discussion of the results at the end of the abstract (now p1 L7-8, P1 L21-25).

2) Pag 2, Line 9: course-resolution→coarse-resolution
Answer: OK, done (now p2 L5)

3) Pag 2, Line 15: "data fusion". I think a more precise term is "data assimilation" or" assimilation of observations".
Answer: Ok, we now use "data assimilation" instead of "data fusion (now p2 L11)

4) Pag 2, Line 19: "dynamical regional climate models"→"regional climate models".
Answer: OK, done (now p2 L12)

5) Pag 3, Lines 3-8 These two paragraphs read as a summary of the methodology. I do not think this is necessary in the introduction.
Answer: Thank you for this hint. We agree, that this information is also given in section "3 Evaluation events and methods" and we therefore removed these two paragraphs from the introduction.

6) Pag 3, Line 10: I was not aware of the concept "two penalty problem". Therefore I was puzzled to read this without either a reference or a couple of lines that briefly summarise what is the deal with this. It is explained later, so I would advise to bring those explanations already here.
Answer: Thank you for this hint. We moved the explanation regarding the "double penalty" to this paragraph to immediately explain this kind of penalty (now p3 L5-9).

7) Pag 4, Line 6:"eleven"→11 (for consistency reasons with the way this is reported for FBR)
Answer: OK, done (now p4 L6).

8) Pag 4: Lines 20-26: Is it really necessary this amount of detail about how the data about temperature and humidity is produced for this system, given that these fields are not used in the manuscript?
Answer: Thank you for this hint, we agree that the gridded fields of temperature, precipitation, and relative humidity are not so relevant for this manuscript. We therefore removed the (too) specific description parts about the lapse rate and the different interpolation methods for the generation of these fields.

9) Pag 5, Lines 15-16: "Therefore the output shows errors in regions with low station density" The model resolution does not imply that there are larger errors in areas with low station

density. Why would it be the case? The validation is more difficult, but it could be that the model does a good job. We just don't know.
Answer: Thanks, we agree that this statement is not correct at this position of the text. We therefore removed this sentence and modified the text in this paragraph to ensure, that the statement is related to the INCA analysis algorithm and not to the RCM's first guess (now p6 L11-12).

10) Pag 5, Line 22: The number of vertical levels in the RCM (not only the driving dataset) is an important parameter worth to mention.
Answer: The COSMO-CLM simulations are provided for 40 vertical levels. The first level is simulated for 10 m above ground and the last level corresponds to the 100 hPa level, whereby the vertical resolution is higher for the boundary layer and decreases towards to the top level. Based on your comment and a note of Referee 2 we now give detailed information about the model characteristics of all models (see also 2) from the Responses to Referee #2: under the Major comments) (WN: p4 L33-34, p5 L1-8), (INCA: now p5 L26-33, p6 L1-8), (CCLM: now p6 L27-34, p7 L1-14).

11) Pag 7, Line 23: the units (m s-1) should not be italic. This applies to several locations through the manuscript. Please review them.
Answer: Ok, corrected (now p9 L11, p10 L10, p10 L26, p11 L1, p11 L2, p11 L3, p13 L14, p13 L15, p13 L25, p13 L26, p13 L27, p13 L34, p14 L12).

12) Pag 8, Line 15 says that wind speeds are systematically underestimated. This is curious, as normally models tend to overestimate wind speed. Indeed, in the conclusions (Page 13, Line 19) this is noted when it is stated that wind speed are overestimated in both types of events. Isn't this contradictory? Please clarify the details.
Answer: Thank for noticing this. The statement "systematically underestimated" is not fully correct in the context of what we try to address in the corresponding section (Pag 9, not Pag 8). In this section we are explaining the behavior of the WFSS for selected wind events and not for event-averaged statistical results. Therefore, the underestimation by the COSMO-CLM model explained in the text refers to a single event. We corrected the corresponding sentence (now p11 L6-8).

13) Page 10, Line 21: "fundamentally able". Do the authors mean "unable"?
Answer: Yes changed it to "unable" (now p12 L18).

14) A bottleneck of WFSS is that it does not allow to disentangle if low skill is driven by problems with wind speed or direction. However in Pag 10, from lines 29, this is somehow solved, and low skill is attributed to errors to these two variables separately. But it is not obvious how these conclusions can be drawn from the shown figures. Is this based on an analysis that is not shown in the manuscript?
Answer: It has to be noted that the WFSS can also be used to separately evaluate the two wind components, for example by classifying the datasets just based on wind direction. In general, the definition of the classes should reflect what a user wants to verify. We used the advantage of the WFSS and evaluated wind speed and wind direction in a combined way. Regarding the separate evaluation of both variables in relation to Fig. 3, we agree that this conclusion cannot be drawn by simply interpreting this figure and that additional information

is needed. The behavior of the influence of wind speed or wind direction on the WFSS is indicated by the results of error measures additionally calculated by traditional statistical methods. These are summarized in Table 4 and the generated mean wind speed bias distribution map, illustrated in Fig.4. Furthermore, we visually interpreted the windroses for most of the events (the windroses for all events are not shown in this manuscript, Fig. 2 just shows windroses for selected events, for good illustration). To make clear of how we draw this conclusion, we modified the corresponding text passages and refer there to the results calculated by traditional methods. The spatial displacement and the biases for the INCAvsWN_therm_JBT case are mainly caused by the differences in wind directions for these thermally induced wind events, indicated by the large mean absolute error of wind direction (MAEdir) (Table 4) sentence (now p12 L32-34).

15) Page 12, Line 21: where→were
Answer: Ok, done (now p14 L26).

16) The conclusions are overly long. They review every single detail of the results and after reading them is not obvious what are the take-home messages. I advise to summarise the conclusions to leave the most important and general conclusions, those that can be exported to other studies/regions.
Answer: Thank you for your advice, we agree that the conclusion gives too detailed information, which especially applies to the discussion of the results. We therefore summarized the explanation of the results and shortly discuss what's relevant for ongoing next steps of work and other studies/regions. In addition, we now provide additional information on possible model improvements (now p15 L19-26, p16 L1-6, p16 L10-12).

17) This may seem as a tiny detail, but the fact that the panels in Fig. 1 do not follow the expected order (a, then b, finally c) puzzled me for a couple of minutes until I realized that FBR (labelled b, and firstly described in the text) is actually the last panel of the figure. Perhaps a trivial re-ordering of the panels following a more intuitive order might facilitate the reading.
Answer: Thank you, we agree that the panel sequence and the corresponding labeling is a bit confusing. We therefore moved the FBR panel to the top of Fig. 1 and labeled it with (a), and the JBT panel to the bottom and labeled it with (b). Furthermore, the overview in the middle of Fig. 1 is denoted as middle panel in the text; the discuss-panels is (a) and (b) and so everything is clear.

Response to Reviewer # 2 from interactive discussion

Answerers to your Major comments:

1) Firstly, it feels that the modeling approaches and the CALMET regridding are just presented as is, with no critical discussions of the pros and cons of the methodologies and how they could affect the analysis here.
Answer: Thank you for this hint, we reconsidered the description about advantages and disadvantages of the different modeling approaches. With regard to the empirical modeling approach, we referred only to former publications and agree, that additional information on this modeling approach should be given in the text. Also the description about the INCA and the CCLM (we now use CCLM instead of COSMO-CLM for the sake of simplicity) model needed to be improved, especially with regard to internal numerical settings and the lateral boundary. We therefore added additional text to the model data sections 2.2, 2.2 and 2.3 (now p4 L33-34, p5 L1-8, p5 L26-33, p6 L1-8, p6 L27-33, p7 L1-14).

With regard to the CALMET re-gridding, the CALMET-based wind fields were not resampled in order to avoid information losses in these high-resolution data. The coarser INCA and CCLM data were resampled and mapped onto the high-resolution WPG grid. In addition, we have performed sensitivity tests for different interpolation methods and found no significant changes in the statistical results. (See paragraph on page 5 from line 31 to 34). We reconsidered also our description related to this; we think that this particular description about the re-gridding of the data is already detailed enough (now p6 L17-20, p7 L15-17).

2) The COSMO model in particular is somewhat of a mystery and there is no speculation as to what the model may be doing wrong to have poorer performance, beyond just saying it is not high enough resolution (even though 1 to 3 km is not that big of a jump). Given the different behavior of the two regimes, the question that sparks most for me is that may be COSMO is poorer at simulating the wind profiles of 'thermal events' versus' strong wind events'. This is particularly pertinent to the study since the conclusions are that we need more observations and no evidence is shown that we may need better models. Thermal events are potentially complex interplays between differential heating and turbulence, which ultimately lead to the wind profile and yet none of the thermodynamic (or even wind) structures are examined from the model to understand this. So, in general an elaboration of the models' shortcomings is needed and more interpretation beyond just a description of the comparison, as this will inform model improvements which I presume is the end goal here.
Answer: We agree, that these flow patterns are influenced by complex interplays of thermodynamic structures. The model behavior of CCLM is also very complex and disentangling the various influences would far exceed the scope of this study. Therefore, at this point, we can only come up with more speculative interpretations.
Based on recent discussions with our internal RCM experts and with ZAMG model developers, we have come to the conclusion that the main argument for medium-term model improvements lies indeed in higher-spatial-resolution simulations. What has not yet been mentioned in the manuscript is that the CCLM model uses an advection scheme, which causes additional smoothing of the terrain. The scheme is implemented to avoid numerical instability, but its diffusion damping causes an effective resolution, which is quite lower than in the INCA model. This ultimately leads to quite low spatial variability in the CCLM wind fields and may explain the high uncertainty in the modelled wind directions, especially under weak synoptic forcing. In addition, flow patterns may significantly divert from the observations. Due to the orographic smoothing, flow-over patterns occur more frequently

than flow-around patterns. However, if flow-over patterns occur more frequently, the influence of the orographic speed-up effect (Taylor et al., 1987) becomes more dominant. In contrast, if mountains and hills are higher, more flow-around patterns and flow-splitting patterns occur, which are favoring even negative orographic speed-up effects (Hewer, 1998). This might be the reason for the overestimation of the wind speed and its improvement under strong wind conditions in FBR. In JBT, however, the underrepresentation of the orography becomes even more striking. The central mountain in this region in CCLM is about 500 m lower than in INCA. This gives a severe deformation of the CCLM wind field and clearly indicates the requirement for improving the treatment of orography in high-resolution simulations. In principal, the ALARO model suffers from similar shortcomings. However, since the model's output is corrected with the help of station data, the wind fields in INCA are much better in agreement with WegenerNet data than CCLM.
Beside higher-resolution simulations, improvements in the CCLM can be expected from using a newly developed advection scheme that allows to circumvent the horizontal diffusive damping. If actually higher-resolution models were evaluated, however, the topographic shading through the terrain becomes increasingly important, especially for the simulation of thermally induced wind events. Such methods are not implemented in the ALARO and were switched off in the CCLM model for the generation of the data used in this study.
Other influences on wind are: (1) misleading land cover properties (e.g., of the roughness lengths), (2) underestimation of land cover heterogeneity, (3) the negligence of the so-called zero-plane displacement (Oke, 2009), and (4) no use of a 3D turbulence parameterization, based for example on large eddy simulations.

We address now these model limitations and possible improvements in the text (now p13 L28-31, p14 L1-2, p14 L18-20, p15 L20-26, p16 L3-6, p16 10-16).

Answerers to your Minor comments:

Abstract:

1 1.14: ‘skill scores’:
Answer: Ok, done (now p1 L12).

2 1.14: 1.16-18: I found the ordering of this confusing:
Answer: Thank you for this hint. We considered to change ordering of the description related to the model intercomparisons; but to be consistent with the defined evaluation cases (INCAvsWN_xxxx_FBR, CCLMvsWN_xxxx_FBR; see Table 1, we preferred it’s better to keep the existing ordering in the text (see text between INCA and WegenerNet than between CCLM and WegenerNet wind fields).

31.14: 1.24: Even if the thermal events are ‘strong events’?:
Answer: A criterion for selecting a day as autochthonous day, which includes thermally induced wind events is generally weak wind speeds (see Table 2). Therefore, the sample of strong wind events in the thermally induced cases is too small, and no statement can be made as to whether a model is better for such strong events under autochthonous weather conditions. Specifically, CCLMvsWN_therm_FBR does not contain strong wind events, INCAvsWN_therm_FBR contains seven strong wind events, CCLMvsINCA_therm_JBT does

not contain strong wind events, and for the INCAvsWN_therm_JBT case we estimated just 16 strong events.

1. Introduction:

5) 2.7: What's the definition of surface wind here – 10 m?
Answer: This statement refers to the first levels within the PBL, which are influenced by the terrain.

6) 2.9: This is potentially possible it just won't be high resolution. And how does it hamper interpolation?
*Answer: Thank you for this hint; with this statement we refer to high-resolution wind field modeling on a regional to local scale. To make clear that the generation of **realistic** high-resolution wind fields is not possible with coarse-resolution models or by an interpolation of wind station data, we modified the text (now p2 L4-6).*

7) 2.28: Are the WegenerNet fields used as part of the INCA analysis and to also validate INCA?
Answer: No, to avoid circularity issues INCA does not use any WegenerNet data and vice versa no INCA data are used in the WPG. Due to the vague description of which data are used in which model and a comment from Referee 1, improved the text related to this (now p4 L29-31, p6 L13-16)

8) 3.7: Given you are referring to COLSMO-CLM as a climate model, I am unsure how to think of actual synoptically overlapping periods with WNet?
The COSMO model in climate mode implements several new features compared to the original COSMO weather model. For example, the vegetation state of soil is not assumed to be constant, or it is able to use not only initial values but also dynamic boundary data. The CCLM simulations where generated during the course of a previous study and cover the period Jan.2006 - Dec. 2009, and they were constrained at synoptic scale by assimilated ECMWF IFS fields – we improved and added more detailed CCLM description (for page and line references see point 2 under Major comments).

9) 3.12: 'and provide'
Ok, will change to "and provide" (now p3 L9).

2. Study Areas and Model Data:

10) 3.26: Sensitive in that it has already experience change?
Yes, in this region climate change is already measurable. For example observational based studies show a strong summer temperature trend of 0.7 °C per decade (Kabas et al. 2011, Hohmann et al.2018).

11) 3.30: Could elaborate a bit her. Katabatic winds, turbulent PBL,...
Answer: Thank you for this hint; we added additional text and use now the term "drainage wind" to refer to small-scale flows (now p3 L28-31).

12) 4.10: Are not both regions subject to synoptic weather conditions given their close proximity?
Answer: Yes, both regions are subject to synoptic weather conditions. With "westerly-flow synoptic weather conditions" we refer to general weather conditions, which lead to airflows with prevailing westerly wind directions and strong wind speeds at higher altitudes in the WegenerNet JBT. In the WegenerNet FBR, the damping effect of the thermal stratification on synoptic winds is larger, which cause a low amplitude between the month with the average strongest winds and the month with the average lowest winds.

13) 4.22: Are there dangers in interpolating both relative humidity and temperature separately since one is a non-linear function of the other, due to saturation temperature being a non-linear function of T?
Answer: Thank you for this comment. The gridded fields of temperature, precipitation, and relative humidity are not used as model input (which uses station data) and are therefore not relevant for this manuscript. For this reason, and because of a comment from Referee 1, we removed the description parts about how these fields are generated (please see also point 8 in the response to Referee #1).

14) 4.28: What are the meteorological fields used? Does this actually include explicit wind observations and what vertical levels are used?
Answer: The main purpose of the generated meteorological fields is to investigate weather and climate as well as evaluating RCMs (please see Page 1, lines 23-26).
Yes, the CALMET model used in the WPG generates mean wind fields based on observed wind speed and wind direction from the WegnerNet stations, among other needs. The INCA system assimilates data from the ZAMG stations.
In this study, we are using the mean wind fields at 10 m height for the model intercomparisions. We've rechecked the manuscript related to this, and noticed, that this important information of which height level is used for the model intercomparisons was missing, so we therefore added additional text related to this to the manuscript (now p4 L32, p, p6 L18, p7 L15).

