Author response for the reviewers regarding the manuscript *"The road weather model RoadSurf driven by the HARMONIE-Climate regional climate model: evaluation over Finland"*

We thank both reviewers for their comments and suggestions (in blue). Please find our detailed point-by-point responses below (in black).

We have made changes in the manuscript, and the changes are visualized at the end of this document. The pages and line numbers as well as the reference numbers for figures used in this response correspond to the ones used in the updated version.

Anonymous Referee #1

GENERAL COMMENTS:

This article presents an evaluation study across Finland of a 13-year long record of road surface temperatures and road ice, snow, and water storage parameters obtained with a road surface weather model driven by output from a regional climate model operated at 12.5 km resolution. The RCM is in turn forced at its lateral boundaries by atmospheric information from the ERA-Interim re-analysis. The emphasis of the analysis is on the performance of the road surface model compared with observations obtained at 25 road weather stations of which 11 were also equipped with optical sensors to establish the prevailing condition of the road surface.

Overall, the paper is coherently written, but in my opinion the scope is too much from the perspective of an NWP. The entire analysis assumes as if the modelling chain can be one-to-one compared with observations and a statistical machinery is applied resulting in skill scores which one usually sees in assessing the forecast performance of a prediction model. This approach does not match with the purpose of the study to evaluate a model system when operated in climate mode but still using observations (ERA-Interim) to constrain the large-scale model circulation to the observed synoptic-scale structure.
We thank the reviewer for the careful and thorough evaluation of our manuscript. We also apologize that we did not emphasize enough the reasoning behind our study that is to assess if RoadSurf can adequately capture the road weather conditions occurred in the current climate when forced with a reanalysis-driven regional climate model (RCM) HARMONIE-Climate (HCLIM). This is the first time such a modeling chain is evaluated and, therefore, we believe that forcing HCLIM with the ERA-Interim reanalysis product suits well this purpose instead of forcing HCLIM directly with global circulation models (GCMs). When forcing HCLIM by a reanalysis product, the large-scale model circulation is constrained by the observed synoptic-scale, as also mentioned by the reviewer. For example, Kotlarski et al. (2014) state that it is a standard procedure to carry out evaluation experiments using the (close to) perfect boundary settings in RCMs, which means using reanalysis product, such as ERA-Interim, to force the regional climate model in the lateral boundaries as it is done in our study. We have therefore clarified the goals of our study in the introduction (see P2–3 L58–64).

Regarding the NWP perspective, it is true that we have used daily data in the analysis of RoadSurf. On the other hand, the regional climate model HCLIM has been evaluated using standard metrics, mean seasonal biases for temperature and mean seasonal relative biases for precipitation using monthly values over 13 years. The daily time scale was chosen for the RoadSurf evaluation because the daily (and even hourly) temporal scales are the most relevant when studying road weather as also mentioned in the manuscript.

However, it is good to note that we have calculated the metrics using daily data obtained for 13 years separately for each month and taken multi-year monthly means of the daily values. This means that for example the mean biases for road surface temperature and mean daily fractions of road surface classes will be the same regardless the time scale in question (e.g. for the biases, the means of the monthly means of daily mean biases and the means of the monthly mean biases are the same; please see Equation 1 below). To better account for the fact that RoadSurf was forced by an RCM, we have performed a part of the analysis (sections 3.2.1, 3.2.3 and 3.2.4) at a (multi-year) monthly scale. The main conclusions stay the same due to the nature of calculating means as explained by the following equation (example for one month):

$$\frac{\sum_{d=1}^{30}\left(\frac{\sum_{h=1}^{24} M_h - O_h}{24}\right)_d}{30} = \frac{\sum_{d=1}^{30}\left(\frac{\sum_{h=1}^{24} M_h}{24}\right)_d}{30} - \frac{\sum_{d=1}^{30}\left(\frac{\sum_{h=1}^{24} O_h}{24}\right)_d}{30} \tag{1}$$

where *d* refers to day, *h* to hour, M_h to *h*th model value, and O_h to *h*th observed value.

The next step, also mentioned by the authors, will be to run the RCM-RoadSurf modelling system driven by GCM output resulting from transient multi-annual simulation under prescribed emission scenarios. The biases then found will presumably be much larger than seen in this evaluation study, and any performance rating as if it were a prediction model will be deemed meaningless. The primary reason for that is huge biases in circulation and regime statistics in the GCM drivers compared to ERA-Interim. So, the authors can better focus on the role of circulation and regime drivers on the performance of their modelling chain, than focus on skill scores like RMSE and Pearson's correlation coefficients. Eventually, they want to draw credible conclusions how climate change information at the large scale will propagate through their RCM to the RoadSurf model.
We agree with the referee that the performance of RoadSurf that is forced by ERA-Interim-driven HCLIM does not mean that we will obtain exactly the same results when HCLIM is forced by GCMs. However, we think the ERA-Interim evaluation is crucial before continuing to use this method further. If RoadSurf would not perform adequately when the input data is coming from an RCM forced by a reanalysis product (i.e. with perfect boundary settings), we see that it would not be appropriate to analyze the effects of climate change on road weather with this method. In our opinion, the evaluation of RoadSurf that is forced by a GCM-driven HCLIM would be a subject of its own study. However, this aspect will be looked at when RoadSurf's inputs are retrieved from a GCM-driven HCLIM. Please see also our comments above (the first paragraph of our response).

As road surface temperature is the main output parameter in RoadSurf, we have, however, added an analysis of the relationships between the road surface temperature biases and the biases in the input parameters of RoadSurf that are retrieved from HCLIM (please see section 3.2.1 starting from P12 L352).

In addition I would argue that the way the experiment has been set up makes it very difficult to conclude how the shortcomings in performance can be attributed to the model components that are used. Several times the authors mention that an issue might be related either to the warm and/or wet bias in their RCM or to features in the RoadSurf model. In that respect, I am wondering why the authors have not carried out a bias adjustment to HCLIM-ALARO temperature and precipitation which serve as forcings to the RoadSurf model. Such an additional experiment would have a twofold benefit: a) to disentangle the bias in HCLIM from issues in RoadSurf, and b) to obtain a measure to what extent the biases in HCLIM affect the performance of RoadSurf. The latter would be very helpful in the analysis and interpretation of future GCM-driven experiments.

We agree with the referee that bias-correction could be helpful in distinguishing the biases caused by the input retrieved from HCLIM and by RoadSurf itself. However, we would like to remind that the purpose of this study was to evaluate the whole modeling chain (i.e., the model biases in RoadSurf when forced by reanalysis-driven HCLIM) and to show that RoadSurf can reliably capture occurred road weather conditions in Finland when driven by an RCM.

We believe that carrying out bias-correction is therefore not in the scope of our study and would require a considerable amount of work, which might not be feasible in the context of this paper. We also see that bias-correction might not be very beneficial as the distributions of e.g. modeled daily minimum, mean, and maximum temperatures as well as precipitation are already relatively close to observations (See Figs. S2 and S3 in the updated manuscript). In addition, it might not always be very straightforward to use bias-correction, such as quantile mapping (that was suggested by the reviewer), for the future simulations as the error correction values might not be stationary (see e.g. Switanek et al., 2017). Moreover, e.g. Maraun et al. (2017) state that if regional feedbacks are not properly taken into account, bias-correction methods, such as quantile mapping, might lead to implausible regional climate change signals. Another problem can be the physical inconsistencies in the bias-corrected data (Schoetter et al., 2012).

But as said earlier, we have now added an analysis of the relationships between the road surface temperature biases and the biases in the input parameters at the road weather stations. Based on this analysis, the road surface temperature bias seems to be mainly explained by the variability of the air temperature bias (section 3.2.1 starting from P12 L352) as speculated in the first version of the manuscript.

MAJOR COMMENTS:

1) I would strongly suggest to focus on the Finland area from the beginning. The discussion of the HCLIM-ALARO model performance for the whole of the Fenno-Scandian domain is distracting. There are always huge issues in the mountainous areas in Norway, for any RCM, and also in E-OBS, but they are not relevant for this study. Focus on Finland in Figs 3 and 5.

Thank you for the suggestion. In the updated manuscript, we show results only over Finland (please see Figs. 3 and 6). Also, the discussion is now adjusted to cover only Finland for section ”3.1 Evaluation of HCLIM38-ALARO”.

2) Do not only examine the bias in the monthly mean temperature, but also at a number of percentiles (e.g. P5,25,75,95). The diurnal amplitude in model temperature compared to observations is relevant here as well.