15) 5.25-29: This is a little confusing here. Do you mean the COSMO model is driven continuously by ECMWF on the domain boundaries for 2008-2010, and you are describing the time stepping numerics? Also, what are setting 'based on shallow convection'?
Answer: For detailed information about numerical settings and driving data see point 2) in the responses to major comments above, where we also give the page and line numbers of the new text passages for a more detailed model description.

3. Evaluation Events and Methods:

16) 6.11: 'autochthounous' I had to look this up! But I am still not sure what is being referred to.
Answer: Weather conditions that are determined by local or regional daily variations in temperature or pressure are referred to as autochthonous conditions. Such conditions are mostly caused in cases of low synoptic influences, by anti-cyclonic weather conditions and favors thermally induced flows.

17) 6.10-15: Is there any presumption of diurnal variations here?
Answer: The selection of autochthonous days is based only on the comparison of daytime and nighttime averages and no assumptions were made regarding daily variations (see page 6, lines 17-29 and Table 2). The results of this method show good agreement with another study, where such days have been manually selected (Oberth, U., 2010: Untersuchung der lokalen Windsysteme im Raum Feldbach unter besonderer Berücksichtigung von Kaltluftabflüssen. (in German). Master theses, 146 pp. [Available online at http://www.wegenernet.org/misc/MA_Oberth_2010_WegernerNet_Wind.pdf].)

18) 6.20: 'daily global radiation'? surface solar?
Answer: Depending on the region, we used the observed global radiation or net radiation as input for the selection method (See paragraph on page 6 from line 17 to 26 and Table 2).

19) 6.30: These 'thermal wind events' have not really been defined yet.
Answer: Thank you for this hint, we re-checked the description and added additional text to ensure what is meant by thermally induced wind events (now p7 L27-28).

20) 6.31-34: Is this the only criteria for the 'strong wind events'. Given it is large scale synoptic would it be more meaningful to have an area coherence footprint or temporal longevity criteria.
Answer: Thank you for this hint; we have noticed that important information on another criterion is missing both in the text and in Table 2.
In order to determine the weather situation during prolonged weather with strong winds, the respective days were selected on the basis of the daily average wind speed. Subsequently we have chosen the hourly events from these days. We therefore added additional text to the corresponding paragraph. Furthermore, we added the additional limit-values for the selection of these days to Table 2.
And yes, further improvements in the selection of such days can be expected through the use of e.g. longevity criteria or frontal detection methods, but these were not applied during the course of this study since considered beyond the scope of due efforts to this end.

21) 7.7: Is this to reduce penalty in both space and time?
Answer: No, the FSS is a spatial and not a spatiotemporal verification metric.

4 Results:

22) Fig 2: This is very confusing indeed. Are these just snapshots of a particular day, even a specific hour, given the time stamp at the top of each plot?
Answer: Yes, this Figure illustrates just single one-hour events, indicated by the hourly period at the top of each plot. We agree that especially the labeling of the hourly periods is somewhat confusing. We improved this labeling (changed for example from 7/29/2009 04:00:00 PM-17:00:00 to 29.07.2009 16:00:00-17:00:00). Furthermore, we have adapted the color map for the representation of the three wind classes from the windroses to the one of the ten classes from the wind fields.

23) 7.33-35: I do not understand this at all. 'Ensemble of events'??

Answer: The event-averaged score values are calculated based on averaging the one-hour event WFSS values over all the hourly events for a specific case. In this study we calculated eight event-averaged score values which are shown in Figure 3. To make this clear, we modified the accompanying text. Moreover, we have noticed that we refer to this value as case-averaged score value here (page 7 line 32) and as event-averaged score value in all other parts of the text. We now uniformly refer to this parameter as event-averaged score value.

24) 8.21: You're implying here that Alpine pumping is a local phenomenon that arises due to local forcings topography. However, wouldn't you expect a model to do well at this if it is simply forced by the analyzed wind at its boundaries?
Answer: Thank you for this hint; we imply here that Alpine pumping is a regional, not a local phenomenon. In contrast to thermally induced local winds, this phenomenon leads to compensating flows on a regional scale, which are called Randgebirgswind and its counterpart, the Antirandgebirgswind. Especially in case of autochthonous weather conditions, the Antirandgebirgswind is influencing the WegenerNet FBR in the afternoon. In our former studies we have evaluated the WPG also for such conditions and found good results, which is mainly due to the dense station network with wind observations. Due to the fact that alpine pumping is a very complex process and INCA has only two station observations available in the WegenerNet FBR, we did not expect any specific results about the quality of the simulated wind fields for such conditions. The analyses of the INCA fields shows, that INCA is able to adequately simulate the significant wind pattern of the Antirandgebirgswind, which affects not only the ridges of the hills but also the valleys in the WegenerNet FBR.

We added additional information on the spatial scale of the Antirandgebirgswind to the corresponding text passage (now p10 L10-12).

25) 8.34: But wasn't this the less challenging terrain compared to the other region?
Answer: In general, in this section we describe the characteristics of example wind fields and the model results for representative hourly events for each evaluation case, first for the WegenerNet FBR and then for the WegenerNet JBT. The results of each individual evaluation case are described and then compared with them from a corresponding case within the same region. In this specific paragraph we describe the results for the CCLMvsWN_therm_FBR case, which are then compared with the INCAvsWN_therm_FBR case. Both cases are defined for thermally induced wind events, which correspond to the WegenerNet FBR. We have rechecked this paragraph and recognized, that corresponding evaluation case definition for the CCLM case is not mentioned in the text and added therefore additional text (now p10 L22).

26) 9.1-10: Although the wind roses do give a good summary of the biases in wind direction, the key thing to understanding the differences of course is the synoptic distribution over the domain. This shows that INCA is not southerly enough mostly in the southern part of the domain. Is this explainable from this perspective?
Answer: We agree that the wind roses in combination with the wind fields give a good intuitive notion how well the INCA wind field matched the WegenerNet field. In this specific example, the large AWFSS and therefore small bias in wind classes over the whole domain is not reflected by the wind classification result shown in the windroses. In this example we

are trying to show the advantage of calculating the WFSS based on an azimuthal class rotation (for explanation of class rotation please see page 8 lines 1-10). A calculation of the WFSS without rotating the classes would lead to a poor AWFSS of about 0.6 (instead of >0.97).
We also agree, that the low WFSS at small neighborhood sizes is mainly caused by the differences in wind sectors, especially in the southern part of the domain. Furthermore, parts of the area differ in wind speed classes. To illustrate this in the text, we added information to the corresponding paragraph (now p10 L30-31).

27) 9.11-12: This is really surprising given that COSMO is all yellow/orange whereas the other fields are seeing weaker speed values in the greens.
Answer: Thank you for this hint. In this particular case, we indeed (inadvertently) used the wrong wind speed limits to create the wind rose. We also re-checked the code for the calculation of the WFSS and could confirm the correct limits are implemented here. We adjusted the lower-middle panel of Fig. 2b and the corresponding text in the manuscript (now p11 L1-5).

28) 9.16: 1th??
Answer: We changed from 1th of August 2012 to 1st of August 2012 (now p11 L9).

29) 9.24-29: This needs more interpretation here. What aspect of the dynamical model is failing? Is it the solution itself or is it the synoptic setup? Why does 8/1/2012 mostly succeed but this day fail?
Answer: Also in this case, we mainly attribute the uniform wind directions simulated with the CCLM to the too strongly smoothed terrain in the model. For such events under low synoptic forcing, both wind fields show a too low spatial variability in wind direction. Regarding wind speed, the INCA wind field shows some variability with higher wind speeds in parts of the summit regions compared to wind speeds at lower altitude. Furthermore, a valley wind in the Enns valley becomes obvious. Probably the analysis part of the INCA model leads to a somewhat better representation of the wind field. We added additional text to draw attention to such effects (now p11 L18-26).

Could you please indicate what you mean with "Why does 8/1/2012 mostly succeed but this day fail"? We checked through the text but were not sure what's meant. For this CCLMvsINCA_therm_JBT event we are analyzing the 31th of Mai 2008 from 13:00-14:00.

30) 10.6: Won't this always be true of COSMO in these synoptic circumstances? However, the scale of the features for the high wind regions here are actually above the coarser grid scales of COSMO, so this lack of resolution reasoning is not correct is it?
Answer: For the WegenerNet FBR the wind fields are systematically overestimated which become obvious in the CCLMvsWN_strong_FBR case and in the statistical evaluation results (cf. also Fig. 4). For the WegenerNet JBT the low wind speeds are probably explained by negative orographic speed-up effects (Hewer, 1998) caused by a too smoothed terrain, compared to the WegnerNet FBR, where speed-up effects are leading to stronger wind speeds. For a detailed information about this speed-up effects see also point 2) in the responses to Major comments above. We added additional text about such effects to the 4.2

Statistical evaluation results section (for the indication of pages and lines of this text see also point 33 below, which deals with a similar question).

31) 10.21: Unable instead of able??
Answer: Yes, ok, done (now p12 L18).

32) 10.25: Again, though isn't this the simpler terrain region?
Answer: Yes, here we describe the performance of the CCLM in comparison to the INCA model for strong wind speeds for the hilly WegenerNet FBR. The influence of the terrain (e.g. channeling of air flow through the valleys) on the synoptic flow field is smaller in this hilly region than in the WegenerNet JBT region. That's why the CCLM shows similar performance as the INCA model despite the lower resolution for this region.

33) Fig 4: It is very surprising that the COSMO model has a widespread systematic bias over the simpler FBR region, but a much reduced systematic bias in general over the much more complex terrain of the JBT region.
Answer: Thank you for this hint; the difference in these bias values between the two regions is probably again attributed to the speed up effects. For more information please see point 2) in the response to Major comments above. Furthermore, it has to be noted that in comparison to the WegenerNet FBR region the INCA data and not the WegenerNet data were used as reference for the evaluation of the CCLM, due to missing WegenerNet data. Since the CCLM wind fields show small bias values for thermally induced wind events, compared to the INCA wind fields, similar results as in the CCLMvsINCA_therm_JBT case can be expected for a comparison of CCLM with WegenerNet data. In case of strong wind events, the intercomparison of the CCLM with the INCA model shows opposite patterns than the INCAvsWN_strong_JBT case, but with smaller bias values. Therefore the same bias values in attenuated form are to be expected for a comparison of the CCLM with WegenerNet data. We added this information to the text (now p14 L8-9, p14 L18-20).

Further changes in the manuscript

1) We now use the abbreviation CCLM instead of COSMO-CLM in the text and in all figures.

2) We have separated section "2.3 INCA and COSMO-CLM" data into "2.3 INCA data and 2.4 CCLM data".

3) Changed "138°E-17°E" to "13.8°E-17°E" (now p6 L10).

4) Corrected "at a defined station locations" to "at defined station locations" (now p8 L7).

5) Corrected "indicates fair weather conditions" to "indicate fair weather conditions" (now p8 L15).

6) Corrected "2N + 1" to "2N – 1" (now p9 L16).

7) Corrected “underpining” to “underpinning” (now p16 L7).

8) Corrected “expect” to “except” (now p16 L3).

9) Added additional text to the Acknowledgements section (now p17 L7-14).

10) Figure 3, (a) and (b): Changed “INCA resolution” to “INCA grid resolution” and “COSMO resolution” to “CCLM grid resolution”.

11) Table 2: Corrected number of events for CCLMvsWN_therm_FBR from 1632 to 264.

A spatial evaluation of high-resolution wind fields from empirical and dynamical modeling in hilly and mountainous terrain

Christoph Schlager[1], Gottfried Kirchengast[1,2], Juergen Fuchsberger[1], Alexander Kann[3], and Heimo Truhetz[1]

[1]Wegener Center for Climate and Global Change (WEGC), University of Graz, Graz, Austria.
[2]Institute for Geophysics, Astrophysics, and Meteorology/Institute of Physics, University of Graz, Graz, Austria.
[3]Department of Forecasting Models, Central Institute for Meteorology and Geodynamics (ZAMG), Vienna, Austria.

Correspondence: Christoph Schlager (christoph.schlager@uni-graz.at)

Abstract. Empirical high-resolution surface wind fields, automatically generated by a weather diagnostic application, the WegenerNet Wind Product Generator (WPG), were intercompared with wind field analysis data from the Integrated Nowcasting through Comprehensive Analysis (INCA) system and with ~~dynamical~~ regional climate model wind field data from the ~~non-hydrostatic climate model COSMO-CLM~~Consortium for Small Scale Modeling Model in Climate Mode (CCLM). The INCA analysis fields are available at a horizontal grid spacing of 1 km × 1 km, whereas the ~~COSMO model~~ CCLM fields are from simulations at a 3 km × 3 km grid. The WPG, developed by Schlager et al. (2017, 2018), generates diagnostic fields at a high resolution grid of 100 m × 100 m, using observations from two dense meteorological station networks: The WegenerNet Feldbach Region (FBR)~~and its alpine sister network, the WegenerNet Johnsbachtal (JBT). The high-density WegenerNet FBR is located in southeastern Styria, Austria, ~~, located in a region predominated by a hilly terrain and ~~small differences in altitude. The networkconsists of more than 150 meteorological stations. The WegenerNet JBT contains eleven meteorological stations at elevations ranging from about 600 m to 2200 m~~ its alpine sister network, the WegenerNet Johnsbachtal (JBT), located in a mountainous region~~in northern Styria~~.

The wind fields of these different empirical/dynamical modeling approaches were intercompared for thermally induced and strong wind events, using hourly temporal resolutions as supplied by the WPG, with the focus on evaluating spatial differences and displacements between the different datasets. For this comparison, a novel neighborhood-based spatial wind verification methodology based on fractions skill ~~socres~~ scores (FSS) is used to estimate the modeling performances. All comparisons show an increasing FSS with increasing neighborhood size. In general, the spatial verification indicates a better statistical agreement for the hilly WegenerNet FBR than for the mountainous WegenerNet JBT. The results for the WegenerNet FBR show a better agreement between INCA and WegenerNet than between ~~COSMO~~ CCLM and WegenerNet wind fields, especially for large scales (neighborhoods). In particular, ~~COSMO-CLM~~ CCLM clearly underperforms in case of thermally induced wind events. For the JBT region, all spatial comparisons indicate little overlap at small neighborhood sizes and in general large biases of wind vectors occur between the ~~dynamical (COSMO~~regional climate model (CCLM) and analysis (INCA) fields and the diagnostic (WegenerNet) reference dataset.

Furthermore, gridpoint-based error measures were calculated for the same evaluation cases. The statistical agreement, estimated for the vector-mean wind speed and wind directions show again a better agreement for the WegenerNet FBR than for the

WegenerNet JBT region. ~~In general, the difference between modeled and observed wind directions is smaller for strong wind speed events than for thermally induced ones.~~ A combined examination of all spatial and gridpoint-based error measures shows that ~~COSMO-CLM~~ CCLM with its limited horizontal resolution of 3 km $\times$ 3 km and hence, a too smoothed orography, is not able to represent small-scale wind patterns. The results for the JBT region indicate ~~that~~ significant biases in the INCA analysis fields~~generally overestimate wind speeds in the summit regions. For~~ , especially for strong wind speed events~~the wind speed in the valleys is underestimated by INCA, however~~. Regarding the WegenerNet diagnostic wind fields, the statistics show ~~decent~~ acceptable performance in the FBR and somewhat overestimated wind speeds for strong wind speed events in the Enns valley of the JBT region.

1 Introduction

Surface wind is often considered as one of the most difficult meteorological variables to model, particularly over areas of complex terrain like the Alps (Whiteman, 2000; Sfetsos, 2002; Abdel-Aal et al., 2009; Gómez-Navarro et al., 2015). Therefore~~a realistic wind field modeling with course-resolution models is not possible and moreover hampers~~ , realistic high-resolution wind fields cannot be generated with coarse-resolution models or by a simple interpolation of wind station data onto regular grids. Innovation in computer sciences, new methods in weather analysis or nowcasting models, advanced software architectures used in regional climate models (RCMs) and the growing power of computers meanwhile led to highly-resolved outputs from such models at the 1-km scale (Awan et al., 2011; Suklitsch et al., 2011; Prein et al., 2013b, 2015; Leutwyler et al., 2016; Kendon et al., 2017).

These models, however, contain various limitations and sources of uncertainties. In case of weather analysis fields, which is a mixed empirical and dynamical modeling from data ~~fusion~~assimilation, they result from too little meteorological station and remote sensing data, and in case of ~~dynamical~~ regional climate models (RCMs) they include deviations in the driving data set, physical and numerical approximations, as well as parameterizations of processes at the sub-grid scale (Gómez-Navarro et al., 2015).

To evaluate and improve these analysis and models, meteorological observations and especially gridded empirical datasets at high spatial and temporal resolutions are needed. The model outputs on their side generally represent the involved processes as areal averages rather than on a point-scale (Osborn and Hulme, 1998; Prein et al., 2015). Therefore, gridded meteorological evaluation datasets, with each (aggregated) grid value being a best-estimate average of the grid cell observations, are the most appropriate evaluation datasets (Haylock et al., 2008; Haiden et al., 2011; Hiebl and Frei, 2016).

To investigate weather and climate on a local 1-km scale as well as evaluating RCMs, the Wegener Center at the University of Graz operates two high-resolution meteorological station networks (Fig. 1): the very high-density WegenerNet Feldbach

Region (FBR) in southeastern Styria, Austria (Fig. 1b) and the high-density WegenerNet Johnsbachtal (JBT) in northern Styria, Austria (Fig. 1c); details introduced in section 2 below.

For both networks, diagnostic windfields at a high resolution grid of 100 m × 100 m are generated by a weather diagnostic application, the WegenerNet Wind Product Generator (WPG). Schlager et al. (2017) introduced the WPG and its performance evaluation for the WegenerNet FBR, which was then advanced by Schlager et al. (2018) to the WegenerNet JBT region and a longer-term evaluation in both the FBR and JBT regions. Jointly these studies established the level of quality of the empirical WPG wind fields. In this study, we ~~intercompare~~ now make use of these empirical high-resolution wind fields as reference data in order to intercompare them with empirical-dynamical wind field analysis data and with dynamical regional climate model data.

In this study, we intercompare the empirical WegenerNet wind fields (Schlager et al., 2017, 2018) with empirical-dynamical wind field analysis data from the Integrated Nowcasting through Comprehensive Analyses (INCA) (Haiden et al., 2011) system and with ~~dynamical~~ regional climate model data from the Consortium for Small Scale Modeling ~~(COSMO)~~ Model in Climate Mode (~~CLM) (Schättler et al., 2016)~~CCLM) (Böhm et al., 2006; Rockel et al., 2008). The intercomparisons aim at getting useful and robust information about performance limits for these empirical and dynamical modeling approaches for regions with very different topographic characteristics and weather situations. Furthermore, we co-analyze the impact of different horizontal resolutions, which inevitably will always be a challenge for the wide diversity of data products typically available.

~~In order to provide more detailed information about the INCA modeling performance, we evaluate the INCA wind fields against the WPG wind fields from the WegenerNet FBR and the WegenerNet JBT for representative types of weather events.~~

~~To analyze the impact of the spatial resolution for the WegenerNet FBR, we use again the WPG wind fields as reference data for the evaluation of the COSMO-CLM wind fields. Due to missing overlapping COSMO-CLM and WegenerNet data from the same period in the WegenerNet JBT, the evaluation of the COSMO-CLM wind fields in this region is performed using INCA datasets.~~

Besides traditional gridpoint based verification methods, we use a novel wind verification methodology, recently developed by Skok and Hladnik (2018). This neighborhood-based spatial verification method avoids the "double penalty" problem and can distinguish forecasts depending on the spatial displacement of wind patterns (Skok and Hladnik, 2018). A 'double penalty' problem arises when using traditional statistical methods for datasets which contain an offset between the modeled and the reference data. In that case, the modeled data are penalized twice: first, for simulating an event where it did not occur and second, for failing to simulate an event where it did actually occur (Roberts, 2008; Prein et al., 2013a; Skok and Hladnik, 2018) . So our primary motivation of this study is indeed to explore ~~an~~ and provide improved insight, by careful intercomparisons the relative performance strength and weakness of empirical an dynamical wind field modeling at high-spatial resolution over complex terrain where actual wind station observations will generally be available at sparse station density.

The paper is structured as follows. Section 2 provides a description of the study areas and basic information about the model data. Section 3 presents defined evaluation cases and the methodology for the automatic selection of typical wind events followed by a description of the methods used to evaluate model results. In the following section 4 results are presented and discussed in detail. Finally, in Section 5 we summarize our results and draw our conclusions.

2 Study Areas and Model Data

2.1 Study Areas

The first study area, the WegenerNet FBR (indicated by the lower white-filled rectangle in Fig. 1a, enlarged in b) lies in the Alpine foreland of southeastern Styria, Austria, centered near the town of Feldbach (46.93°N, 15.90°E). It covers a dense grid of 154 meteorological stations within an area of about 22 km × 16 km, in a hilly terrain, characterized by small differences in altitude (Kirchengast et al., 2014). The typical difference in altitude between the valleys and the crests is about 100 m and the highest peak is the Gleichenberg Kogel, with an elevation of 598 m.

This region, with a more Alpine climate at the valley floors and more Mediterranean climate along hillsides is quite sensitive to climate change (Wakonigg, 1978; Kabas et al., 2011; Hohmann et al., 2018). Furthermore, it exhibits rich weather variability, especially through strong convective activity and severe weather in summer (Kirchengast et al., 2014; Kann et al., 2015a; O et al., 2017, 2018; Schroeer and Kirchengast, 2018). The wind fields in this study area are characterized by thermally induced local flows and influenced by thermally-driven regional wind systems with weak wind speeds, caused by a dynamical process called Alpine pumping (Lugauer and Winkler, 2005). Furthermore, nocturnal drainage winds, which are leading to cold air pockets, are relevant for this region, which is dominated by agriculture. Especially in fall and winter, the nocturnal cold air production is amplified by temperature inversions in relation to high-pressure weather conditions. In the WegenerNet FBR, hillside locations are thermally preferred to valley locations at night. Results related to the WPG-diagnosed empirical wind fields in the WegenerNet FBR can be found in Schlager et al. (2017, 2018).

The second study area, the WegenerNet JBT (indicated by the upper white-filled rectangle in Fig. 1a, enlarged in c), is named after the Johnsbachtal river basin (location of village Johnsbach 47.54°N, 14.58°E) and situated in the eastern Alpine region, in the *Ennstaler Alps* and the *Gesäuse National Park*, in the northern Styria, Austria. The terrain of this mountainous region is characterized by large differences in elevation. The Hochtor, with an elevation of 2369 m, is the highest summit, and the valleys are roughly at a height from 600 m to 800 m (Strasser et al., 2013; Schlager et al., 2018). This region spans an area of about 16 km ×17 km and comprises ~~eleven~~ 11 irregularly distributed meteorological stations including two summit stations at altitudes of 2.191 m and 1.969 m (Schlager et al., 2018).

The climate is Alpine with mean annual temperatures of around 8 °C to 0 °C and an annual precipitation of about 1.500 mm to 1.800 mm from the valley to the summit regions (Wakonigg, 1978; Prettenthaler et al., 2010). Typical for this region are thermally induced local flows and westerly-flow synoptic weather conditions. Details related to first studies and their results as well as to the cooperation and partnerships can be found in Strasser et al. (2013), and in most up-to-date form in Schlager et al. (2018). Recently, Schlager et al. (2018) computed and evaluated WPG-generated empirical wind fields in this region.

2.2 WegenerNet data

The data acquired from the two WegenerNet regions FBR and JBT are automatically quality controlled and processed by the WegenerNet Processing System (WPS), consisting of four subsystems (Kirchengast et al., 2014): The Command Receive Archiving System transfers raw measurement data via wireless transmission to the WegenerNet database in Graz, the Quality

Control System checks the data quality, the Data Product Generator (DPG) generates regular station time series and gridded fields of weather and climate products, and the Visualization and Information System offers the data to users via the WegenerNet data portal (www.wegenernet.org).

Besides weather and climate time series, the DPG generates, based on a spatial interpolation of the station observations, gridded fields of the variables temperature, precipitation and relative humidity for the WegenerNet FBR. ~~In case of temperature, lapse rates are estimated from the temperature observations at the numerous different station altitudes. These lapse rates are used for generating temperature fields over the hilly terrain. For temperature and relative humidity fields, inverse-distance weighted interpolation is used, and for precipitation fields, inverse-distance-squared weighted interpolation is used. The~~ These gridded products of the WegenerNet FBR are available to users in near-real time with a latency of about 1-2 hours. Kirchengast et al. (2014) and Kabas (2012) provide detailed information about the subsystems of the WPG.

The DPG furthermore includes a newly developed wind field application, the Wind Product Generator (WPG), as briefly introduced in Sect.1. The WPG provides high-resolution wind fields for the WegenerNet FBR as well as for the WegenerNet JBT. The WPG uses the freely available empirical California Meteorological Model (CALMET) as core tool and generates, based on meteorological observations, terrain elevations and information about land use, mean wind fields at 10 m and 50 m height levels with a spatial resolution of 100 m $\times$ 100 m and a temporal resolution of 30 minutes, again with a maximum latency of about 1-2 hours. In order to keep the meteorological input data of the WPG independent from the data pertaining to the other operational station networks, observations from the ZAMG stations (violet stars in Fig. 1a and Fig. 1b) and other external stations are not used as WPG input. For the WegenerNet FBR, the gridded wind fields are available starting in 2007 and for the WegenerNet JBT starting in 2012. The wind fields at 10 m height level are used for the model intercomparisons.

The CALMET model is a diagnostic model that omits time-consuming integrations of nonlinear equations, such as the governing equations of dynamical models (Truhetz, 2010; Seaman, 2000; Ratto et al., 1994). It is hence not capable of simulation of dynamic processes such as flow splitting and grid-resolved turbulence, or to deliver prognostic information. Specific parameterizations allow the model to empirically take into account conditions such as kinematic effects of terrain, slope flows, and terrain-blocking effects (Scire et al., 1998; Cox et al., 2005; Seaman, 2000). We enhanced the model by implementing methods developed by Bellasio et al. (2005) to as well take into account topographic shading through relief, topographic slope and aspect, and the sun position for the estimation of solar radiation. In addition, the modeling of temperature fields is now based on vertical temperature gradients, calculated from meteorological station observations located at different altitudes, and the influence of vegetation cover is taken into account. Details about these advanced algorithms can be found in Bellasio et al. (2005) .

The quality of the generated wind fields depends above all on the quality and the spatial and temporal resolution of the meteorological observations and surface-related datasets, which are used as model input (Schlager et al., 2017, 2018; Morales et al., 2012; C . A detailed description of the WPG application and the statistical results for the WegenerNet FBR can be found in Schlager et al. (2017). More information regarding statistical results related to the WegenerNet JBT as well as information regarding evaluation results from five-year climate data of the WegenerNet JBT in comparison to nine-year climate data from the WegenerNet FBR can be found in Schlager et al. (2018).

2.3 INCA ~~and COSMO-CLM~~ data

The INCA system has been developed at the Central Institute for Meteorology and Geodynamics (ZAMG) in Vienna, Austria, to provide realistic analyses and nowcasts of quantities of several meteorological variables for the highly mountainous and the overall complex terrain of Austria. In case of the variable wind, the system operationally generates spatially distributed analysis wind fields in 3D and for 10-m height above ground with a horizontal grid spacing of 1 km × 1 km and a temporal resolution of 1 hour.

The basic idea of the INCA wind module is to statistically correct a numerical weather prediction (NWP) model first guess (i.e., in operational mode the latest available NWP model run) with latest observational data which are not entering the NWP data assimilation. Thus, the skill of the INCA analysis depends on the station density, their representativeness and the spatial distribution of station observations, as well as on the skill of the NWP model providing the first guess. The impact of the NWP model on the analysis skill is further discussed in Kann et al. (2015b).

In this study, NWP model outputs used as first guess ~~are~~ for INCA were generated by the ~~revised version of~~ the spectral ARPEGE-ALADIN (ALARO) ~~(Wang et al., 2006) with~~ model in revised version (Wang et al., 2006). ALARO has a horizontal grid spacing of 4.8 ~~km × km~~ × 4.8 ~~km . Therefore the output shows errors in regions with low station density (Prein, 2013; Haiden et al., 2011).~~km (600 x 540 grid points) and includes 60 vertical layers up to the 2 hPa level (about 43 km altitude), covering Central Europe, Eastern France and the Northern part of the Mediterranean Sea. It is run with a temporal resolution of 180 s using a hydrostatic semi-implicit semi-Lagrange dynamical solver (Bubnová et al., 1995) and the ALARO-0 physics package, which includes the 3MT microphysics-convection scheme (Gerard and Geleyn, 2005), the ISBA force restore 2L soil scheme (Noilhan and Planton, 1989), and the ACRANEB radiation scheme (Ritter and Geleyn, 1992). Soil temperature and moisture are initialized by a 6h-cycle optimal interpolation data analysis taking into account the latest ALARO forecast as first guess and 2 m relative humidity and temperature observations from SYNOP and national stations. The 2 m values are transferred to soil variables via empirical relations (Giard and Bazile, 2000). To reduce initial spin-up a digital filter initialization is applied.

The model gets its lateral boundary and atmospheric initial conditions from the high-resolution deterministic operational global integrated forecast system (IFS) of the European Centre for Medium-range Weather Forecasts (ECMWF) model in lagged mode (i.e., ALARO 00 UTC is linked to IFS 18 UTC of the day before, ALARO 06 UTC to IFS 00 UTC, etc.). This is due to the rather late availability of the IFS data. Coupling is achieved by one-way nesting via Davies relaxation (Davies, 1976) . Sea surface temperature is interpolated from the deterministic IFS model to the ALARO grid. More details about ALARO development and configurations can be found in Termonia et al. (2018).

The INCA wind fields have already been evaluated for a moderately hilly region in the north of Austria (47.78°N–49°N, ~~138~~13.8°E–17°E), where the wind analyses ~~shows~~ show significantly higher errors compared to the statistical results from other meteorological variables (Haiden et al., 2011). These higher errors mainly root in the limited representativeness of station data, as well as on the low station density, which can be only partly compensated by INCA's analysis algorithms (Haiden et al., 2011) .

~~Regarding the COSMO-CLM (Rockel et al., 2008), the climate version of the "Lokalmodell", we~~ In the WegenerNet FBR area, INCA assimilates observations from the ZAMG Feldbach and Bad Gleichenberg stations (violet stars in Fig. 1b) to the NWP's first guess, and in the WegenerNet JBT region, observations from the ZAMG Admont station are used. However, data from WegenerNet FBR and JBT stations are not used in INCA data assimilation and hence the WPG fields can be used for independent evaluation (Haiden et al., 2011).

The coordinate system of the INCA datasets is transformed into WGS84 / UTM zone 33N coordinates. Furthermore, we resampled the wind fields at 10 m height levels from the INCA gridding onto the WegenerNet FBR and WegenerNet JBT grids, using a bilinear interpolation method. Based on extensive sensitivity tests regarding different interpolation methods, we concluded that the statistical results do not significantly depend on the interpolation method.

2.4 CCLM data

Regarding the CCLM (Rockel et al., 2008), we use available wind fields generated with the model version 5.0. These wind fields were generated during the course of a previous study and are available for the period 2008 to 2010. The data have a comparatively coarse horizontal resolution of 3 km $\times$ 3 km on an hourly basis that is nevertheless the highest resolution proper available to this study. This limited resolution leads to a smoothed orography, which may result in different wind patterns with errors in wind speed or direction. Furthermore, the winds may be displaced by an incorrect position of the topographic slopes (Skok and Hladnik, 2018; Prein et al., 2013b).