The temperature percentiles (P5, P22, P75 & P95) have been added in the Supplement (Fig. S5) and are discussed in the new section 3.1.2. Moreover, we have included figures of the biases in daily minimum and maximum temperatures to account for the diurnal cycle (Fig. 5 in the new section 3.1.2).

It is still good to note that the purpose of our paper is not to evaluate very thoroughly the performance of HCLIM, but rather focus on the performance of RoadSurf.

3) Similarly for precipitation. In addition to mean precipitation look at wet-day frequency (threshold 0.3 or 1.0 mm/day), and perhaps some exceedance percentiles. It provides much more insight than an RMSE score.

Initially, we did not use RMSE scores for precipitation, but showed mean relative seasonal precipitation biases (Figs. 6 and 7). We have included a new figure on the wet-day frequency (Fig. S6) and added a discussion regarding this figure in section 3.1.3.

As said above, it is good to keep in mind that the purpose of our paper is not to evaluate very thoroughly the performance of HCLIM, but rather focus on the performance of RoadSurf.

4) Can there be said anything about the accuracy of the RCM inputs other than near-surface temperature and precipitation that are used to drive the RoadSurf model.
We have now added a brief evaluation of other input parameters (relative humidity, wind speed, as well as shortwave and longwave radiation) by comparing HCLIM model results and ERA5 reanalysis product (please see section 3.1.4). In addition, we have briefly evaluated the modeled total cloud fraction.

5) As mentioned in the general comments it would be useful to apply a bias-adjustment on daily mean temperature and precipitation, also frequency of occurrence, to bring the HCLIM-ALARO temperature (e.g. quantile-quantile) and precipitation forcing in the same "statistical" ballpark as the observations.
Please see our response above (the last paragraph of general comments).

6) As the RCM is operated at 12.5km resolution there should be reference to the efforts within EuroCordex in conducting evaluation (ERA-Interim driven) and transient (GCM driven) experiments at 12.5 km resolution across Europe with a variety of RCMs. For the evaluation study you best cite Kotlarski et al. (2014; doi:10.5194/gmd-7-1297-2014).
Thank you for pointing out this relevant reference. The EURO-CORDEX initiative has been mentioned in the introduction as well as in the discussion of the results obtained with HCLIM.

The following phrases were added:

"*Although high-resolution climate projections for Europe are currently available through the international climate downscaling initiative EURO-CORDEX that provides RCM data at 50 km (EUR-44) and 12.5 km (EUR-11) resolution (Jacob et al., 2014), the EURO-CORDEX dataset does not publicly include reanalysis-driven RCM simulations at very high temporal resolutions, such as 1-hourly.*" (P3 L64–67)

"*On the other hand, the HCLIM38-ALARO results for mean seasonal T_{air} were in agreement with EURO-CORDEX RCMs that were run at 12.5 km grid resolution. For instance, Kotlarski et al. (2014) showed that some of the ERA-Interim-driven EURO-CORDEX RCMs had a warm (cold) bias especially over the northern parts of Finland during the winter (summer).*" (P8 L230–233)

"*The results obtained for HCLIM38-ALARO showed similar magnitude and spatial patterns of the precipitation biases compared to other EURO-CORDEX RCMs that are mainly overestimating seasonal precipitation over Finland during the winter and summer as shown by Kotlarski et al. (2014).*" (P9 L282–285)

"*Overall, the HCLIM38-ALARO results were found to be in line with other EURO-CORDEX RCMs.*" (P19 L587–588)

7) Section 3.2.1 ("Road surface temperature"), after line 240 bothers me most. Why are all discrepancies blamed on the bias in temperature forcing, and not on potential issues with downwelling radiation, in particular biases in downwelling long wave radiation due to biases in cloud amount or cloud base.

We have now evaluated the biases in the downwelling radiation (both shortwave and longwave). The variability in the biases of downwelling longwave radiation seems to play a small role in explaining the variability of road surface temperature biases. However, the biases in the air temperature have a much larger impact (please see section 3.2.1 starting from P12 L352) as speculated in the first version of the manuscript.

The following phrases were added:

"*The analysis shown in Fig. 10 revealed that the variability of the monthly biases in T_{air} explained on average 57 % (range 19–84 % in October–April) of the variability of the monthly biases in T_{road} while the LW_d biases explained on average 16 % (range 2–34 % in October–March). Furthermore, the variability in SW_d biases was found to explain a small amount (4 %) of the variability in T_{road} biases during April.*" (P12 L355–358)

8) Page 9, L260-266. The authors argue that the better skill obtained with the forcing from the NWP compared to this study can be ascribed to the higher resolution at which the NWP is operated. I tend to disagree on that, in my opinion the use of data-assimilation when operating in NWP-mode will keep the model atmospheric state across the Finland region much closer to the observed state.

Thank you for pointing this out. However, we have compared the results only to the ones that are obtained without any data assimilation. This has been clarified in the text as follows:

"*For example, Karsisto et al. (2016) found that the biases in the simulated T_{road} varied between –1 and 1 ºC (mostly ±2 ºC in our study) at 20 stations in Finland during October and December 2013 when RoadSurf was driven by a high-resolution NWP version of HARMONIE (cy36h1.4) with a grid resolution of 2.5 km without any data assimilation. However, it is good to note that the results obtained in our study and by Karsisto et al. (2016) are not directly comparable since in their study RoadSurf was initialized using road weather station measurements for 48 hours and only the first forecasted hour was analyzed.*" (P13 L392–398).

9) The statistical methods used in sections 3.2.2. are not suitable for evaluation purposes, they belong to the realm of NWP verification. I advise to take this section out or move it to the supplement.

This old section 3.2.2 has been removed, and some parts of the discussion on the model performance from this section have been moved to section 3.2.1 (starting from P12 L382).

10) The same applies to section 3.2.5 although I find the message (i.e. over-representing of storage of ice, under-representation of storage of water) quite useful. So I would advise to move the technical method to the supplement but keep the message in the main body of the manuscript.

The technical method has been moved to the supplement, and only the results from POD-FAR analysis are kept (see section 3.2.4). In addition, we have added a figure and discussion on the multi-year sums of the occurrence of the storages for both model and observations thus avoiding a day-to-day comparison (Fig. S7; please see section 3.2.4 starting from P17 L528).

OTHER COMMENTS:

1) It must be mentioned in the abstract that the HCLIM-ALARO simulation is driven by ERA-Interim

This is now mentioned in the abstract as follows:

"*RoadSurf was driven by meteorological input data from the cycle 38 of the high-resolution regional climate model (RCM) HARMONIE-Climate (HCLIM38) with ALARO physics (HCLIM38-ALARO) and ERA-Interim forcing in the lateral boundaries.*" (P1 L10–11).

2) Abstract, L 13: remove "precisely"
Removed (P1 L15).

3) Abstract, L 14, 18: replace "lack" by "absence" According to the text in Line 99 "the model does not take into account wintertime road maintenance operations ...". From that line I conclude that there is no maintenance at all in the model. "Lack" may imply there is still some maintenance left. Please, adjust everywhere in the text, if needed.
Corrected (P1 L16, P1 L22, P4 L119, P16 L501, P18 L553, P18 L557, & P19 L608).

4) Abstract, L 17: remove "simulated", it is already implied by "warm bias".
Removed (P1 L19).

5) Introduction, L 24: "climate and weather information" → "weather and climate information".
Corrected (P1 L28).

6) Introduction, L34: "Finish temperatures ..." → " Finish temperature records ..."
Corrected as "Finnish temperature records" (P2 L38).

7) Introduction, L42: replace "reliable" by "plausible" or "credible". It is not a prediction.
Changed as "credible" (P2 L46)

8) Introduction, L65: "13 year long simulations" → "13-year long simulations".
Corrected (P3 L78).

9) Page 3, L85-88: mention the source of the sea-surface boundary conditions (SST and sea-ice extent (probably also ERA-Interim)
Yes, these parameters (SST and sea-ice concentration) are taken from ERA-Interim. This information has been added in the text as follows:

"*The sea-surface (sea-surface temperature and sea-ice concentration) and lateral boundary conditions of HCLIM38-ALARO were taken from ERA-Interim reanalysis (Dee et al., 2011) every 6 hours.*" (P4 L102–104).

10) Page 3, L92: "transfer in the ground ..." → "transfer into the ground ..."
Corrected (P4 L110).

11) Page 4, L95: ".. the elevation is taken into account ..." The elevation of what or with respect to what?
By elevation, we mean topography in general as RoadSurf otherwise assumes a flat surface. Modified as follows:

"*However, topography in general is taken into account implicitly through the input data.*" (P4 L114).