~~The numerical settings for the simulations are based on the *third order, two time-level, Runge–Kutta, split-explicit* scheme~~ The CCLM fields are provided for 40 vertical levels. The first level is simulated for 10 m above ground and the last level corresponds to the 100 hPa level (about 16 km altitude), whereby the vertical resolution is higher in the boundary layer and decreases towards to the top level.

CCLM is a non-hydrostatic model with a Runge-Kutta dynamical core, which makes use of a 3rd order scheme with diffusion damping to discretize the advection term in the compressible Euler equations (Wicker and Skamarock, 2002). In order to avoid numerical instability, the model's orography is additionally smoothed via a 10th-order Raymond (1988) filter. The vertical coordinate system is a terrain-following, time-invariant Gal-Chen pressure-based sigma coordinate (Gal-Chen and Somerville, 1975) Deep and shallow convection are parametrized following Tiedtke (1989) and turbulence is parameterized based on Mellor and Yamada (Raschendorfer, 2001). Vertical mixing comes from a prognostic formulation of the turbulent kinetic energy (TKE) with a 2.5 closure that accounts for grid- and ~~the physical settings~~ subgrid-scale water and ice clouds. It uses a statistical cloud scheme for cloud cover and cloud water content (so-called Gaussian closure scheme). Horizontal diffusion follows the Smagorinsky approach.

Land cover data for CCLM are based on ~~shallow convection. In this model setup~~ the Global Land Cover 2000 project (EEA, 2016) from SPOT4 satellite products (Bartalev et al., 2003). In the model setup used (3 km resolution), deep convection is resolved explicitly, which means that ~~the used~~ parameterization for deep convection was switched off. Shallow convection is still parameterized. In climate research, such simulations are referred to as ~~convection permitting~~ convection-permitting climate simulations (CPCSs) ~~(Prein et al., 2013b). Furthermore, the simulations are driven by the integrated forecast system (IFS) of the~~

~~European Centre for Medium-Range Weather Forecasts (ECMWF). These driving data are provided in a~~ (Prein et al., 2013a) . To minimize decoupling effects from model-internal variability, that usually occur in large model domains and if nudging techniques are not used (Kida et al., 1991), CCLM is operated in a small domain encompassing the Greater Alpine Region and it is also driven by ECMWF's IFS. The data assimilation system IFS includes a wide range of observations and is assumed to provide perfect boundary conditions with a horizontal grid spacing of about 25 ~~-~~km at mid latitudes, and ~~are calculated for~~ on 91 vertical ~~layers (Bechtold et al., 2008).~~ levels (Bechtold et al., 2008). Every 6 h (00, 06, 12, 18 UTC) of the IFS data is an analysis field from the assimilation system and every alternate 6 h (03, 09, 15, 21 UTC) is a short-range forecast field. This procedure has already been used by Suklitsch et al. (2011) and keeps the modeled synoptic patterns in agreement with the observed ones.

~~The coordinate system of the INCA and the COSMO datasets are~~ In course of the data preparation for the study, CCLM data at 10 m height level were also transformed into WGS84/UTM zone 33N coordinates~~. Furthermore, we resampled the wind fields from these two models onto the WegenerNet FBR and WegenerNet JBT grid, using a bilinear interpolation method. Based on extensive sensitivity tests regarding different interpolation methods, we concluded that the statistical results are not significantly dependent on the interpolation method~~ , resampled and mapped onto the high-resolution WPG grid, and checked for sensitivity with respect to the interpolation method which was as well found weak (see Section 2.3 above).

3 Evaluation events and methods

3.1 Events for wind field evaluation

The WegenerNet, INCA and ~~COSMO~~ CCLM wind fields are intercompared for two representative types of wind events: thermally induced wind events and strong wind events. For this purpose, we defined eight evaluation cases, four for each of the two study areas (Table 1). For the cases shown in Table 1 we use the WegenerNet data as reference, ~~expect~~ except for evaluating the ~~COSMO-CLM~~ CCLM wind fields for the WegenerNet JBT. The reason for this is, that the ~~COSMO-CLM~~ CCLM data used in this study are available from 01/2008-12/2009, but the WegenerNet JBT data are available only as of 01/2012, since this latter network was sufficiently completed for long term monitoring only since 2012 (Table 1, cases ~~COSMOvsINCA~~CCLMvsINCA_therm_JBT and ~~COSMOvsINCA~~CCLMvsINCA_strong_JBT).

In both study areas, autochthonous weather conditions mainly lead to thermally induced wind systems, meaning that the wind fields are controlled by small-scale temperature and pressure gradients. These small-scale gradients lead to characteristic interacting systems of air motion, like slope winds and mountain-valley winds, and create complex everyday flow patterns. The autochthounous days are characterized by small synoptic influences, cloudless or nearly cloudless skies, low relative humidity and increased radiation fluxes between the Earth surface and the atmosphere (Prettenthaler et al., 2010). Due to frequently occurring temperature inversions in relation to clear sky and high pressure weather conditions in winter, which often leads to a stable atmospheric stratification in the whole WegenerNet FBR and in the valley regions of the WegenerNet JBT, autochthonous days are only selected from spring, summer and fall (March to October).

The automatic selection of thermally induced and strong wind events is done based on thresholds, that we defined based on sound physical and careful sensitivity checks summarized in Table 2. For the estimation of autochthonous days, we compared the observed daytime mean values of wind speed (v) ~~,~~ and relative humidity (rh) ~~, and~~ as well as the nighttime mean values of net radiation (Q_n) from selected stations with the respective thresholds. A further criterion for the selection of such days is a high daily global radiation, which indicates fair weather conditions. For this purpose, we compared the daily mean modeled global radiation ($Q_{g,m}$) for clear sky conditions with the observed daily mean net radiation ($Q_{n,o}$) for the WegenerNet FBR and with the observed daily mean global radiation ($Q_{g,o}$) for the WegenerNet JBT at ~~a~~ defined station locations (Table 2, reference data). The reason for the comparison of $Q_{g,m}$ with $Q_{n,o}$ for the WegenerNet FBR is that this station network includes no global radiation sensors. Due to the almost linear relationship between the daytime $Q_{g,o}$ and $Q_{n,o}$ for clear sky conditions we find that the same selection method can robustly be applied to both study areas by defining different thresholds (Table 2, $Q_{g,m}$-$Q_{n(g),o}$). If all criteria are fulfilled for a given day, the data from the entire day are added to the thermally induced wind events dataset, leading to 24 hourly events (i.e., 24 hourly-mean wind speed values).

The modeling of the global radiation is done based on ESRI's ArcGIS Area Solar Radiation Tool. This tool is designed for local landscape scales and derives the incoming solar radiation based on a digital elevation model. Small differences of daily mean values between $Q_{g,m}$ and $Q_{n(g),o}$ ~~indicates~~ indicate fair weather conditions and high global radiations during the day. If all criteria are fulfilled for a given day, the data from that day are included into the thermally induced wind events dataset.

The strong wind events, caused by synoptic weather conditions such as cyclones and frontal system at larger scale, are selected on an hourly basis from preselected days, by comparing hourly mean values from ~~the~~ gridded reference datasets with defined minimum wind speeds. These preselected days were estimated by comparing the daily average wind speed from the gridded datasets with a defined minimum average wind speed (Table 2, $\overline{v}$ and v for strong wind speed cases). If the hourly-mean value of this reference dataset is larger than the defined minimum wind speed, the data of this reference dataset and the corresponding model dataset are used as part of the hourly event data for evaluating strong wind events.

3.2 Statistical evaluation methods

In order to account for spatial differences and displacements between the model and the reference data and to analyze wind speed and direction in a combined way, we apply a novel wind verification methodology. This methodology extends the Fractions Skill Score (FSS), a spatial verification metric developed by Roberts (2008), which is classified as neighborhood-based approach and originally used for verifying precipitation. The FSS is based on the assumption that a model is useful when the model data and the corresponding reference data show a similar spatial frequency of precipitation events, which alliviates the requirement of the models to predict the events at exactly correct positions, which is an unduly strict assumption. Furthermore, this metric avoids the 'double penalty' problem and provides a scale dependent information about the level of model skill (Gilleland et al., 2010; Roberts, 2008).

~~A 'double penalty' problem arises when using traditional statistical methods for datasets which contain an offset between the modeled and the reference data. In that case, the modeled data are penalized twice: first, for simulating an event where it did not occur and second, for failing to simulate an event where it did actually occur (Roberts, 2008; Prein et al., 2013a; Skok and Hladnik, 2018)~~

. In In order to obtain information of how the skill of a model varies with the spatial scale, the FSS is calculated for different neighborhood sizes. A neighborhood size of n defines a square consisting of $n \times n$ grid points, i.e., it denotes the side length of the square (e.g., for n=5 the square contains 25 grid points). These squares of defined neighborhood sizes are moved as sliding windows over the datasets, centered successively at each grid point, whereby the area outside the domain is assumed to contain no windclass. In terms of FSS value, it is a 0-to-1 normalized metric, i.e., the lower limit of the FSS is 0 and the upper limit +1, with values approaching +1 representing a better degree of model performance.

The extended version of the FSS is named the Wind Fractions Skill Sore, denoted as WFSS which has been developed by Skok and Hladnik (2018). The score is calculated based on user defined wind classes. The definition of these classes is partly subjective and can significantly affect the WFSS. The wind vector field should be classified in such a way, that the definition reflects what a user wants to analyze. For example, a complex terrain leads to strong changes in wind directions, therefore it is reasonable to define smaller class intervals regarding the wind directions. For upper level winds the focus could be more on the magnitude of wind speed.

We defined eight wind direction classes with an interval size of 45° for a range of three wind speed categories, as shown in the windroses of Fig. 2. Wind speeds $< 0.5\ m\ s^{-1}$ $< 0.5\ \mathrm{m\ s^{-1}}$ were classified *calm*, independently of the wind direction. The small interval size of 45° was chosen to be able to capture the varying wind directions in the study areas, especially for thermally induced winds. Because of the generally much lower wind speeds in WegenerNet FBR we defined a smaller interval size of the wind speed categories for this region (Fig. 2a and b, lower panels) than for the WegenerNet JBT (Fig. 2c and d, lower panels).

Table 3 includes the equations for the calculation of the WFSS and the asymptotic WFSS (AWFSS). This AWFSS will always be reached for a neighborhood size $\geq 2N + 1$ $\geq 2N - 1$, where N is the number of grid points of the largest domain size. At such a large neighborhood size, the estimated fractions within the domain are the same at all locations and further enlarging of this size will not affect the WFSS. A bias always leads to a AWFS AWFSS value < 1, which indicates systematic differences in the frequency of wind classes between the model and the reference wind classification.

The WFSS is calculated for each hourly field for the selected events. The final case-averaged event-averaged score values are then estimated calculated based on averaging these hourly one-hour event WFSS values over the full ensemble of all the hourly events within the analyzed multi-year period, for each evaluation case listed in Table 1.

As briefly explained above, the chosen thresholds of the classification and the number of classes are influencing the score. We found that especially the wind direction thresholds can have a strong impact on the score values. For example, a small change in the wind direction value from prevailing northwesterly winds, which are close to a threshold to distinguish between W and N, could dramatically change the WFSS value. Such a small error in the model data could indicate a poor modeling performance, whereas a human analyst would asses the forecast as reasonably good. To avoid this problem we calculated the hourly WFSS for every rotation between 0° and 45° with an interval size of 5° (nine trail classes), in addition to the original class definition. As next step, always the maximum values of the hourly score values at each neighborhood size are used for computing the final case-averaged score values. We applied this approach for in total 7597 selected events (Table 2, number of

events) and estimated the case-averaged score value for each of the eight evaluation cases. A more detailed description of the wind fractions skill score metric can be found in (Skok and Hladnik, 2018).

Furthermore, we applied traditional gridpoint-based statistical performance measures such as bias, root-mean-square error and others to each evaluation case. All statistical performance metrics used in this study are summarized in Table 3.

4 Results

4.1 Evaluation for selected wind events

Figure 2a-d illustrates typical examples of modeled wind fields (upper rows of panels) and the corresponding windroses of relative frequency of wind directions divided by wind speed categories (lower rows of panels) from selected representative evaluation events. Each panel depicts modeled and the associated reference data for the WegenerNet FBR (Fig. 2a and b) and the WegenerNet JBT (Fig. 2c and d), for thermally induced (Fig. 2a and c) and strong wind events (Fig. 2b and d). Figure 2e shows the WFSS values of these examples, estimated as explained in Section 3.2 above.

The thermally induced wind event on the 29th of July 2009 from 16:00-17:00 UTC for the WegenerNet FBR (Fig. 2a) shows thermally driven regional flows caused by the Alpine pumping. This flow is called Antirandgebirgswind which arises usually in the afternoon as southerly wind with maximum wind speeds of about ~~2.5 $m\ s^{-1}$~~ 2.5 m s^{-1}. The Antirandgebirgswind is a compensating flow between the bordering mountains of the eastern Alps, and the hilly countryside region of southeastern Styria (called Riedelland), which is comprising the WegenerNet FBR (Wakonigg, 1978). The INCA (upper-left panel of Fig. 2a) and the WegenerNet wind fields (upper-right panel of Fig. 2a) show a similar distribution with generally low wind speeds and prevailing southerly wind directions. The intercomparison of these INCA with WegenerNet data for this event shows the largest WFSS values for all neighborhood sizes, which indicates a good overlap of the wind classes (Fig. 2e, INCAvsWN_therm_FBR). Furthermore, it shows a nearly perfect asymptotic value of about 0.99. This large AWFSS indicates a very small bias, which is also reflected by the similar wind classification results (lower-left and lower-right panel of Fig. 2a).

The ~~COSMO~~ CCLM wind field shows similarly low wind speeds, compared to the WegenerNet wind field, but a shift in wind directions from the S sector mainly to the SE and partly to E and NE sectors between the ~~COSMO~~ CCLM and the WegenerNet data can be observed (lower-middle and lower-right panel of Fig. 2a). This shift is reflected by small WFSS values at all neighborhood sizes, especially below a scale of 10 km. The AWFSS shows the largest value of all ~~COSMO-CLM~~ CCLM intercomparisons, but is still low compared to INCA evaluation cases, which indicates a large bias ~~.~~ (Fig. 2e, CCLMvsWN_therm_FBR). Evidently, this ~~dynamical~~ regional climate wind field modeling at 3-km gridsapcing is not adequately representative for the given challenging hilly terrain.

The strong wind speed event in the WegenerNet FBR on the 30th of October 2008 from 10:00-11:00 UTC led to southwesterly to southerly winds (Fig. 2b). The INCA model data and the WegenerNet reference data show maximum 1-hour vector-mean wind speeds of around ~~9 – 10 $m\ s^{-1}$~~ 9 – 10 m s^{-1} (upper-left and right panel of Fig. 2b). Regarding the wind directions, differences in the wind sectors can be observed (lower-left and right panel of Fig. 2b). The INCA data show wind directions mainly from the SW sector (lower-left panel) while in the WegenerNet data show wind directions from the S and

SW sectors (lower-right panel). The WFSS for this case shows small values at small neighborhood sizes and increases with increasing neighborhood size (Fig. 2e, INCAvsWN_strong_FBR). These low WFSS values are mainly caused by the differences in wind direction classes, especially in the southern part of the domain and through some spatial displacements in wind speed classes. Despite low WFSS values at small neighborhood sizes caused by differences in wind sectors, the AWFSS shows a high asymptotic value (AWFSS>0.97). This high value is caused by the prevailing wind directions in the WegenerNet data, which are close to the threshold values to distinguish between S and the SW. In this case, the 5° azimuthal class rotation procedure hence avoids lower score values.

Regarding the ~~COSMO~~ CCLM data (lower-middle panel of Fig. 2b), ~~no wind is assigned~~ the whole wind field shows wind speeds from about 6.5 $\mathrm{m\ s^{-1}}$ to ~~a~~ 7.5 $\mathrm{m\ s^{-1}}$ and is therefore assigned to the wind class with wind speeds higher than ~~6 $m\ s^{-1}$,~~ ~~whereas~~ 6 $\mathrm{m\ s^{-1}}$. Whereas, for the WegenerNet wind fields, a large proportion is assigned to the ~~highest wind speed class~~ class with wind speeds from 3 $\mathrm{m\ s^{-1}}$ to 6 $\mathrm{m\ s^{-1}}$ of this region (~~lower-right panel of Fig. 2b)~~ Fig. 2e, CCLMvsWN_strong_FBR) and indicates that the dynamically modeled CCLM wind speeds are systematically overestimated relative to the empirically diagnosed wind speeds.

This discrepancy leads to the smallest WFSS values at all neighborhood sizes for this region (Fig. 2e, ~~*COSMOvsWN_strong_FBR*~~CCLMv and indicates that the dynamically modeled ~~COSMO~~ CCLM wind speeds are systematically underestimated relative to the empirically diagnosed wind speeds for this hourly event.

On the ~~1th~~ 1st of August 2012 the winds were thermally driven and the local pressure and temperature gradients were causing varying wind speeds and wind directions in the WegenerNet JBT. This is illustrated for the late afternoon INCA and WegenerNet wind fields in the upper left panels of Fig. 2c. The WFSS for the evaluation of the INCA wind field shows the second largest value at the 1 km neighborhood size, which indicates overlapping areas at this neighborhood size, equal to the horizontal resolution of the INCA analysis (Fig. 2e, INCAvsWN_therm_JBT). The large AWFSS value indicates again a small bias, which is also reflected by the similar wind classification shown in the windroses of the corresponding lower left panels of Fig. 2c. The high asymptotic value (AWFSS>0.9) indicates a small bias and that the WFSS is mostly influenced by the spatial displacement.

In a further example of a thermally induced wind event on the ~~31th~~ 31st of May 2008, we intercompare the ~~COSMO~~ CCLM with INCA wind fields (right panels of Fig. 2c). ~~In that case, the~~ Especially in the CCLM, the smoothed terrain leads to uniform wind speeds and directions. Regarding the INCA wind fields, some variability in wind speed, with higher values in the summit regions and lower values at lower altitudes in the valleys of this region, can be observed. Furthermore, a valley wind in the Enns valley is simulated by INCA. Probably the analysis part of the INCA model with its higher-resolved DEM and assimilated ZAMG observations leads to a somewhat better representation of the wind field. Comparing wind directions, the largest part of the ~~COSMO-CLM~~ CCLM modeled flow is from the N and NE sectors, while the INCA system estimated wind directions mainly from the E sector and partly from the NE and SE sectors (bottom right panels of Fig. 2c). This ~~shift in wind directions~~ simplistic pattern of wind directions in CCLM leads to low WFSS values for all neighborhood sizes, including the lowest asymptotic value of all examples, indicating a very poor representation of the wind field by the dynamical modeling of the CCLM in this challenging mountainous terrain.

The strong wind speed event for the WegenerNet JBT on the 7th of December 2013 is caused by northwesterly weather conditions. These synoptic scale flow conditions led to strong wind speeds with maximum 1-hour mean wind speeds of around ~~20 $m\ s^{-1}$~~ 20 $\mathrm{m\ s^{-1}}$ from 17:00-18:00 UTC. Both the INCA and the WegenerNet wind fields show wind directions mainly from the NW, with some proportions from the N and the W sectors, caused by a channeling of the air flow through the pronounced valleys of this study area. The INCA wind fields show much lower wind speeds in the valley regions compared to the WegenerNet wind fields, resulting from the observations of the ZAMG ADM station that flow into the INCA analysis but are considered far off the area and not used by the diagnostic modeling (upper left panels of Fig. 2d). As the neighborhood size increases, the WFSS also increases, but due to spatial displacements, the values are generally low (Fig. 2e, INCAvsWN_strong_JBT). The low AWFSS value is caused by the differences in wind speed categories (lower left panels of Fig. 2d).

For the 5th of November 2008 we intercompare the ~~COSMO-CLM~~ CCLM wind fields with INCA wind fields from 01:00-02:00 UTC, for a strong wind speed event (right panels of Fig. 2d). In this example, the influence of the smoothed terrain caused by the course horizontal resolution of the ~~COSMO-CLM model~~ CCLM becomes obvious. This smoothed topography results in systematically lower wind speeds compared to the INCA wind fields. The WFSS shows similar results like for the previous INCAvsWN_strong_JBT evaluation, with small values at all neighborhood sizes (Fig. 2e, ~~*COSMOvsINCA_strong_JBT*~~CCLMvsINCA_strong indicating the clear limits of the ~~COSMO~~ CCLM dynamical modeling fields also for strong wind events.

4.2 Statistical evaluation results

The statistical event-averaged WFSS values from the large ensemble of events over multiple years are represented for each evaluation case in Fig. 3. Overall, it shows a monotonic increase with neighborhood size for all cases so that the AWFSS is the largest value, indication relatively the best performance at large scales.

For the WegenerNet FBR, the statistical WFSS values, calculated for the INCA wind fields compared to the WegenerNet wind fields, shows for both the thermally induced and strong wind events nearly the same behavior (Fig. 3a, INCAvsWN_therm_FBR and INCAvsWN_strong_FBR). The WFSS values for these cases indicate a reasonably good spatial matching at all neighborhood sizes. Furthermore, the AWFSS values are higher than 0.8, reflecting generally small INCA biases of wind classes.

The statistical WFSS estimated for the ~~COSMOvsWN~~CCLMvsWN_therm_FBR case indicates that the ~~COSMO-CLM model~~ CCLM clearly and systematically underperforms in case of thermally induced wind events for the WegenerNet FBR. Evidently, due to the coarse horizontal resolution of the wind fields, the ~~COSMO-CLM~~ CCLM wind fields appear fundamentally ~~able~~ unable to capture the varying wind directions for such events in this region. For the ~~COSMOvsWN~~CCLMvsWN_strong_FBR case, however the results indicate a similar spatial matching as for the INCAvsWN_therm_FBR and INCAvsWN_strong_FBR cases, just with a somewhat higher bias of wind class differences. This similar performance despite the coarser horizontal resolution of the ~~COSMO~~ CCLM model is explained through a weaker influence of the terrain on the wind fields under strong wind conditions in this region.

Because of the challenging terrain of the WegenerNet JBT, the statistical WFSS values are generally low for this region, signalling large biases (Fig. 3b). These biases are indicated by low asymptotic values, which tend to be between 0.61 and 0.64, except for the INCAvsWN_strong_JBT case, which shows an even lower value (AWFSS=0.39).

The spatial displacement and the biases for the INCAvsWN_therm_JBT case are mainly caused by the differences in wind directions for these thermally induced wind events. Especially at small neighborhood sizes at the 1-km scale, WFSS values indicate large spatial displacements.

The INCAvsWN_strong_JBT case shows the lowest values at all neighborhood sizes, but this time caused by the differences in the wind speed categories. These low values are caused by the INCA-analyzed wind speeds, which, in case of strong winds, are overestimated in the summit regions and underestimated in the valley regions. Slightly overestimated WegenerNet wind speeds in the Enns valley are somewhat reinforcing the difference between the INCA and the WegenerNet wind speeds. These differences in wind speed especially in the valley and the summit regions become obvious from Fig. 4c and 4d and are discussed in further detail below.

The intercomparison of the ~~COSMO-CLM~~ CCLM wind fields with INCA delivers nearly the same (low) WFSS values for both type of wind events. In case of thermally induced events (~~COSMOvsINCA~~CCLMvsINCA_therm_JBT) the spatial displacements and biases are mainly caused due to differences in wind directions. For strong wind events (~~COSMOvsINCA~~CCLMvsINCA_str the smoothed terrain caused by the coarse resolution of the ~~COSMO-CLM model~~ CCLM leads to systematically underestimated wind speeds.

Table 4 summarizes, in addition to the AWFSS values, the results estimated with traditional statistical methods of the INCA analysis and ~~COSMO~~ CCLM dynamical fields. Due to the less challenging region of the WegenerNet FBR, all traditional statistical parameters show better performance for this region compared to the WegenerNet JBT. The absolute-value statistical metrics (bias *B*, standard deviation SD_o, root-mean-square-error *RMSE*) applied to the hourly vector-mean wind speeds, show higher values for the WegenerNet JBT, resulting from the generally higher wind speeds in addition to effects of the complex mountainous terrain on the wind fields in this region. The *B* values are slightly positive for the WegenerNet FBR and negative for the WegenerNet JBT. The substantially negative *B* value for the INCAvsWN_strong_JBT case again reflects the underestimation of wind speed in the valleys, as explained above. Furthermore, the ~~COSMOvsINCA~~CCLMvsINCA_strong_JBT intercomparison also shows a negative bias, caused by the coarse resolution of the ~~COSMO-CLM model~~CCLM, which leads to lower wind speeds for strong wind events.

The *RMSE* values range from 0.79 m s^{-1} to 1.85 m s^{-1} for the WegenerNet FBR and from 1.3 to 8.6 m s^{-1} for the WegenerNet JBT. The high value of 8.6 m s^{-1} for the INCAvsWN_strong_JBT case is caused by the underestimation of wind speed in the valleys as well as the overestimation in the summit regions by the INCA model. The mean *R* values show a better correlation for the WegenerNet FBR than for the WegenerNet JBT. The mean absolute error of wind direction (MAE_{dir}) applied to hourly vector-mean wind directions also shows better performance (INCA and ~~COSMO~~ CCLM fields) for the WegenerNet FBR. Due to the varying wind directions caused by thermally induced circulations, the MAE_{dir} is higher for such events for both study areas, with the highest value of 68° for the INCAvsWN_therm_JBT case.

Figure 4 shows the mean wind speed bias spatial distributions for all evaluation cases, for the WegenerNet FBR (Fig. 4a and b) and the WegenerNet JBT (Fig. 4c and d). The distribution for the INCAvsWN_therm_FBR case, the case for thermally induced wind events for the WegenerNet FBR, shows large areas with nearly no B values (left panel of Figure 4a). The maximum B value for this case can be observed in the area of the *Gleichenberger Kogel*, north of the ZAMG Bad Gleichenberg station with a value of around 1.4 m s^{-1}. The evaluation of the ~~COSMO-CLM model~~ CCLM for the same thermally induced events shows an overestimation of wind speeds for the whole study area, with B values from 0.5 m s^{-1} to 1 m s^{-1} in the western part, and from 1 m s^{-1} to about 1.4 m s^{-1} in the eastern part of the study area (right panel of Figure 4a).

The overestimation of the wind speeds for the WegenerNet FBR can be explained by the too frequent flow-over patterns simulated for this region, which lead to a more dominant orographic speed-up effect. Due to the orographic smoothing, flow-over patterns are generally more frequent than flow-around patterns, especially for the WegenerNet FBR with its small differences in altitude (Taylor et al., 1987).

The evaluation of the INCA model for strong wind speeds illustrates the strong influence of the terrain on this model. The results show a good agreement in the valleys of the study area, with partly small negative B values (left panel of Figure 4b). The hilltop regions exhibit positive B values, with maximum values of around 5 m s^{-1} again in the area of the *Gleichenberger Kogel*. Overall positive B values of ~~COSMO-CLM~~ CCLM dynamical wind speed fields for strong wind events are seen in the right panel of Figure 4b, showing the systematic overestimating by ~~COSMO~~ CCLM fields in this case. These large B-values are probably also due to the speed-up effect explained for the above case CCLMvsWN_therm_FBR.

For the WegenerNet JBT, the strong influence of the terrain on the INCA-analyzed wind speeds can be observed in all evaluation cases in this region. The evaluation of the INCA model for thermally induced wind events exhibit negative B values in the valleys, whereby positive values are partly occurring in the summit regions (left panel of Figure 4c). At lower elevations, the intercomparison of the ~~COSMO-CLM~~ CCLM with the INCA model shows nearly no B values for thermally induced events (right panel of Figure 4c). Furthermore, small negative B values at the summit regions and some spots with positive values can be observed for this case. Due to these small bias values, similar results as these ones can be expected for a comparison of CCLM with WegenerNet data.

Similar bias distribution patterns as for the INCA evaluation for thermally induced wind events are present for strong wind events in the INCAvsWN_strong_JBT case, but this time with strong negative and positive B values ranging from ~~−14.4~~−14.4 m s^{-1} to ~~4.9 $m\ s^{-1}$~~ 4.9 m s^{-1} (left panel of Figure 4d). These strong negative B values are again caused by the severely underestimated INCA wind speeds and the somewhat overestimated WegenerNet wind speeds in the valley regions of the WegenerNet JBT.

Opposite patterns can be seen for the intercomparison of the ~~COSMO-CLM~~ CCLM with the INCA model. This intercomparison exhibits small positive values in the valley regions and strong negative values in the summit regions (right panel of Figure 4d). Main reason for these strong negative B values is the course resolution of the ~~COSMO-CLM~~ CCLM data and the resulting underestimation of wind speeds for strong wind events, as explained above. The negative B-values are likely caused by negative orographic speed-up effects, which are preferred in flow-around patterns and flow-splitting patterns that occur especially when the differences in the altitude of ridges of mountains are large (Hewer, 1998).

5 Conclusions

In this work we evaluated wind fields generated by two different modeling systems against empirically diagnosed wind fields from WegenerNet high-density network data: the INCA analysis system of the Austrian weather service ZAMG (Haiden et al., 2011) and the non-hydrostatic ~~COSMO-CLM model~~ CCLM (Schättler et al., 2016). The INCA wind fields have a horizontal resolution of 1 km $\times$ 1 km, and in case of ~~COSMO-CLM~~CCLM, 3 km $\times$ 3 km horizontal resolution was available, both on an hourly basis. The empirical high-resolution wind fields from the WegenerNet ~~where~~ were generated by the WegenerNet Wind Product Generator (WPG), recently developed by Schlager et al. (2017, 2018).

The WPG-diagnosed gridded wind fields are available with a temporal resolution of 30 minutes and a spatial resolution of 100 m $\times$ 100 m and can therefore well serve as reference. The WegenerNet Feldbach Region (FBR) was used as study area, characterized by generally small differences in altitude in the hilly terrain of this region. The second study area was the WegenerNet Johnsbachtal (JBT) region, which is a mountainous region characterized through a very complex terrain.

The evaluation of the INCA and the ~~COSMO-CLM~~ CCLM wind fields was based on classifying the data separately into thermally induced and strong wind events. In case of the INCA evaluation, we could select wind events within the period 2008-2017 for the WegenerNet FBR and within 2012-2017 for the WegenerNet JBT. For evaluating the ~~COSMO-CLM~~ CCLM data, events from the period 2008-2010 were selected for both study areas. Due to WegenerNet JBT wind fields being not yet available within 2008-2010, we intercompared the ~~COSMO-CLM~~ CCLM wind fields with INCA wind fields in this region.

Besides traditional performance measures such as bias, root-mean-square error, correlation coefficient, and mean absolute error of wind direction, we in particular applied a spatial wind verification methodology named the Wind Fractions Skill Score (WFSS) (Skok and Hladnik, 2018). This new score was used to detect spatial displacements of wind patterns and biases based on predefined wind speed and direction classes. The WFSS avoids the 'double penalty' problem and is able to distinguish between a 'near miss' and large displacements between modeled and reference wind fields. Furthermore, a spatial scale-dependent skill is determined by this score.

Due to the less challenging terrain of the Alpine foreland region, all statistical performance measures showed better INCA and ~~COSMO~~ CCLM performance for the WegenerNet FBR than for the challenging mountainous region WegenerNet JBT. The spatial verification of all evaluation cases indicates an increasing skill with increasing by larger scale (neighborhood size). For both study areas, the traditional statistical performance measures, applied to the wind speed, mostly show better performance of INCA and ~~COSMO~~ CCLM for thermally induced wind events than for strong wind events. On the other hand, the results related to wind direction indicate a better performance for strong wind events than for thermally induced events.