12) Page 4, L107: "... we did not include any forecast periods". Suggest to add the phrase "implying that no in-situ observations are used to initialize and force RoadSurf."
This has been added in the text (P4 L125–126).

13) Page 4, L119: Mention the version of the E-OBS dataset.
In the first version of the manuscript, we used the E-OBS version 17.0. In the updated manuscript, we have utilized the minimum and maximum temperatures provided by E-OBS version 19.0e (as suggested in the reviewer comment 14). Thus, we redid the analysis of mean temperatures and precipitation using the new ensemble version 19.0e. This information is now added in the text (P5 L140). In addition, we updated all the figures where E-OBS data was used (Figs. 3–4 & 6–7).

14) Page 4, L120: In addition to daily mean temperature, E-OBS also contains daily minimum/maximum temperature. Why not using these parameters for evaluation?
The minimum and maximum temperatures provided by the E-OBS version 19.0e have now been employed in the evaluation of HCLIM38-ALARO (please see section 3.1.2).

15) Page 4, L 122: remove "some"
Removed (P5 L146).

16) Page 5, L 127: remove "some"
Removed (P5 L151).

17) Page 6, L 180: the phrase "... such as from the possible biases in the input parameters ERA-Interim ..." is confusing. Do you mean that land-surface information from ERA-Interim is used in forcing HCLIM-ALARO, or does this statement refer to the lateral/sea-surface boundary conditions specified from ERA-Interim?
ERA-Interim's SST and sea-ice concentration are used to force HCLIM. If these input parameters from ERA-Interim are biased, it might affect the performance of HCLIM. This has been clarified in the text. This matter would need further evaluation, which is unfortunately not in the scope of this study.

The following phrases were added:
"*A prognostic lake model was included in the model version used in this study, and thus the warm bias might have stemmed from other reasons, such as from SURFEX's own features or the possible biases in ERA-Interim's sea-surface temperatures or sea-ice concentrations that are used to force the sea-surface in HCLIM.*" (P8 L227–230)

18) Page 7, L197: "Similarly than in ..." → "Similar to ..."
Corrected (P9 L269).

19) Page 8, L227: "... during different months" → "... for different months"
Corrected (P11 L335).

20) Page 9, L268: "earlier" → "before"
This part of the old section 3.2.2 is removed, and therefore this correction was not made.

21) Page 10, L296: "further" → "hence"
Corrected (P13 L386).

22) Page 10, L300-311: "the stations" → "stations" (about 11x)
This part of the old section 3.2.2 is removed, and therefore this correction was not made.

23) Page 10, L307: “It could be expected ...”→”It might be expected ...”
This part of the old section 3.2.2 is removed, and therefore this correction was not made.

24) Page 10, L316: “between the different stations”→“between stations”
This part of the old section 3.2.2 is removed, and therefore this correction was not made.

25) Page 10, L317: “hypothesized”→“speculated”
This part of the old section 3.2.2 is removed, and therefore this correction was not made.

26) Page 11, L343: “class occurred within a month”→”class occurring within a month”
Corrected (P15 L483).

27) Page 11, section 3.2.4 and Fig. 9 Perhaps you could briefly repeat that the road surface classes in the observations and the model do not entirely match.
This information has been repeated as follows:

“*It is good to remember that the observed and modeled road surface classes might not fully match as they are defined differently.*” (P16 L485–486).

”*The definitions of road surface classes differ slightly for the observations and model (e.g. the partly icy class is included only in the model)*.” (Fig. 12; P35 L1085–1086)

28) Page 12, L356-357: rephrase last part of sentence as “i.e., the tendency of the model to underestimate frost and to overestimate ice with the same magnitude.”
Rephrased as suggested (P16 L499–500).

29) Page 12, L360-361: “where much less maintenance”→“where maintenance” and “is performed compared to ...”→” is performed far less frequently compared to …”
Corrected (P16 L502–504).

30) Page, 12, L 362: “In real life,”→“In reality”
Corrected (P16 L505).

31) Page, 12, L 365-367: That is precisely the problem, because the bias in forcing temperature has not been adjusted the distinction between those two error sources cannot be made
We are discussing the possible reasons for the bias, so we would not call this a problem but rather a speculation. Most likely, both factors are contributing to the errors. No changes made.

32) Page, 12, L 375: No threshold used? Just, plainly 0 when the mean was 0?
Correct. This is because the daily mean values for storages were very small and very often the storages are plainly zero. We tested the same method using daily median and daily maximum values, but this did not considerably affect the results. However, we decided to use daily maximum values in the updated manuscript to avoid any confusion (see section 3.2.4).

33) Page, 12, L 379: “...storages might be slightly displaced or mistimed”. That is typical for NWP verification, but should not be relevant in an evaluation study.
This part has been modified as follows:

“*However, this method might penalize the model more than it should because the modeled storages were compared with observations using day-to-day values. For this reason, we additionally calculated the multi-year sums of all the modeled and observed daily cases with daily maximum more than zero or zero.*” (P17 L526–529).

34) Page, 14, L417-419: Second part of this sentence, “however ...” is unclear. Please rephrase.
Rephrased as:

”*However, underestimated frequency of snow cannot be explained by the snowpacks that are depleting too fast in the model. This is because the majority of the stations with an optical sensor utilized in this study are located in the southern parts of Finland where the modeled snowpacks might actually stay longer compared to the measurements as discussed before.*” (P18 L569–571).

35) Conclusions, L 420-423. Like in the abstract it should be stated that HCLIM-ALARO is driven by ERA-Interim re-analyses.
Thank you for pointing this out – this statement is included (P18 L580–581).

36) Conclusions, L 422: “the skill of HCLIM- ...” → ”the skill of the HCLIM- ...”
Corrected (P18 L579).

37) Conclusions, L 427: “undercath” → ”undercatch”
Corrected (P19 L587).

38) Conclusions, L 427-428: “the modeled domain” → ”the model domain”
Changed as ”Finland” (P19 L587).

39) Conclusions, L 432-433: Remove “However,”. Moreover, the absence of data-assimilation is most probably at least as relevant as the difference in horizontal resolution for explaining the poorer performance.
This sentence is modified as:

”*The coarser grid resolution of the HCLIM38-ALARO compared to the NWP model input used in the earlier studies might be the main reason for this outcome as no data assimilation was used for HCLIM38-ALARO or the NWP model.*” (P19 L595–597).

40) Conclusions, L 439: “This is of a great importance” → “This is of great importance”
Corrected (P19 L604).

41) Conclusions, L 439: “... are the most slippery...” → ”... are most prone to slippery conditions ...”
Corrected (P19 L604).

42) Conclusions, L442: “... than what the observations showed” → “ than is indicated by observations”
Corrected as ”*than indicated by observations*” (P19 L607–608).

43) Conclusions, L447: “the 13 year long ... period” → “the 13-year long ... period”.
Corrected (P19 L612).

44) Figure caption 1: Does the displayed domain include or exclude the boundary relaxation zone? How wide is the zone in terms of grid points? The color “yellow” for Northern Finland is very hard to distinguish.

The extension zone of 11 grid points has been removed from the figure. However, the figure includes an 8-point wide relaxation zone. This information has been added to the figure caption (see Fig. 1). The colors are changed for Fig. 1 and 2, so that yellow color is not used.

The following phrases were added:

”*Figure 1 depicts the HCLIM38-ALARO simulated domain along with the model’s 8-point wide relaxation zone as well as the regions of Finland that are analyzed in more detail in this study.*” (P4 L101–102)

”*The transparent areas depict the model’s 8-point wide relaxation zone.*” (Fig. 1; P26 L826)

Anonymous Referee #2

This paper evaluates the RoadSurf model forced with output from a regional climate model (HARMONIE-Climate). The RoadSurf is used operationally to simulate road conditions for the benefit of the public. Here, the authors extend RoadSurf by forcing it with output from a regional climate model. This successful endeavor then paves the way to make assessments of future road conditions under climate change by forcing RoadSurf with output from a projection-period regional climate simulation.

The paper is easy to read and understand. I am not an expert in road modeling, so it is difficult to criticize anything about the RoadSurf model. I certainly couldn’t identify any glaring deficiencies. Much of the paper is devoted to assessing the skill of the regional climate model. There are biases and problems, as one would expect, but even with these biases, the RoadSurf model is able to reasonably replicate what is observed at the observed road sites. Clearly, it would be even more powerful if the simulation forced with regional climate model output could be compared to results with bias-corrected forcing or local forcing, but that may not really be feasible. So, in the context of the purpose of the paper, which is to assess whether or not RoadSurf forced with a regional climate model has the potential to provide useful information on Road conditions now and in the future, I would say that the authors have demonstrated this to be the case.