More specifically, the verification for the WegenerNet FBR shows that the INCA analysis wind fields are more skillful than the ~~COSMO-CLM~~ CCLM dynamical wind fields in this region. The INCA verification indicates a reasonably good performance for both thermally induced and strong wind events~~, with some spatial displacements at smaller scales and small biases in wind classes. Regarding wind speed, we found a generally good modeling performance by these analysis fields for both types of wind events with slight overestimates in the summit regions of the WegenerNet FBR.~~ .

The ~~COSMO-CLM dynamical model~~ CCLM clearly performs less well in case of thermally induced wind events for this region. The reason for this weak performance is the limited resolution of the wind field dataset from this model. ~~With the resolution~~ Although the difference in the numerical resolution between INCA (1 km grid spacing) and CCLM (3 km grid spacing) is only a factor of 3, CCLM is not able to resolve small-scale wind patterns. This occurs for multiple reasons: 1) due to the 3rd-order advection scheme with its horizontal diffusion damping, the effective resolution in CCLM is several times coarser than the numeric grid spacing (Ogaja and Will, 2016); 2) the orography is smoothed as well, so that individual mountain ridge and valley structures are removed. For example, the mountain peak of the Hochtor with its 2396 m elevation in the center of the WegenerNet JBT region is lowered by about 500 m in the CCLM model.

Hence, with the resolution of 3 km $\times$ 3 km, the fields are not able to resolve the varying wind speeds and directions caused by thermally driven circulations. The wind speeds are overestimated by this model for both, thermally ~~and strong~~ induced and strong wind events, and large differences in wind directions are found for thermally induced events.

For the WegenerNet JBT region, the verification shows generally large spatial displacements at all scales and strong biases in wind classes for all evaluation cases. In case of the INCA evaluation, large wind direction deviations for thermally induced wind events indicate that the analysis fields are not able to adequately capture the varying wind patterns such as slope and valley winds, which roots in the sparse station density that INCA can anchor to and the coarse horizontal resolution of the first guess provided by the ALARO model in this complex-terrain region. Furthermore, the statistics show an substantial underestimation of wind speeds in the valleys and overestimated wind speeds in parts of the summit regions for both type of wind events. ~~Especially for strong wind events we found a large negative bias in the valley regions and a positive bias in parts of the summit regions. Somewhat overestimated WegenerNet empirical wind speeds in the region of the Enns valley contribute in a minor way to the strong negative bias.~~

The intercomparison of the ~~COSMO-CLM~~ CCLM dynamical fields with the INCA analysis fields for thermally induced wind events ~~shows large displacements and a bias from wind speed class differences. The error measure applied to the wind direction for this intercomparison indicates large differences in wind directions for thermally induced wind events between the COSMO and INCA wind fields. Furthermore, large spatial displacements, biases in populating the wind classes, underestimated wind speeds in the summit regions, and slightly overestimated wind speeds in the valleys,~~ are reflecting the disadvantage of smoothed terrain, which is caused by the limited ~~resolution of~~ effective resolution being several times coarser than the 3 km $\times$ 3 km ~~of the COSMO-CLM model.~~ grid spacing of the CCLM as already noted above. Improvements can be expected from latest developments in the numerical core of CCLM by Ogaja and Will (2016), which have enabled an improvement of the effective resolution by a factor of 2 via introducing a 4th-order advection scheme that allows to circumvent the horizontal diffusion damping.

Based on these finding we suggest, ~~underpining~~ underpinning the results of Haiden et al. (2011), that additional observed wind information in the summit and valley regions, especially in a complex terrain like the WegenerNet JBT, and a more comprehensive use of wind-constraining satellite data as well as a higher-resolution RCM could help to systematically improve the INCA-analyzed wind fields. At higher resolutions, the topographic shading through the terrain becomes increasingly

important, especially for the simulation of thermally induced wind events. Such methods have not yet been implemented into the ALARO model, but may help to generate more realistic wind fields in the future.

Related to the ~~COSMO~~ CCLM dynamical modeling, a verification of ~~COSMO-CLM-generated~~ CCLM-generated higher-resolution 1 km $\times$ 1 km wind fields and the application of the new 4th-order advection scheme from Ogaja and Will (2016) in a convection-permitting configuration would also be a promising issue for further investigations of how this may improve the modeling of wind patterns in a complex terrain.

Investigations regarding the WegenerNet JBT wind fields showed, that an additional wind-observing station in the Enns valley would improve the results for this region (Schlager et al., 2018). Such an additional station would avoid the overestimation of WegenerNet wind speeds in the Enns valley, especially for strong wind events. In the WegenerNet FBR region just recently (in May 2018) another wind station was added in the Raab valley (station Nr. 155 1b), which will further improve the WPG-derived fields in future. This adds further value to valuable reference for evaluation of important other data products such as the INCA operational analysis and dynamical climate model fields.

Code availability. The CALMET 6.5.0 model code is available from the website www.src.com/calpuff/. The INCA and the WPG code is not in the public domain and cannot be distributed. The source code for the CCLM is available on request via the website https://www.clm-community.eu. The code for the calculation of FSSwind score is available as part of the SpatialVx package (function calculate_FSSwind). SpatialVx is a R software package made by Eric Gilleland that enables the calculation of a large number of spatial verification scores (https://cran.r-project.org/package=SpatialVx).

Data availability. CORINE Land Cover data for the study area were taken from www.eea.europa.eu, digital elevation model data from www.gis.steiermark.at, and WegenerNet data from www.wegenernet.org. The WegenerNet data contain the WPG wind field output data as introduced in this study. The INCA data are available on request from the Central Institute for Meteorology and Geodynamics (klima@zamg.ac.at). CCLM data are available on request from the Wegener Center, University of Graz (heimo.truhetz@uni-graz.at).

Author contributions. C. Schlager collected the data, performed the analyses and modeling, created the figures, and wrote the first draft of the manuscript. G. Kirchengast provided guidance and advice on all aspects of the study and significantly contributed to the text. J. Fuchsberger provided guidance on technical aspects of the WegenerNet networks, and its data characteristics and contributed to the text. A. Kann provided INCA-related advice and contributed to the INCA part of the text and H. Truhetz provided information and advice on the CCLM setup and characteristics and contributed in particular to the CCLM part of the text. All authors commented on the final version of the manuscript.

Competing interests. We declare that no competing interests are present.

Acknowledgements. The authors thank Gregor Skok (~~University of Ljubljana, Department of Physic~~Dept. of Physics, Univ. of Ljubljana), for providing the R code to calculate the Wind Fractions Skill Score. Furthermore, the authors acknowledge the data providers at the Central Institute for Meteorology and Geodynamics (ZAMG) for the Integrated Nowcasting through Comprehensive Analysis (INCA) dataset. Andras Csaki (Wegener Center, Univ. of Graz) is thanked for performing the CCLM modeling and extracting the wind field data. The authors are as well grateful to the Julich Supercomputing Centre (JSC) and the Vienna Scientific Cluster (VSC) for providing the necessary HPC resources. WegenerNet funding is provided by the Austrian Ministry for Science and Research, the University of Graz, the state of Styria (which also included European Union regional development funds), and the city of Graz; detailed information can be found online (www.wegcenter.at/wegenernet). The CCLM simulation was funded by the Austrian Science Fund (FWF) under project NHCM-2 (project number P24758-N29).

References

Abdel-Aal, R., Elhadidy, M., and Shaahid, S.: Modeling and forecasting the mean hourly wind speed time series using GMDH-based abductive networks, Renewable Energy, 34, 1686–1699, https://doi.org/10.1016/j.renene.2009.01.001, 2009.

Awan, N. K., Truhetz, H., and Gobiet, A.: Parameterization-induced error characteristics of MM5 and WRF operated in climate mode over the alpine region: An ensemble-based analysis, J. Climate, 24, 3107–3123, https://doi.org/10.1175/2011JCLI3674.1, 2011.

Bartalev, S., Belward, A., Ershov, D., and S. Isaev, A.: A new SPOT4-VEGETATION derived land cover map of Northern Eurasia, Int. J. Remote Sens., 24, 1977–1982, https://doi.org/10.1080/0143116031000066297, 2003.

Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M. J., Vitart, F., and Balsamo, G.: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, Quart. J. Roy. Meteor. Soc., 134, 1337–1351, https://doi.org/10.1002/qj.289, http://dx.doi.org/10.1002/qj.289, 2008.

Bellasio, R., Maffeis, G., Scire, J. S., Longoni, M. G., Bianconi, R., and Quaranta, N.: Algorithms to Account for Topographic Shading Effects and Surface Temperature Dependence on Terrain Elevation in Diagnostic Meteorological Models, Bound.Layer Meteor., 114, 595–614, https://doi.org/10.1007/s10546-004-1670-6, 2005.

Böhm, U., Kücken, M., Ahrens, W., Block, A., Hauffe, D., Keuler, B., Rockel, B., and Will, A.: CLM-The Climate Version of LM: Brief Description and Long-Term Applications, pp. 225–236, Deutscher Wetterdienst (DWD), 6 edn., [Available online at http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter06/newsLetter_06.pdf.], 2006.

Bubnová, R., Hello, G., Bénard, P., and Geleyn, J.-F.: Integration of the Fully Elastic Equations Cast in the Hydrostatic Pressure Terrain-Following Coordinate in the Framework of the ARPEGE/Aladin NWP System, Mon. Wea. Rev., 123, 515–535, https://doi.org/10.1175/1520-0493(1995)123<0515:IOTFEE>2.0.CO;2, 1995.

Cox, R. M., Sontowski, J., and Dougherty, C. M.: An evaluation of three diagnostic wind models (CALMET, MCSCIPUF, and SWIFT) with wind data from the Dipole Pride 26 field experiments, Meteor. Appl., 12, 329–341, https://doi.org/10.1017/S1350482705001908, 2005.

Davies, H. C.: A lateral boundary formulation for multi-level prediction models, Quart. J. Roy. Meteor. Soc., 102, 405–418, https://doi.org/10.1002/qj.49710243210, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49710243210, 1976.

EEA: Global land cover 2000 - Europe, Tech. rep., European Environment Agency (EEA), https://www.eea.europa.eu/data-and-maps/data/global-land-cover-2000-europe, 2016.

Gal-Chen, T. and Somerville, R. C.: On the use of a coordinate transformation for the solution of the Navier-Stokes equations, J. Comput. Phys., 17, 209 – 228, https://doi.org/https://doi.org/10.1016/0021-9991(75)90037-6, 1975.

Gerard, L. and Geleyn, J.-F.: Evolution of a subgrid deep convection parametrization in a limited-area model with increasing resolution, Quart. J. Roy. Meteor. Soc., 131, 2293–2312, https://doi.org/10.1256/qj.04.72, 2005.

Giard, D. and Bazile, E.: Implementation of a New Assimilation Scheme for Soil and Surface Variables in a Global NWP Model, Mon. Wea. Rev., 128, 997–1015, https://doi.org/10.1175/1520-0493(2000)128<0997:IOANAS>2.0.CO;2, 2000.

Gilleland, E., Ahijevych, D. A., Brown, B. G., and Ebert, E. E.: Verifying Forecasts Spatially, Bull. Amer. Meteor. Soc., 91, 1365–1376, https://doi.org/10.1175/2010BAMS2819.1, 2010.

Gómez-Navarro, J. J., Raible, C. C., and Dierer, S.: Sensitivity of the WRF model to PBL parametrisations and nesting techniques: evaluation of wind storms over complex terrain, Geosci. Model Dev., 8, 3349–3363, https://doi.org/10.5194/gmd-8-3349-2015, https://www.geosci-model-dev.net/8/3349/2015/, 2015.

Gross, G.: On the applicability of numerical mass-consistent wind field models, Bound.Layer Meteor., 77, 379–394, https://doi.org/10.1007/BF00123533, 1996.

Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B., and Gruber, C.: The Integrated Nowcasting through Comprehensive Analysis (INCA) system and its validation over the eastern alpine region, Wea. Forecasting, 26, 166–183, https://doi.org/10.1175/2010WAF2222451.1, 2011.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006, J. Geophys. Res., 113, n/a–n/a, https://doi.org/10.1029/2008JD010201, 2008.

Hewer, F.: Non-Linear Numerical Model Predictions of Flow Over an Isolated Hill of Moderate Slope, Bound.Layer Meteor., 87, 381–408, https://doi.org/10.1023/A:1000944817965, 1998.

Hiebl, J. and Frei, C.: Daily temperature grids for Austria since 1961—concept, creation and applicability, Theor. Appl. Climatol., 124, 161–178, https://doi.org/10.1007/s00704-015-1411-4, 2016.

Hohmann, C., Kirchengast, G., and Birk, S.: Alpine foreland running drier? Sensitivity of a drought vulnerable catchment to changes in climate, land use, and water management, Climatic Change, 147, 179–193, https://doi.org/10.1007/s10584-017-2121-y, 2018.

Kabas, T.: WegenerNet climate station network region Feldbach: Experimental setup and high resolution data for weather and climate research (in German), Scientific Rep. 47-2012, Wegener Center Verlag, Graz, Austria, [Available online at http://wegcwww.uni-graz.at/publ/wegcreports/2012/WCV-WissBer-No47-TKabas-Jan2012.pdf.], 2012.

Kabas, T., Foelsche, U., and Kirchengast, G.: Seasonal and annual trends of temperature and precipitation within 1951/1971-2007 in south-eastern Styria, Austria, Meteor. Z., 20, 277–289, https://doi.org/10.1127/0941-2948/2011/0233, 2011.

Kann, A., Meirold-Mautner, I., Schmid, F., Kirchengast, G., Fuchsberger, J., Meyer, V., Tuechler, L., and Bica, B.: Evaluation of high-resolution precipitation analyses using a dense station network, Hydrol. Earth Syst. Sci., 19, 1547–1559, https://doi.org/10.5194/hess-19-1547-2015, 2015a.

Kann, A., Wittmann, C., Bica, B., and Wastl, C.: On the Impact of NWP Model Background on Very High–Resolution Analyses in Complex Terrain, Wea. Forecasting, 30, 1077–1089, https://doi.org/10.1175/WAF-D-15-0001.1, 2015b.

Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J. P., Fosser, G., and Wilkinson, J. M.: Do convection-permitting regional climate models improve projections of future precipitation change?, Bull. Amer. Meteor. Soc., 98, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1, 2017.

Kida, H., Koide, T., Sasaki, H., and Chiba, M.: A New Approach for Coupling a Limited Area Model to a GCM for Regional Climate Simulations, J. Meteor. Soc. Japan, 69, 723–728, https://doi.org/10.2151/jmsj1965.69.6_723, 1991.

Kirchengast, G., Kabas, T., Leuprech, A., Bichler, C., and Truhetz, H.: WegenerNet: A pioneering high-resolution network for monitoring weather and climate, Bull. Amer. Meteor. Soc., 95, 227–242, https://doi.org/10.1175/BAMS-D-11-00161.1, 2014.

Leutwyler, D., Fuhrer, O., Lapillonne, X., Lüthi, D., and Schär, C.: Towards European-scale convection-resolving climate simulations with GPUs: A study with COSMO 4.19, Geosci. Model Dev., 9, 3393–3412, https://doi.org/10.5194/gmd-9-3393-2016, 2016.

Lugauer, M. and Winkler, P.: Thermal circulation in South Bavaria – climatology and synoptic aspects, Meteor. Z., 14, 15–30, https://doi.org/10.1127/0941-2948/2005/0014-0015, 2005.

Morales, L., Lang, F., and Mattar, C.: Mesoscale wind speed simulation using CALMET model and reanalysis information: An application to wind potential, Renewable Energy, 48, 57–71, https://doi.org/10.1016/j.renene.2012.04.048, 2012.

Noilhan, J. and Planton, S.: A Simple Parameterization of Land Surface Processes for Meteorological Models, Mon. Wea. Rev., 117, 536–549, https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2, 1989.

O, S., Foelsche, U., Kirchengast, G., Fuchsberger, J., Tan, J., and Petersen, W. A.: Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria, Hydrol. Earth Syst. Sci., 21, 6559–6572, https://doi.org/10.5194/hess-21-6559-2017, 2017.

O, S., Foelsche, U., Kirchengast, G., and Fuchsberger, J.: Validation and correction of rainfall data from the WegenerNet high density network in southeast Austria, J. Hydrol., 556, 1110–1122, https://doi.org/10.1016/j.jhydrol.2016.11.049, 2018.

Ogaja, J. and Will, A.: Fourth order, conservative discretization of horizontal Euler equations in the COSMO model and regional climate simulations, Meteor. Z., 25, 577–605, 2016.

Osborn, T. J. and Hulme, M.: Evaluation of the European daily precipitation characteristics from the atmospheric model intercomparison project, Int. J. Climatol., 18, 505–522, https://doi.org/10.1002/(SICI)1097-0088(199804)18:5<505::AID-JOC263>3.0.CO;2-7, 1998.

Prein, A. F.: Added Value of Convection Resolving Climate Simulations, Scientific Report 53-2013, Wegener Center Verlag, Graz, Austria, [Available online at http://wegcwww.uni-graz.at/publ/wegcreports/2013/WCV-SciRep-No53-APrein-Jul2013.pdf.], 2013.

Prein, A. F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N. K., Keuler, K., and Georgievski, G.: Added value of convection permitting seasonal simulations, Climate Dyn., 41, 2655–2677, https://doi.org/10.1007/s00382-013-1744-6, https://doi.org/10.1007/s00382-013-1744-6, 2013a.

Prein, A. F., Holland, G. J., Rasmussen, R. M., Done, J., Ikeda, K., Clark, M. P., and Liu, C. H.: Importance of regional climate model grid spacing for the simulation of heavy precipitation in the colorado headwaters, J. Climate, 26, 4848–4857, https://doi.org/10.1175/JCLI-D-12-00727.1, 2013b.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Toelle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., Van Lipzig, N. P., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Rev. Geophys., 53, 323–361, https://doi.org/10.1002/2014RG000475, 2015.

Prettenthaler, F., Podesser, A., and Pilger, H.: Climate Atlas Styria, Period 1971-2000: An Application-Oriented Climatology (in German), vol. 4, Verlag der Oesterreichischen Akademie der Wissenschaften, Wien, 2010.

Raschendorfer, M.: The New Turbulence Parameterization of LM, pp. 89–97, Deutscher Wetterdienst (DWD), 1 edn., [Available online at http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter01/newsLetter_01.pdf.], 2001.

Ratto, C., Festa, R., Romeo, C., Frumento, O., and Galluzzi, M.: Mass-consistent models for wind fields over complex terrain: The state of the art, Environ. Software, 9, 247 – 268, 1994.

Raymond, W. H.: High-Order Low-Pass Implicit Tangent Filters for Use in Finite Area Calculations, Mon. Wea. Rev., 116, 2132–2141, https://doi.org/10.1175/1520-0493(1988)116<2132:HOLPIT>2.0.CO;2, 1988.

Ritter, B. and Geleyn, J.-F.: A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations, Mon. Wea. Rev., 120, 303–325, https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2, 1992.

Roberts, N.: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model, Meteor. Appl., 15, 163–169, https://doi.org/10.1002/met.57, http://dx.doi.org/10.1002/met.57, 2008.

Rockel, B., Will, A., and Hense, A.: The Regional Climate Model COSMO-CLM (CCLM), Meteor. Z., 17, 347–348, https://doi.org/10.1127/0941-2948/2008/0309, http://dx.doi.org/10.1127/0941-2948/2008/0309, 2008.

Schättler, Doms, G, U., and Baldauf, M.: A Description of the Nonhydrostatic Regional COSMO Model; Part VII: User's Guide, Deutscher Wetterdienst, 3004 Offenbach, Germany, 2016.

Schlager, C., Kirchengast, G., and Fuchsberger, J.: Generation of high-resolution wind fields from the dense meteorological station network WegenerNet in south-eastern Austria, Wea. Forecasting, 32, 1301–1319, https://doi.org/10.1175/WAF-D-16-0169.1, 2017.

Schlager, C., Kirchengast, G., and Fuchsberger, J.: Empirical high-resolution wind field and gust model in mountainous and hilly terrain based on the dense WegenerNet station networks, Atmos. Meas. Tech., 2018, 1–32, https://doi.org/10.5194/amt-2018-31, https://www.atmos-meas-tech-discuss.net/amt-2018-31/, in press, 2018.

Schroeer, K. and Kirchengast, G.: Sensitivity of extreme precipitation to temperature: the variability of scaling factors from a regional to local perspective, Climate Dyn., 50, 3981–3994, 2018.

Scire, J. S., Robe, F. R., Fernau, M. E., and Roberto, Y. J.: A User's Guide for the CALMET Meteorological Model (Version 5), Earth Tech, Inc, 196 Baker Avenue, Concord, MA 01742, 1998.

Seaman, N. L.: Meteorological modeling for air-quality assessments, Appl. Energy, 34, 2231–2259, https://doi.org/10.1016/S1352-2310(99)00466-5, 2000.

Sfetsos, A.: A novel approach for the forecasting of mean hourly wind speed time series, Renewable Energy, 27, 163 – 174, https://doi.org/http://dx.doi.org/10.1016/S0960-1481(01)00193-8, 2002.

Skok, G. and Hladnik, V.: Verification of Gridded Wind Forecasts in Complex Alpine Terrain: A New Wind Verification Methodology Based on the Neighborhood Approach, Mon. Wea. Rev., 146, 63–75, https://doi.org/10.1175/MWR-D-16-0471.1, 2018.

Strasser, U., Marke, T., Sass, O., Birk, S., and Winkler, G.: John's creek valley: A mountainous catchment for long-term interdisciplinary human-environment system research in Upper Styria (Austria), Environ. Earth Sci., 69, 695–705, https://doi.org/10.1007/s12665-013-2318-y, 2013.

Suklitsch, M., Gobiet, A., Truhetz, H., Awan, N. K., Göttel, H., and Jacob, D.: Error characteristics of high resolution regional climate models over the Alpine area, Climate Dyn., 37, 377–390, https://doi.org/10.1007/s00382-010-0848-5, 2011.

Taylor, P. A., Mason, P. J., and Bradley, E. F.: Bound.Layer Meteor., Boundary-Layer Meteorology, 39, 107–132, https://doi.org/10.1007/BF00121870, 1987.

Termonia, P., Fischer, C., Bazile, E., Bouyssel, F., Brožková, R., Bénard, P., Bochenek, B., Degrauwe, D., Derková, M., El Khatib, R., Hamdi, R., Mašek, J., Pottier, P., Pristov, N., Seity, Y., Smolíková, P., Španiel, O., Tudor, M., Wang, Y., Wittmann, C., and Joly, A.: The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1, Geosci. Model Dev., 11, 257–281, https://doi.org/10.5194/gmd-11-257-2018, 2018.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Monthly Weather Review, 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Truhetz, H.: High resolution wind field modelling over complex topography : analysis and future scenarios, Scientific Rep. 32-2010, Wegener Center Verlag, Graz, Austria, [Available online at http://wegcwww.uni-graz.at/publ/wegcreports/2010/WCV-SciRep-No32-HTruhetz-Apr2010.pdf.], 2010.

Wakonigg, H.: Weather and Climate in Styria (in German), Verlag fuer die Technische Universitaet Graz, Graz, 1978.

Wang, Y., Haiden, T., and Kann, A.: The operational Limited Area Modelling system at ZAMG: ALADIN-AUSTRIA, vol. 37, Österreichische Beiträge zu Meteorologie und Geophysik, Wien, 2006.

Whiteman, C.: Mountain Meteorology: Fundamentals and Applications, Oxford University Press, Graz, 2000.

Wicker, L. J. and Skamarock, W. C.: Time-Splitting Methods for Elastic Models Using Forward Time Schemes, Mon. Wea. Rev., 130, 2088–2097, https://doi.org/10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2, 2002.

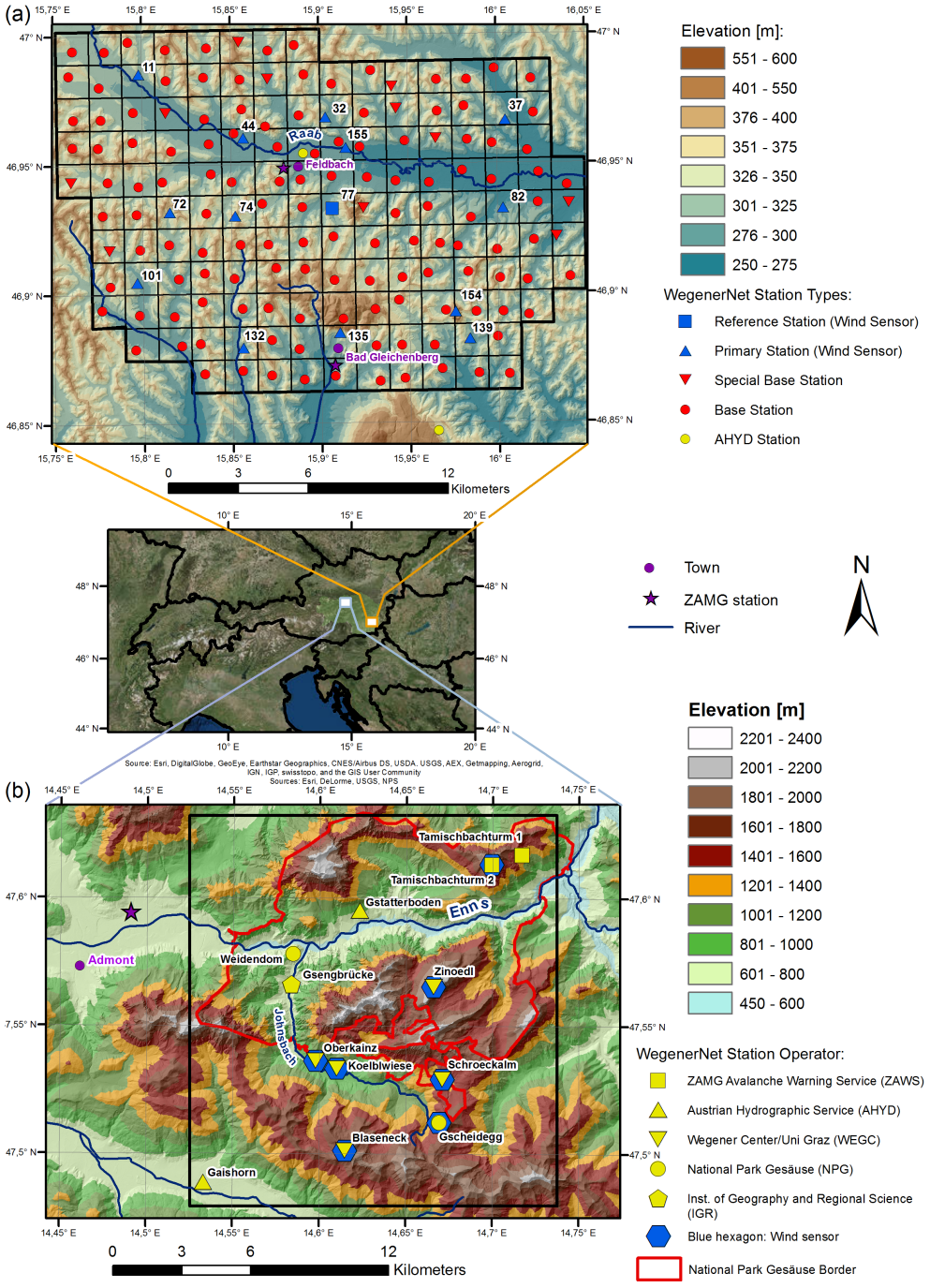

Figure 1. ~~(a)~~ Location of the study areas in Austria (middle panel), the WegenerNet ~~Felbach~~ Feldbach Region (FBR) in the southeast of the state of Styria ~~, Austria~~ and the WegenerNet Johnsbachtal (JBT) region in the north of Styria ~~, Austria~~ (white-filled rectangles, enlarged in (~~b~~a) and (~~e~~b)). (~~b~~a) The WegenerNet FBR with its ~~154~~ 155 meteorological stations, with the legend explaining map characteristics and station types. (~~e~~b) Map of the WegenerNet JBT region (black rectangle) including its meteorological stations, with the legend explaining map characteristics and station operators.

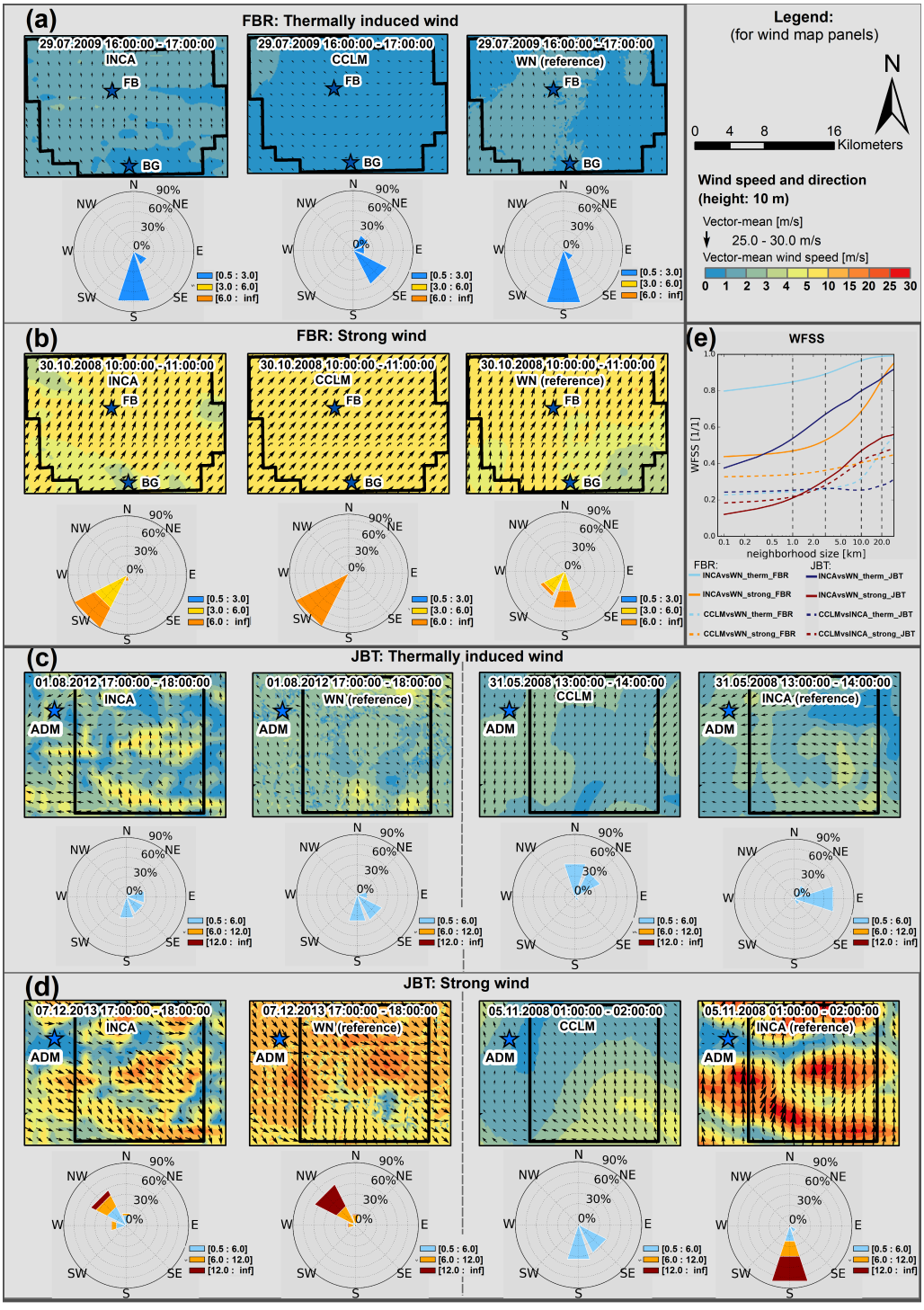

Figure 2. Wind Fractions Skill Score (WFSS) analysis for selected one-hour wind fields for the WegenertNet FBR (a, b) and the WegenerNet JBT region (c, d). (a-d) Modeled and reference wind fields (first row) and corresponding relative frequency of wind directions for a range of wind speed categories (second row), in each panel. (e) WFSS results for the modeled versus reference wind fields from (a)-(d). See Table 1 for more information on the evaluation cases.