So, overall, I find this paper suitable for publication in close to it’s current form. Will be interesting to see what happens when they run with climate change scenarios.

We thank the referee for the positive feedback on our manuscript. Evaluation of RoadSurf using local forcing would be interesting, but this is not feasible as the road weather stations in Finland do not observe solar radiation and precipitation measurements are considered unreliable (Kangas et al., 2015).

However, we have now added an analysis of the relationships between the road surface temperature biases and the biases in the input parameters at the road weather stations. Based on this analysis, the variability in the road surface temperature biases seems to be mainly explained by the variability in the air temperature biases (please see section 3.2.1 starting from P12 L352) as speculated in the first version of the manuscript.

References:

Kangas, M., Heikinheimo, M., and Hippi, M.: RoadSurf: a modelling system for predicting road weather and road surface conditions, Meteorol. Appl., 22, 544–553, https://doi.org/10.1002/met.1486, 2015.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297-1333, https://doi.org/10.5194/gmd-7-1297-2014, 2014.

Maraun, D., Shepherd, T., Widmann, M., Zappa, G., Walton, D., and Gutiérrez, J. et al.: Towards process-informed bias correction of climate change simulations, Nat. Clim. Change, 7, 764–773, https://doi.org/10.1038/nclimate3418, 2017.

Schoetter, R., Hoffmann, P., Rechid, D., and Schlünzen, K.: Evaluation and Bias Correction of Regional Climate Model Results Using Model Evaluation Measures, J. Appl. Meteorol. Clim., 51, 1670–1684, https://doi.org/10.1175/jamc-d-11-0161.1, 2012.

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., and Demaria, E. M. C.: Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes, Hydrol. Earth Syst. Sci., 21, 2649-2666, https://doi.org/10.5194/hess-21-2649-2017, 2017.

The road weather model RoadSurf (v6.60b) driven by the ~~HARMONIE-Climate~~ regional climate model HCLIM38: evaluation over Finland

Erika Toivonen[1], Marjo Hippi[1], Hannele Korhonen[1], Ari Laaksonen[1,2], Markku Kangas[1], and Joni-Pekka Pietikäinen[1]

[1]Finnish Meteorological Institute, Helsinki, Finland
[2]Department of Applied Physics, University of Eastern Finland,~~ (PL 1627) 70211~~ Kuopio, Finland

Correspondence to: Erika Toivonen (erika.toivonen@fmi.fi) and Joni-Pekka Pietikäinen (joni-pekka.pietikainen@fmi.fi)

Abstract. In this paper, we evaluate the skill of the road weather model RoadSurf to reproduce present-day road weather conditions in Finland. RoadSurf was driven by meteorological input data from the cycle 38 of the high-resolution regional climate model (RCM) HARMONIE-Climate (HCLIM38) ~~utilizing~~with ALARO physics (HCLIM38-ALARO) and ERA-Interim forcing in the lateral boundaries. Simulated road surface temperatures and road surface conditions were compared to observations between 2002 and 2014 at 25 road weather stations located in different parts of Finland. The main characteristics of road weather conditions were accurately captured by RoadSurf in the study area. For example, the model ~~precisely~~ simulated road surface temperatures with a mean monthly bias of –0.3 ºC~~,~~ and mean absolute error ~~RMSE~~of ~~2.1~~0.9 ºC~~, and Pearson's correlation coefficient of 0.93~~. The RoadSurf's output bias most probably stemmed from the ~~lack~~absence of road maintenance operations in the model, such as snow ploughing and salting, and the biases in the input meteorological data. The biases in the input data were most evident in northern parts of Finland, where the regional climate model HCLIM38-ALARO overestimated precipitation and had a warm bias in ~~simulated~~ near-surface air temperatures during the winter season. ~~In turn~~Moreover, the~~se~~ variability of the ~~input data~~ biases in air temperature ~~seemed~~was found to explain on average 57 % of the variability of the biases in road surface temperature.~~to result in a warm bias in simulated road surface temperatures.~~ ~~Furthermore~~On the other hand, the ~~lack~~absence of road maintenance operations in the model might have affected RoadSurf's ability to simulate road surface conditions: ~~t~~The model tended to overestimate icy and snowy road surfaces and underestimate the occurrence of water on the road. However, the overall good performance of RoadSurf implies that this approach can be used to study the impacts of climate change on road weather conditions in Finland by forcing RoadSurf by future climate projections from RCMs, such as HCLIM.

1 Introduction

Road traffic sector is one field benefiting from improved regional ~~climate and~~ weather and climate information, especially at northern high latitudes. These regions do not only experience frequent wintertime snow and ice conditions~~,~~ but also rapidly changing road weather due to, for instance, the onset of snowfall (Juga et al., 2012) or during temperature variations around the freezing point (Kangas et al., 2015). Systematic consideration of upcoming weather events helps the general public in their every-day commute and, furthermore, road maintenance authorities to attend the roads in a cost-effective manner (Nurmi et al., 2013). In Finland, the Finnish Meteorological Institute (FMI) has a duty to issue warnings of hazardous traffic conditions to the general public. To support this, the institute has developed a road weather model RoadSurf which has been in operational use since 2000 (Kangas et al., 2015).

Road weather conditions are expected to be affected by ongoing anthropogenic climate change (e.g. Jaroszweski et al., 2014) throughout the inhabited northern high latitudes. This region is strongly impacted by the Arctic amplification of climate warming (Screen, 2014), which can clearly be seen, for instance, in the Finnish temperature~~s~~ records of the past 170 years (Mikkonen at al., 2015). The expected warmer and wetter future climate implies new challenges for road maintenance and traffic safety, especially in the southern parts of Finland: Precipitation events are likely to shift towards less snowfall and more frequent rain and sleet episodes (Räisänen, 2016). This kind of change in climate will decrease snowy road conditions, but at the same time increase the occurrence of wet road surfaces, which could lead to more frequently observed slippery and icy road conditions during the coldest times of a day, such as nighttime (Andersson and Chapman, 2011a). Moreover, the events of temperature change around the freezing point might become more frequent in the northern parts of Finland (Makkonen et al., 2014) leading to an increased occurrence of black ice conditions and making the roads more vulnerable to erosion. Therefore, policymakers and other stakeholders should have ~~an~~ access to ~~reliable~~credible regional climate projections that can provide a solid basis for informed impact assessments and adaptation measures in the road weather sector. A central tool for producing such projections are high-resolution regional climate models (RCMs).

Although the impacts of climate change on road weather, safety, and design have been assessed in many studies (see e.g. Koetse and Rietveld, 2009), most of these studies have only considered relative changes in air temperature and precipitation and related these to the possible impacts on the roads (e.g. Andersson and Chapman, 2011a; Andersson and Chapman, 2011b; Hambly et al., 2013; Hori et al., 2018; Makkonen et al., 2014). It would be beneficial to study the climate change impacts on, for instance, road surface temperatures (T_{road}) or road surface conditions using an approach in which these impacts can be accessed more directly. Furthermore, as ~~snowy and icy road~~ slippery road conditions, such as snowy or icy roads, are the major cause for the wintertime and weather-related road accidents in Fenno-Scandia (Andersson and Chapman, 2011b; Malin et al., 2019; Salli et al., 2008), it is essential to estimate how frequently these conditions will occur in the future.

The main goal of this paper is to evaluate the skill of RoadSurf to reproduce present-day road weather conditions in Finland when driven by a state-of-the-art high-resolution RCM, the cycle 38 of the HIRLAM-ALADIN Regional Mesoscale Operational Numerical Weather Prediction (NWP) In Europe (HARMONIE) Climate (HCLIM) (Lindstedt et al., 2015). HCLIM is forced by the ERA-Interim reanalysis product (Dee et al., 2011) in the lateral boundaries since it is a standard procedure to carry out evaluation experiments using the (close to) perfect boundary settings in RCMs (e.g. Kotlarski et al., 2014). This is the first time such a modelling chain is evaluated, and therefore ~~T~~this evaluation is needed in order to build and study future scenarios of road weather in this area with ~~a~~ larger confidence. Although high-resolution climate projections for Europe are currently available through the international climate downscaling initiative EURO-CORDEX that provides RCM data at 50 km (EUR-44) and 12.5 km (EUR-11) resolution (Jacob et al., 2014), the EURO-CORDEX dataset does not publicly include reanalysis-driven RCM simulations at very high temporal resolutions, such as 1-hourly. Therefore, ~~M~~meteorological input data for RoadSurf is taken from ~~the HIRLAM-ALADIN Regional Mesoscale Operational Numerical Weather Prediction (NWP) In Europe (HARMONIE) Climate (~~HCLIM~~) (Lindstedt et al., 2015) regional climate model~~ which is run for the years 2002–2014 with ALARO physics (Gerard, 2007; Gerard et al., 2009; Piriou et al., 2007) at 12.5 km resolution. These ~~RCM~~ HCLIM simulations are evaluated against ~~a~~ standard meteorological datasets, E-OBS v19.0e (Cornes et al., 2018) and the ERA5 reanalysis product (C3S, 2017), over F~~enno-Scandia~~inland.

In the previous studies, mainly NWP model outputs have been used to force RoadSurf. The simulated road weather parameters, such as T_{road}, have been verified against observations over Finland (Karsisto et al., 2016) and the Netherlands (Karsisto et al., 2017). In addition, Kangas et al. (2015) have studied RoadSurf's ability to simulate the amount of water, snow, frost, and ice on the road (called storage terms in RoadSurf) as well as road surface conditions and friction values, although only for two road weather stations in Finland. These studies have considered relatively short verification periods varying from 1 week to some months. In this paper, we concentrate on 13-year long simulations of HCLIM and HCLIM-driven RoadSurf. First, the performance of HCLIM is evaluated by comparing the model results with ~~a gridded observation dataset~~ E-OBS v19.0e dataset of near-surface air temperature and precipitation and with ERA5 reanalysis for downwelling shortwave and longwave radiation, relative humidity, and wind speed. All of these parameters are used as inputs for RoadSurf. This comparison is followed by an evaluation of HCLIM-driven RoadSurf ~~-HCLIM configuration~~ against observations at 25 road weather stations located in Finland. The focus is on T_{road}, but also the simulated road surface conditions and storage terms are compared to the observations. In addition, this study investigates the role of the biases in the HCLIM data ~~road weather station's local features, such as location, surrounding characteristics, and road maintenance class,~~ on the ~~model~~ biases in road surface temperature produced by HCLIM-driven RoadSurf.

2 Models and data

2.1 Models

2.1.1 HARMONIE-Climate (HCLIM)

HARMONIE is a seamless NWP model framework developed in collaboration with several European national meteorological services (Bengtsson et al., 2017). The nonhydrostatic and spectral dynamical cores in HARMONIE are provided by the ALADIN–NH (Bénard et al., 2010) which solves the fully compressible Euler equations using a two-time level, semi-implicit, semi-Lagrangian discretization on an Arakawa A grid. This study applied a model setup using the cy38h1 climate model version of HARMONIE with ALARO physics (HCLIM38-ALARO hereafter), as mentioned before,

~~and~~ a hydrostatic version of the dynamical core as well as a timestep of 300 seconds. The HCLIM38-ALARO version used in this study includes a lake model Flake (Mironov, 2008; Mironov et al., 2010) and a surface parameterization framework, surface externalisée (SURFEX) (Masson et al., 2013). A more thorough description of HCLIM can be found in Lindstedt et al. (2015).

For this study, HCLIM38-ALARO was run from January 2002 to December 2014 (years 2000 and 2001 as a spin up) over the Fenno-Scandian domain (151 x 181 grid boxes) with 12.5 km x 12.5 km horizontal grid resolution and 65 vertical layers. Figure 1 depicts the HCLIM38-ALARO simulated domain along with the model's 8-point wide relaxation zone as well as the regions of Finland that are analyzed in more detail in this study. The sea-surface (sea-surface temperature and sea-ice concentration) and lateral boundary conditions of HCLIM38-ALARO were taken from ERA-Interim reanalysis (Dee et al., 2011) every 6 hours~~, and the HCLIM-ALARO's output data was used to force RoadSurf offline~~. In this study, the HCLIM38-ALARO output parameters were produced every full hour and were used to force RoadSurf offline.

2.1.2 RoadSurf

The road weather model RoadSurf used in this study is a 1D model based on solving the energy balance at the ground surface. This study employed the RoadSurf version 6.60b which is the operational version of the FMI's research department with slight I/O changes made for this study. The model takes into account the conditions at the road surface and beneath it, and calculates the vertical heat transfer into the ground as well as at the interface of ground and atmosphere. Hydrological processes, such as accumulation of rain and snow, run-off from the surface, sublimation, freezing, melting, and evaporation, are parameterized. The model estimates road surface friction using a numerical-statistical equation (Juga et al., 2013). RoadSurf assumes a flat horizontal surface which does not have any shading elements, such as trees. However, ~~the elevation~~topography in general is taken into account implicitly through the input data. Thermodynamic properties of the road surface and ground are assumed to be similar for all simulated points, and the first two layers of the surface are always described as asphalt. In addition, the effect of traffic on the road surface is included: The model assumes that traffic packs some part of the snow into ice whereas the remaining part is assumed to be blown away from the road. However, the model does not take into account wintertime road maintenance operations, such as salting and snow ploughing, because RoadSurf is also used to plan and optimize these maintenance actions. The ~~lack~~absence of road maintenance in the model implies that there will be unavoidable discrepancies when comparing the modeled and observed road weather conditions.

As inputs, RoadSurf needs near-surface air temperature (T_{air}), near-surface relative humidity (*RH*), 10-meter wind speed (*WS*), precipitation (*Pr*) as well as ~~incoming~~downwelling shortwave (SW_d) and longwave (LW_d) radiation. In the operational use, the model employs observations from road weather stations, meteorological SYNOP weather stations, and radar precipitation networks to initialize road conditions while the road weather is predicted for the upcoming days utilizing forecasts produced by NWP models. In this study, we did not include any forecasted periods implying that no in-situ observations were used to initialize and force RoadSurf. Instead, RoadSurf was modified so that it utilizes the RCM data, in this case, the output of reanalysis-driven HCLIM38-ALARO. In addition to the above-mentioned inputs needed by

RoadSurf, we utilized the bottom layer ground temperature (at the depth of 4.28 m) produced by HCLIM38-ALARO. Using the simulated ground temperature instead of climatological one was motivated by the fact that although in the original RoadSurf version this temperature is assumed to vary sinusoidally, it is estimated by an equation in which some of the parameter values are based on measurements retrieved from only one FMI observatory located in Southern Finland. RoadSurf's main outputs are T_{road} and a traffic index describing driving conditions, but the model produces also surface friction, prevailing road conditions, and the sizes of water, snow, and ice storages on the road. RoadSurf divides the road surfaces into eight classes: 'dry', 'damp', 'wet', 'wet snow', 'frosty', 'partly icy', 'icy', and 'dry snow'. This classification is mainly based on the storage terms and T_{road}. The model physics of RoadSurf are described in more detail in Kangas et al. (2015).

2.2 Evaluation data

2.2.1 E-OBS dataset of ~~G~~gridded daily precipitation and near-surface air temperature ~~dataset~~

The HCLIM38-ALARO simulated daily precipitation and near-surface air temperatures were compared with the E-OBS dataset, version 19.0e (~~Haylock et al., 2008~~Cornes et al., 2018), which consists of daily precipitation and 2 m air temperature (daily minimum, mean, and maximum) data retrieved from stations located in Europe. The data is available as a~~n~~ regular ~~interpolated~~ grid which covers the pan-European domain with a resolution of 0.~~25~~11° (approximately ~~27.5~~12 km). This E-OBS version 19.0e consists of a 100-member ensemble of realizations for each daily field. We utilized ensemble means that can be taken as grid box averages (Cornes et al., 2018) and that are comparable to the best guess grid in the earlier versions of E-OBS (Haylock et al., 2008).

In general, gridded datasets, such as E-OBS, include ~~some~~ uncertainties due to the use of point measurements (e.g. rain gauges) and interpolation procedures. For example, the undercatch of precipitation can lead to high biases especially in winter at high latitudes as well as in the areas of rough topography (e.g. Prein and Gobiet, 2017). These undercatch errors are typically between 3 and 20 % for rainfall and up to 40 % (for shielded) or even up to 80 % (for non-shielded gauges) for snow (Goodison et al., 1998). Moreover, the accuracy and success of the E-OBS dataset depend~~s~~ on the number of stations used in the ~~interpolation~~gridding process (Cornes et al., 2018): The sparse station density can introduce ~~some~~ errors into the ~~interpolated~~gridded dataset (e.g. Prein and Gobiet, 2017). For Finland, the station density is sparser in the northern parts compared to the south (Fig. S1). Although these observational uncertainties are not in the scope of this study, they should be kept in mind when analyzing the results.