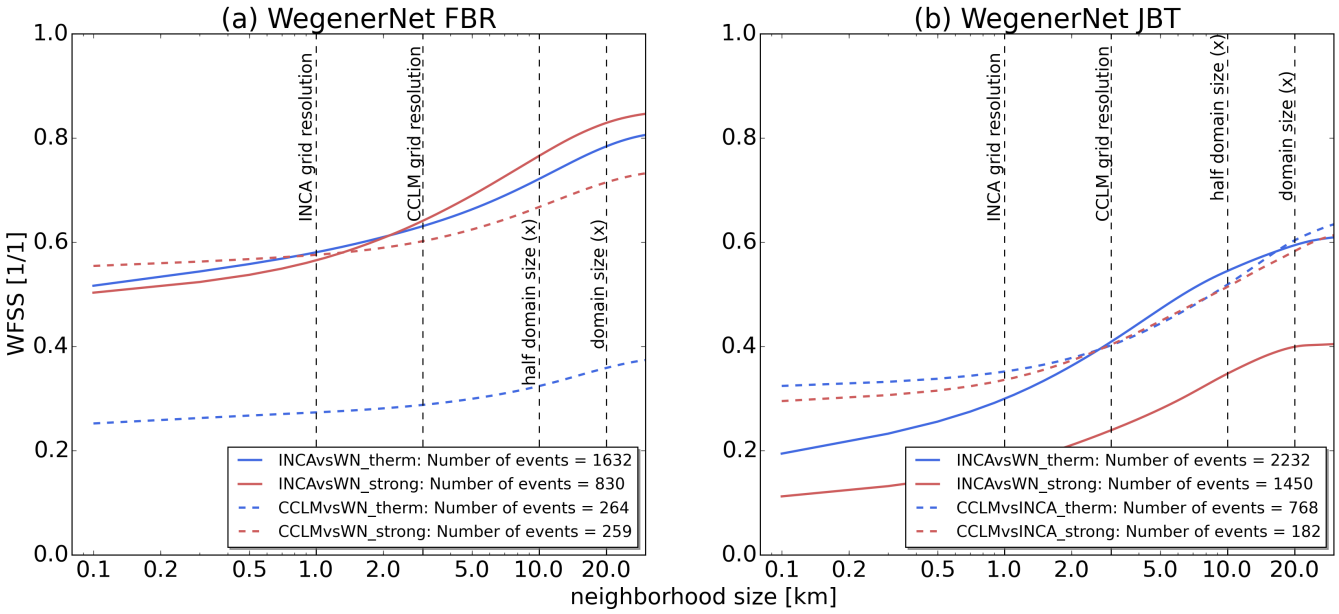

Figure 3. The event averaged Wind Fractions Skill Score (WFSS) results for the WegenerNet FBR (a), compared to the WegenerNet JBT (b), for the four defined evaluation cases in each region (see legend; indicating also the number of events included). See Table 1 for more information on the evaluation cases.

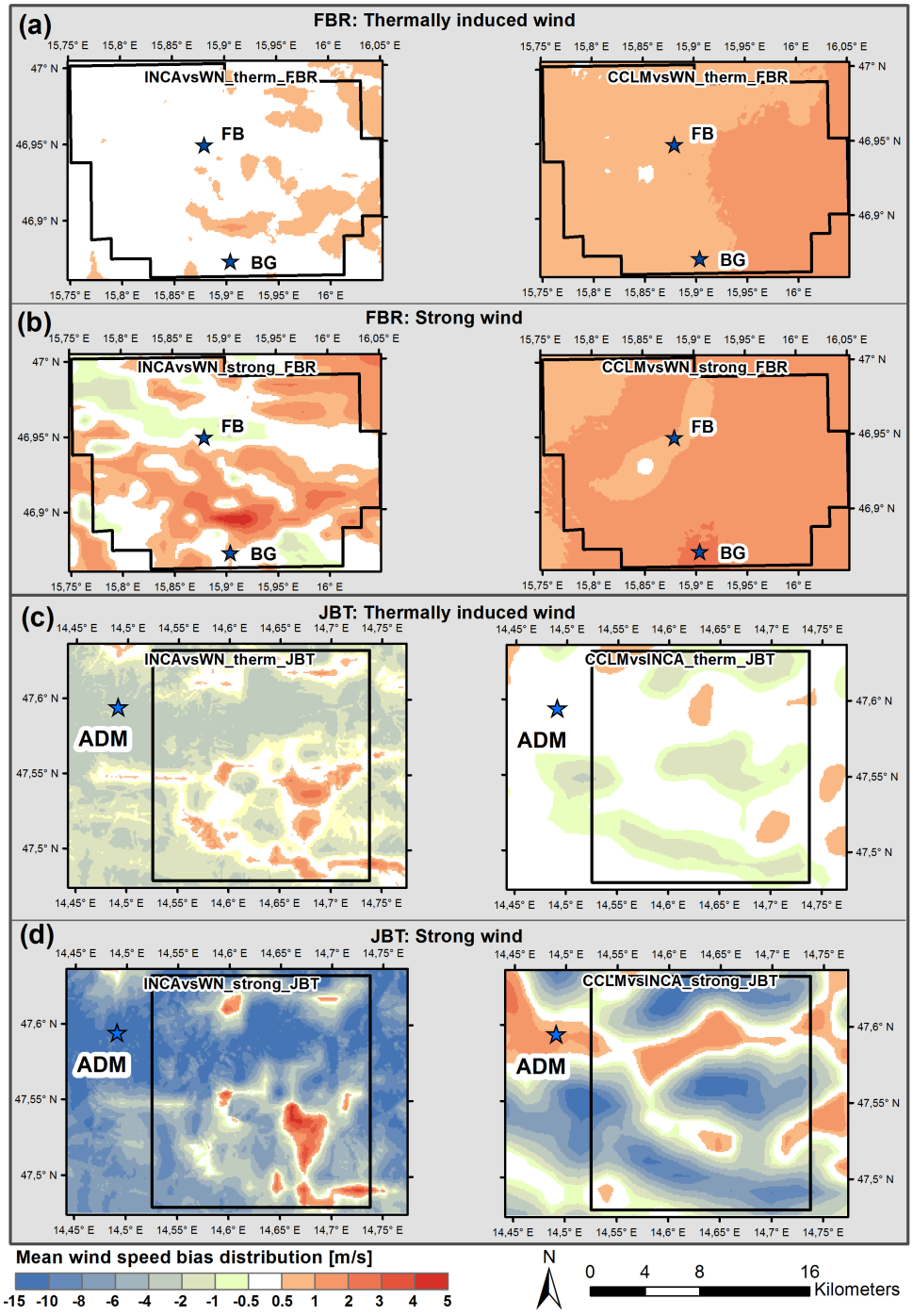

Figure 4. Mean wind speed bias distribution for the WegenertNet FBR (a, b) and the WegenerNet JBT (c, d): (a, b) INCA versus WegnerNet (left) and ~~COSMO~~ CCLM versus WegenerNet (right) for (a) thermally induced wind events and (b) strong wind events, and (c, d) WegenerNet versus INCA (left) and ~~COSMO~~ CCLM versus INCA (right) for (c) thermally induced wind events and (d) strong wind events.

Table 1. Characteristics of wind field evaluation cases used for the WegenerNet, INCA, and ~~COSMO~~ CCLM intercomparisons (top half for the WegenerNet FBR; bottom half for the WegenerNet JBT region).

Evaluation Case	*Type*	*Region*	*Modeled dataset*	*Reference dataset*	*Period*
			WegenerNet FBR		
INCAvsWN_therm_FBR	thermally	FBR	INCA	WN	2008-2017
INCAvsWN_strong_FBR	strong	FBR	INCA	WN	2008-2017
~~COSMOvsWN~~CCLMvsWN_therm_FBR	thermally	FBR	~~COSMO~~ CCLM	WN	2008-2010
~~COSMOvsWN~~CCLMvsWN_strong_FBR	strong	FBR	~~COSMO~~ CCLM	WN	2008-2010
			WegenerNet JBT		
INCAvsWN_therm_JBT	thermally	JBT	INCA	WN	2012-2017
INCAvsWN_strong_JBT	strong	JBT	INCA	WN	2012-2017
~~COSMOvsINCA~~CCLMvsINCA_therm_JBT	thermally	JBT	~~COSMO~~ CCLM	INCA	2008-2010
~~COSMOvsINCA~~CCLMvsINCA_strong_JBT	strong	JBT	~~COSMO~~ CCLM	INCA	2008-2010

Table 2. Limits for the selection of thermally induced or strong wind events for the defined evaluation cases shown in Table 1 (top half for the WegenerNet FBR; bottom half for the WegenerNet JBT).

Evaluation Case	*Variables*[a]	$\overline{v}$ [m s^{-1}] (*reference data*[b])	v [m s^{-1}] (*reference data*[b])	rh [%] (*reference data*[b])	$Q_{g,m}$ - $Q_{n(g),o}$ [W m^{-2}] (*reference data*[b])	Q_n [W m^{-2}] (*reference data*[b])	*Number of events*[c]
				WegenerNet FBR			
INCAvsWN_therm_FBR		–	<1.5 (RS_{dm})	<65.0 (RS_{dm})	<100.0 (RS_{dm})	<30.0 (RS_{nm})	1632
INCAvsWN_strong_FBR		>2.5 (WN_{dm})	>3.0 (WN_{hm})	–	–	–	830
~~COSMOvsWN~~CCLMvsWN_therm_FBR		–	<1.5 (RS_{dm})	<65.0 (RS_{dm})	<100.0 (RS_{dm})	<30.0 (RS_{nm})	~~1632~~ 264
~~COSMOvsWN~~CCLMvsWN_strong_FBR		>2.5 (WN_{dm})	>3.0 (WN_{hm})	–	–	–	259
				WegenerNet JBT			
INCAvsWN_therm_JBT		–	<2.0 (SCH_{dm})	<65.0 (SCH_{dm})	<20.0 (SCH_{dm})	<30.0 (SCH_{nm})	2232
INCAvsWN_strong_JBT		>9.5 (WN_{dm})	>9.0 ($INCA_{hm}$)	–	–	–	1450
~~COSMOvsINCA~~CCLMvsINCA_therm_JBT		–	–	<65.0 (WEI_{dm})	<20.0 (WEI_{dm})	–	768
~~COSMOvsINCA~~CCLMvsINCA_strong_JBT		>6.0 ($INCA_{dm}$)	>6.0 ($INCA_{hm}$)	–	–	–	182

[a] $\overline{v}$: average wind speed; v: wind speed; rh: relative humidity; $Q_{g,m}$ - $Q_{(n)g,o}$: difference between mean modeled global radiation ($Q_{g,m}$) and observed net radiation ($Q_{n,o}$) for the WegenerNet FBR and difference between $Q_{g,m}$ and observed global radiation ($Q_{g,o}$) for the WegenerNet JBT; Q_n: net radiation.

[b] RS: Reference Station (Nr. 77); SCH: Schroeckalm station; WEI: Weidendom station; dm: daytime mean value from observations (from sunrise till sunset); nm: nighttime mean value from observations (from sunset till sunrise); WN_{dm}: daily mean value from gridded WegenerNet wind speed; WN_{hm}: hourly mean value from gridded WegenerNet wind speed; $INCA_{dm}$: daily mean value from gridded INCA wind speed; $INCA_{hm}$: hourly mean value from gridded INCA wind speed.

[c] Hourly wind events; i.e., hours are used as the base period for the statistical analysis.

Table 3. Statistical performance parameters used for the intercomparison of the wind field modeling results.

Parameter	*Equation*	*Remarks*
Wind fractions skill score	$WFSS = 1 - \frac{\sum_k \sum_{i,j} [O_k(i,j) - M_k(i,j)]^2}{\sum_k \left\{ \sum_{i,j} O_k(i,j)^2 + \sum_{i,j} M_k(i,j)^2 \right\}}$	O_k: fraction values for observations for wind class ~~k~~ k at location ~~i~~i,~~j~~j; M_k: fraction values for ~~forecasts~~ model data for wind class ~~k~~ k at location ~~i~~i,~~j~~ j (Skok and Hladnik, 2018; Roberts, 2008)
Asymptotic WFSS	$AWFSS = 1 - \frac{\sum_k \left(f_k^O - f_k^M\right)^2}{\sum_k \left\{ \left(f_k^O\right)^2 + \left(f_k^M\right)^2 \right\}}$	f_k^O: frequency of wind class ~~k~~ k in the observations; f_k^M: frequency of wind class ~~k~~ k in the ~~forecast~~ model data (Skok and Hladnik, 2018)
Bias	$B = \frac{1}{N} \sum_{i=1}^{N} (v_{m,i} - v_{o,i})$	v_m: modeled wind speed; v_o: observed wind speed
Standard deviation of observed wind speed	$SD_o = \sqrt{\frac{1}{(N-1)} \sum_{i=1}^{N} (v_{o,i} - \overline{v_o})^2}$	v_o: observed wind speed; $\overline{v_o}$: mean observed wind speed
Root-mean-square error	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (v_{m,i} - v_{o,i})^2}$	v_m: modeled wind speed; v_o: observed wind speed
Correlation coefficient	$R = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(\frac{v_{m,i} - \overline{v_m}}{\sigma_m} \right) \left(\frac{v_{o,i} - \overline{v_o}}{\sigma_o} \right)$	v_m: modeled wind speed; $\overline{v_m}$: mean modeled wind speed; v_o: observed wind speed; $\overline{v_o}$: mean observed wind speed; σ_m: standard deviation of modeled wind speed; σ_o: standard deviation of observed wind speed
Mean absolute error of wind direction	$MAE_{\mathrm{dir}} = \frac{1}{N} \sum_{i=1}^{N} \{\arccos[\cos(\phi_{m,i} - \phi_{o,i})]\}$	ϕ_m: modeled wind direction; ϕ_o: observed wind direction

Table 4. Statistical performance measures calculated for the evaluation cases from Table 1, for the WegenerNet FBR and the WegenerNet JBT region. See Table 3 for more information on the calculation of the parameters.

Evaluation Case	$AWFSS$ ~~[1/1]~~ [1]	B [m s^{-1}]	SD_o [m s^{-1}]	$RMSE$ ~~[1/1]~~ [1]	R ~~[1/1]~~ [1]	MAE_{dir} [°]
			WegenerNet FBR			
INCAvsWN_therm_FBR	0.81	0.34	0.74	0.79	0.67	38
INCAvsWN_strong_FBR	0.85	0.50	1.04	1.66	0.34	14
~~COSMOvsWN~~CCLMvsWN_therm_FBR	0.38	1.32	0.72	1.85	0.37	55
~~COSMOvsWN~~CCLMvsWN_strong_FBR	0.74	1.01	1.03	1.28	0.57	14
Mean Value	0.70	0.79	0.88	1.40	0.49	30
			WegenerNet JBT			
INCAvsWN_therm_JBT	0.61	-1.37	2.37	2.97	0.20	68
INCAvsWN_strong_JBT	0.39	-6.69	3.97	8.60	0.16	39
~~COSMOvsINCA~~CCLMvsINCA_therm_JBT	0.64	-0.23	1.32	1.31	0.40	56
~~COSMOvsINCA~~CCLMvsINCA_strong_JBT	0.63	-3.79	4.24	5.52	0.08	25
Mean Value	0.57	-3.04	2.98	4.62	0.20	47