The comparison of modeled and observed data was performed using the coarsest grid resolution. The HCLIM38-ALARO model results ~~for the whole simulated domain~~ covering Finland~~enno-Scandia~~ were thus compared with E-OBS by remapping the ~~modeled~~ E-OBS values into the ~~E-OBS~~ grid of HCLIM38-ALARO: temperature data by using bilinear and precipitation data by using first-order conservative remapping. ~~The analysis does not include the model's relaxation zone where the latelar forcing influences the model results. In addition, t~~The areas with a lake fraction greater than or equal to 0.5 have been excluded from the analysis because E-OBS data over the lakes is based on the interpolation of the measurements over land. Moreover, the modeled 2 m air temperature values have been corrected using a lapse rate of 0.0064 ºC m^{-1} to account for the differences between the orography in the E-OBS dataset and the model. A standard Student's t-test at a 95 percent confidence level was used to assess the significance of the differences between the modeled and observed monthly averages (in case of temperature) or monthly sums (in case of precipitation).

2.2.2 ERA5 reanalysis product

Reanalysis is a scientific method that is based on a combination of data assimilation and numerical models. The fifth generation of the ECMWF's atmospheric reanalyses of the global climate, ERA5, provides hourly atmospheric data estimates at a grid horizontal resolution of approximately 30 km (Hersbach et al., 2018). This product was created using 4D-Var data assimilation and the ECMWF's Integrated Forecast System (IFS) cycle 41r2 that was used as the operational medium-range forecasting system in 2016. The model includes 137 levels in the vertical reaching to 1 Pascal. Overall, ERA5 assimilates more observations compared to ERA-Interim reanalysis product.

We utilized the monthly means of daily means for SW_d, LW_d, 10-meter WS, and near-surface RH to evaluate the performance of HCLIM38-ALARO. Monthly means of daily mean RH were computed employing ERA5 product of hourly near-surface T_{air} and dew point temperature (T_{dew}) ($RH = 100 * e_s(T_{dew})/e_s(T_{air})$) as RH is not archived directly in the ERA5 dataset. Saturation vapor pressure (e_s) was calculated based on the Magnus formula and with respect to water (WMO, 2008). Modeled near-surface RH was directly available and used as such.

Similarly to the comparison with the E-OBS data, the evaluation was carried out using the coarsest grid resolution by remapping HCLIM38-ALARO model results into the ERA5 grid using bilinear interpolation. Again, a standard Student's t-test at a 95 percent confidence level was used to assess the significance of the differences between the modeled and observed monthly averages (in case of LW_d, WS, and RH) or seasonal averages (in case of SW_d).

2.2.~~2~~3 Road weather stations

The results obtained by RoadSurf-HCLIM configuration were compared with observations retrieved from 25 road weather stations located in different regions of Finland. Table 1 describes the features of these stations, such as location, surrounding characteristics, road maintenance class, and the monthly average air temperatures, during October and April from 2002 to 2014. Stations 1–8 are located in Southern Finland, stations 9–13 in Western and Central Finland, stations 14–16 in Eastern Finland, stations 17–21 in Northern Finland, and stations 22–25 in Lapland (Fig. 2). The model grid cell closest to each of these stations was selected for evaluation. However, it needs to be noted that the model output represents an areal average over the whole model grid cell whereas the road weather observations are point measurements.

The road weather stations are equipped with the Vaisala ROSA road weather package and Vaisala DRS511 sensors (Vaisala, 2018a) which are installed in the road surface. Thirteen of the selected stations included also the Vaisala DSC111 optical sensor (Vaisala, 2018b) which provides information on, for instance, water, snow, and ice storages on the road. Two of the stations with an optical sensor had a large amount of missing data and, therefore, only eleven of them were included in this study. This study employs the road surface temperature and the information on the road surface classes provided by the ROSA stations and the storage terms provided by the stations with the additional optical sensors. Data availability was on average 79 % (range 57–91 %) at ROSA stations and 32 % (range 18–38 %) at stations with the optical sensor during the study period of 2002–2014.

The classification of observed and modeled road surface conditions differ slightly. For example, the observations included ’damp and salty’ as well as ’wet and salty’ road surface classes. These classes were combined with ’damp’ and ’wet’, respectively~~,~~ because RoadSurf does not include information on salting of the roads. The ’wet snow’ and ’dry snow’ classes provided by RoadSurf were also grouped together considering that observations did not have a directly comparable class for wet snow. In addition, observations do not include a ’partly icy’ class which is defined in the model. Therefore, these divergent definitions of road condition classes might cause some discrepancies when comparing the modeled and observed road conditions.

3 Results and discussion

3.1 Evaluation of HCLIM38-ALARO

3.1.1 Mean ~~N~~near-surface air temperature

The HCLIM38-ALARO model accurately captured the daily and seasonal mean 2 m air temperatures (T_{air}) over ~~the Fenno-Scandian domain~~Finland between 2002 and 2014. This is confirmed by Fig. S2 which illustrates the probability density functions (PDF) of the daily T_{air} for the observations and model during different seasons over all the grid points falling over Finland. Overall, the general shapes of T_{air} distributions were correctly reproduced by HCLIM38-ALARO with the largest deviations found in the winter season (December–February).

Also, ~~Fig. 3 which illustrates~~ the multi-year mean seasonal T_{air} was well captured by HCLIM38-ALARO. Figure 3 shows the seasonal means from E-OBS as well as the mean biases in the HCLIM38-ALARO simulated mean seasonal T_{air} with a reference to E-OBS. The stippled areas depict significant differences indicated by the Student’s t-test ($p < 0.05$). The mean biases averaged over ~~the whole domain~~Finland~~-~~ were slightly positive in the autumn and winter (September–February) and negative in the spring ~~all seasons of which~~and ~~the~~ summer ~~season~~ (~~June~~March–August). The autumn season had the smallest domain-averaged bias of ~~–0.25~~0.004 ºC and the ~~spring~~summer season ~~(March–May)~~ the highest domain-averaged bias of –0.~~68~~40 ºC. The biases were statistically significant mainly over ~~the mountainous areas in Norway~~the northern parts of Finland where the model had an enhanced warm bias in the winter and cold bias in the summer. These~~is~~ ~~error~~biases might ~~have~~ partly be caused by ~~the complex topography and~~ the lower station density in the northernmost domain, which might decrease the accuracy of the E-OBS data. On the other hand, the model was in good agreement with the observations ~~over~~

~~Sweden, Finland, and the Baltic countries~~during the spring and autumn~~, where~~ when most of the differences were not statistically significant.

~~The summer season was especially well captured by HCLIM-ALARO, but, interestingly, there was a statistically significant positive bias in the winter season in the northern Sweden and Finnish Lapland.~~ It is good to note that Lindstedt et al. (2015) encountered similar warm biases in their HCLIM-ALARO simulations with the cycle 36 over Sweden during the wintertime and they suggested it might originate from the non-prognostic lake surface temperatures. A prognostic lake model was included in the model version used in this study, and thus the warm bias might have stemmed from other reasons, such as from SURFEX's own features or the possible biases in ~~the input parameters (~~ERA-Interim's~~)~~ sea-surface temperatures or sea-ice concentrations that are used to force the sea-surface in HCLIM~~ or from SURFEX's own features~~. On the other hand, the HCLIM38-ALARO results for mean seasonal T_{air} were in agreement with EURO-CORDEX RCMs that were run at 12.5 km grid resolution. For instance, Kotlarski et al. (2014) showed that some of the ERA-Interim-driven EURO-CORDEX RCMs had a warm (cold) bias especially over the northern parts of Finland during the winter (summer). However, a more detailed analysis of the causes of the model biases is out of the scope of this study.

Figure 4 demonstrates that the mean monthly biases in the simulated daily T_{air} with a reference to the E-OBS dataset were generally between ±1 ºC when the biases were averaged over different regions of Finland for the period of 2002–2014. The highest positive biases occurred in the winter season and the highest negative biases in the summer as discussed before. However, some regional differences were apparent. For example, in Southern Finland, the biases were mainly negative during the autumn and winter months (October–February). Similarly, the biases were negative at the beginning of the winter season in Western and Central Finland, but the biases during the late winter and early spring season were positive as opposed to the negative biases in Southern Finland (excluding March when the bias in Southern Finland was also positive). In Eastern Finland, the mean biases resembled Western and Central Finland but were slightly higher for every month except for July, November, and December. The monthly biases were even higher in Northern Finland and Lapland compared to the other parts of Finland. In the northernmost areas, the biases were mostly positive during the autumn and winter seasons and negative during the spring and summer.

3.1.2 Minimum and maximum near-surface air temperature and percentiles of mean temperature

Similarly to the mean near-surface T_{air}, we assessed the differences between the observed and modeled daily PDFs as well as the multi-year seasonal means of daily minimum and maximum near-surface temperatures ($T_{air,min}$ and $T_{air,max}$, respectively) in 2002–2014 over Finland. Again, the PDFs of both $T_{air,min}$ and $T_{air,max}$ were adequately represented in HCLIM38-ALARO with the largest deviations in the winter season (not shown). Figure 5 shows that the multi-year seasonal means of $T_{air,min}$ were mainly overestimated and, contrarily, $T_{air,max}$ underestimated. The stippled areas in Fig. 5 depict significant differences pointed out by the Student's t-test ($p < 0.05$). The differences between HCLIM38-ALARO and E-OBS were significant mainly in the winter and summer season for $T_{air,min}$ with the largest domain-averaged difference of 1.73 °C found in the winter. For $T_{air,max}$, the differences were significant mostly in the summer with also the largest domain-averaged difference of –2.03 °C occurring in the summertime.

In addition to daily minimum and maximum temperatures, the differences in the 5th, 25th, 75th as well as 95th percentiles of the daily mean T_{air} between the model and observations were computed for different seasons (Fig. S5). The spatial differences for each season and over all the percentiles were similar to each other with generally more positive biases found for the 5th percentile and more negative biases for the 95th percentile (excluding the autumn), which is in line with the results for $T_{air,min}$ that is overestimated and $T_{air,max}$ that is underestimated. In the winter, Finland could clearly be divided into two regions as the biases were positive in the northern parts of Finland and negative in the south (excluding the 5th percentile). For all seasons, the maximum biases in the 5th, 25th, and 75th percentiles occurred in the winter with a maximum domain-averaged difference of 4.9 °C for the 5th percentile. For the 95th percentiles, the largest biases appeared in the summer with a maximum domain-averaged difference of –2.2 °C.

3.1.~~2~~3 Precipitation and wet-day frequency

Also multi-year mean seasonal precipitation sums were reliably simulated by HCLIM38-ALARO although slight overestimation was evident. Figure ~~5~~6 depicts both observed multi-year mean seasonal precipitation sums from E-OBS dataset over ~~the model domain~~Finland in 2002–2014 as well as the differences between HCLIM38-ALARO with a reference to E-OBS. Similarly ~~than in~~to the figures shown before ~~Fig. 3~~, the stippled areas represent significant differences confirmed by the Student's t-test ($p < 0.05$). Overall, precipitation was overestimated rather than underestimated throughout the year. The biases were the smallest in ~~autumn~~the winter ~~(September–October)~~ with a domain-averaged bias of ~~12.7~~16.1 % and highest in the spring ~~(March–May)~~ with a domain-averaged bias of ~~31.9~~42.2 %. The ~~highest~~ largest biases in simulated precipitation occurred in ~~the Norwegian mountains~~the north of Finland, especially over Lapland, where the biases were also statistically significant for every season. ~~as well as over the mountainous areas, which may penalize the model in the areas with the most complex topography. We stress that E-OBS might suffer from undercatch errors during the winter and spring.~~ The biases were statistically significant over the whole model domain during the spring and summer season. We stress that E-OBS might suffer from undercatch errors during the winter and spring. The larger biases in the northern parts of Finland might again originate from the sparser observation network in the northernmost domain. ~~During the winter and autumn seasons, the biases were significant mainly in the northern parts of the model domain (e.g. the northernmost Finland) and in Latvia in addition to Norway. Again, some part of the biases might have been caused by the lack of a dense observation network in the northernmost domain. Statistically significant differences during the spring season occurred almost in the whole Finland, the northern part of European Russia, northern Sweden, partly the Baltic countries, and Norway.~~ The results obtained for HCLIM38-ALARO showed similar magnitude and spatial patterns of the precipitation biases compared to other EURO-CORDEX RCMs that are mainly overestimating seasonal precipitation over Finland during the winter and summer as shown by Kotlarski et al. (2014).

The overall overestimation of spring and summertime precipitation in HCLIM38-ALARO might be due to too frequent low and moderate intensity precipitation events as Lindstead et al. (2015) and Lind et al. (2016) pointed out in their stud~~y~~ies of HCLIM36. Also the wet-day frequency with a 1 mm day^{-1} threshold was slightly overestimated especially during the spring and summer with the highest domain-averaged bias of 4.6 days season^{-1} (Fig. S6). Contrarily, HCLIM38-ALARO slightly underestimated wet-day frequency during the winter (excluding the most northern and southern parts of Finland) with the domain-averaged bias of –0.2 days season^{-1}. In addition, HCLIM38-ALARO slightly overestimated the relative frequency of daily precipitation over Finland for the intensities that were approximately between 10 and 40 mm day^{-1} in the spring season and 10 and 80 mm day^{-1} in the summer reason (Fig. S3). Otherwise, the PDFs of daily precipitation were adequately captured by HCLIM38-ALARO.

Figure ~~6~~7 further confirms that precipitation was ~~mainly~~ overestimated over different regions of Finland ~~although a slight underestimation was found in January in Western and Central, Eastern, and Northern Finland.~~throughout the year. The mean monthly biases between the regions did not substantially differ ~~substantially~~ from each other. However, the biases were the smallest in ~~the~~Northern Finland during the winter (December–March) and in the southern parts of Finland during ~~most of~~ the other months (April–November). ~~and, c~~Consistently, the largest biases were found in ~~the northern parts of Fin~~Lapland. As already seen in Fig. ~~5~~6, the largest biases appeared during the spring season (especially between April and May) and the second largest biases during the summer and early autumn season (from June to September).

3.1.4 Other variables

The modeled seasonal averages of total cloud fraction (*clt*), SW_d, LW_d, *RH*, and *WS* were compared against the ERA5 reanalysis product over 2002–2014 since these parameters were used as inputs for RoadSurf together with T_{air} and precipitation. Again, the stippled areas in Fig. 8 illustrate significant differences revealed by the Student's t-test ($p < 0.05$). *Clt* was significantly underestimated throughout the year with the highest domain-averaged bias of –16.1 % in the winter (Fig. 8a–d). Consequently, LW_d was significantly underestimated during the winter, summer (in the north), and autumn with the largest domain-averaged bias of –15 W m^{-2} occurring in the wintertime (Fig. 8e–h). SW_d was, in turn, mostly significantly overestimated, especially during the autumn when the domain-averaged bias was 10.3 W m^{-2} (Fig. 8i–l). The biases in SW_d during the winter were small as the received actual SW_d is, in general, limited during this time of the year at the high latitudes. However, negative biases in SW_d were found over the southern parts of Finland during the spring, although the differences were significant at only over restricted areas. These results are in agreement with the previous comparison of *clt*, LW_d, and SW_d between HCLIM36-ALARO and ERA-Interim reanalysis product over North Europe shown by Lindstedt et al. (2015).

In addition, *RH* was underestimated in the winter and autumn with a domain-averaged bias of –4.3 % during the winter and overestimated during the summer with a domain-averaged bias of 6.3 % (not shown). *WS* was mainly underestimated during all seasons with the largest domain-averaged negative bias of –0.6 m s^{-1} appearing in the winter and autumn seasons (not shown).

3.2 Evaluation of HCLIM-driven RoadSurf~~-HCLIM~~

3.2.1 Road surface temperature

The meteorological data from HCLIM38-ALARO was used as an input to RoadSurf ~~which~~that was further evaluated against 25 road weather stations in Finland. Here, we mostly concentrate on the evaluation of road surface temperature as it is the main output of RoadSurf. Only the results obtained for ~~the~~an extended winter season from October to April were explored because this period is the most relevant for road maintenance (e.g. salting of the roads and snow ploughing) and road safety in Finland. Road surface temperature produced by RoadSurf was evaluated against the observations by calculating the PDFs of observed and modeled daily T_{road} at the road weather stations as well as computing mean monthly biases~~, root-mean-square-errors (RMSE) as well as Pearson's correlation coefficients (*R*)~~ and mean absolute errors (MAE) using the average ~~daily~~monthly road surface temperature values. It is good to keep in mind that the ~~daily and~~ hourly and daily ~~time~~temporal resolutions are the most crucial for road weather because the accident rates might increase rapidly in case of a sudden change of the prevailing weather (Juga et al., 2012). The monthly time scale was selected for the evaluation to account for the fact that RoadSurf was driven using an RCM that was forced by a reanalysis product only in the lateral boundaries. This implies that the modeled day-to-day variability might not entirely match with observations at all locations. However, calculating monthly statistics ~~of the above-mentioned metrics using daily data~~ gives us a clear understanding of the model performance ~~during~~for different months during the study period from 2002 to 2014.

Figure ~~7~~9 makes evident that the HCLIM-driven RoadSurf was able to simulate the monthly means of T_{road} with ~~a~~ high accuracy and with most of the biases falling between ±2 ºC. The mean monthly bias at all 25 stations was –0.3 ºC (range –2.1–2.~~8~~7 ºC)~~,~~ and MAE ~~the average monthly RMSE 2.1~~0.9 ºC (range ~~1~~0.3–~~4.6~~2.9 ºC)~~ and the average monthly *R* 0.93 (range 0.8–1)~~. Some regional and seasonal differences were apparent. In January and February, most of the stations located in Southern, Western, and Central Finland had mainly negative mean biases whereas the biases were predominantly positive at the stations located in Eastern and Northern Finland, and Lapland. When looking at the results for all stations, most of the positive mean biases occurred in January and March~~ and October~~ whereas negative biases occurred in April, November, and December. Eleven stations had negative mean biases throughout all the analyzed months while the rest of the stations had both negative and positive mean biases depending on the month. ~~The RMSE values were the lowest in March, October, and November and the highest in January, February, April, and December.~~ Overall, the MAE values were the lowest in March and October while ~~T~~the highest ~~RMSE~~MAE values occurred in Lapland in January and February.~~where the correlations were also worse compared to the stations located in the south. Interestingly, the lowest correlations occurred in April at almost every station. The statistical significance of the differences between the stations is discussed in more detail in Sect. 3.2.2.~~ Despite the apparent mean monthly biases, the shapes of the daily T_{road} PDFs were sufficiently reproduced by RoadSurf with the largest deviations found in the winter (Fig. S4) in accordance with the PDFs of daily T_{air}.

~~Two p~~Probable reasons for the seasonal and regional differences in the model performance are ~~(1)~~ the biases in the HCLIM38-ALARO data~~ (mainly T_{air} and precipitation),~~ and~~ (2)~~ the fact that RoadSurf works well in the vicinity of 0 degrees. ~~For example~~To address the impact of the biases in the input parameters on the T_{road} biases, we computed the monthly mean biases in the HCLIM38-ALARO model outputs with a reference to E-OBS (in case of T_{air} and precipitation) and ERA5 (in case of LW_d, SW_d, RH, and WS) at the grid cell closest to the road weather station in question. The monthly biases in the input parameters were plotted against the monthly biases in T_{road}. The analysis shown in Fig. 10 revealed that the variability of the monthly biases in T_{air} explained on average 57 % (range 19–84 % in October–April) of the variability of the monthly biases in T_{road} while the LW_d biases explained on average 16 % (range 2–34 % in October–March). Furthermore, the variability in SW_d biases was found to explain a small amount (4 %) of the variability in T_{road} biases during April. The comparison between other input parameters and T_{road} did not reveal clear linear responses and are thus not discussed here. Also Karsisto et al. (2017) noted that a part of the T_{road} biases is caused by the biases in the input parameters used to force road weather models. In their study, the input was provided by a forecast produced with a high-resolution NWP version of HARMONIE (cy36h1.4) with a grid resolution of 2.5 km over the Netherlands. In that study, the KNMI road weather model (a 1D heat balance model similar to RoadSurf) was run by removing the bias of one of the model inputs, 2 m T_{air}. This reduced the T_{road} bias during the nighttime by 50 % indicating that the biases in the input parameters clearly affect road weather model outcomes.

Moreover, the comparison of the simulated and observed T_{air} in the wintertime (December–February) revealed a warm bias ranging from 0.~~2~~1 to 1.1 ºC in the northern parts of Finland (Northern Finland and Lapland) while Southern Finland had negative biases ranging between –0.~~5~~4 and –0.~~1~~04 ºC (see Fig. 4). Thus, the larger and more positive biases in the simulated T_{air} in Northern Finland and Lapland compared to Southern Finland ~~could~~seem to explain the larger positive biases in the modeled T_{road} at the northernmost stations. ~~On the other hand, the errors in the precipitation input might have caused the higher RMSE values and lower correlations in April compared to the other months: The biases in the HCLIM38-ALARO simulated precipitation were the highest in April.~~ In addition, Kangas et al. (2015) noted that RoadSurf is designed to work especially well when temperatures are close to zero. Based on the monthly statistics obtained for the study period (2002–2014), road surface temperatures were crossing zero degrees particularly often during March, April, and October (see Sect. 3.2.~~3~~2). This good model performance near 0 ºC could, in turn, partly explain why the ~~RMSE~~ MAE values were ~~actually~~ lower in October and March compared to other months. ~~December at all stations in 2013 and also in almost every simulated year (not shown) as opposed to the findings by Karsisto et al. (2016). In their study, RMSE values of the simulated T_{road} were larger in October 2013 compared to December 2013. They stated that this might be due to difficulties in simulating the highest and lowest T_{road} because the estimation of T_{road} is very sensitive to the total radiation values. Unfortunately, the road weather stations included in our study do not observe radiation or cloudiness; Therefore, the inaccuracy in the simulated radiation could not be evaluated here.~~

Some part of the biases in T_{road} might originate from the RoadSurf model itself. For instance, the absence of snow removal and salting in the model might keep the road surface colder than what it would be with the maintenance actions. In addition, traffic is assumed to pack some part of the snow into ice while the remaining part is assumed to be blown away from the road. For example, the real traffic amounts are higher in Southern Finland compared to the other parts of the country, which can lead to an overestimation of the simulated icy and snowy conditions in the south and, hence, to colder road surface conditions than what is observed. On the other hand, the snowpack that is observed might actually stay longer than what is simulated by the model leading to positive biases in T_{road} at locations with less traffic: this could especially happen at stations such as 23 (Sieppijärvi). The biases in T_{road} might also stem from the absence of shading effects as this effect is not taken account by RoadSurf.

Although the results obtained in this study indicated a good skill of RoadSurf to realistically capture T_{road}, the mean biases ~~and RMSE values~~ were slightly larger compared to the previous studies of RoadSurf. For example, Karsisto et al. (2016) found that the biases in the simulated T_{road} varied between –1 and 1 ºC (mostly ±2 ºC in our study) ~~and RMSE values between 0.3 and 1.9 ºC~~ at 20 stations in Finland during October and December 2013 when RoadSurf was driven by a high-resolution NWP version of HARMONIE (cy36h1.4) with a grid resolution of 2.5 km without any data assimilation. However, it is good to note that the results obtained in our study and by Karsisto et al. (2016) are not directly comparable since in their study RoadSurf was initialized using road weather station measurements for 48 hours and only the first forecasted hour was analyzed. ~~In their study, the input forecast was produced by a high-resolution NWP version of HARMONIE (cy36h1.4) with a grid resolution of 2.5 km. Thus~~However, one possible reason for the slightly larger errors obtained in the present study might be the coarser grid resolution of HCLIM38-ALARO as compared to the NWP version: Coarser grid resolution implies that not all the local features, such as ~~elevation~~topography, are described as in detail as they are in higher resolution NWP models. Increasing the grid resolution of HCLIM38-ALARO might therefore yield better performance of RoadSurf although increasing the grid resolution of a climate model will also increase the computational cost. However, the longer time period used in this study makes the results more robust compared to the previous studies in which only short time periods were analyzed.

~~3.2.2 The role of station characteristics on the simulated road surface temperature~~

~~As mentioned earlier, the performance of RoadSurf to simulate T_{road} differed between the studied regions of Finland. Thus, a nonparametric Kruskal–Wallis test with an alpha of 0.01 was used to investigate the statistical significance of the differences in the monthly mean biases, RMSE values, and correlation coefficients of the stations and whether these differences stemmed from the station's different characteristics. The Kruskal–Wallis test can be performed to determine if all groups of a certain dataset are identical or if at least one group is differing from the rest (Helsel and Hirsch, 2002). Therefore, the stations were divided into different groups based on the region, surrounding characteristics, and road maintenance class. Before using the Kruskal–Wallis test, the normality of the data was tested using an Anderson–Darling normality test and the equality of variances using a Levene's test, both tests with an alpha of 0.05. One-way ANOVA could not be used as the biases, RMSE values, and the correlation coefficients were not normally distributed among all the tested groups. Furthermore, not all the variances of the tested groups were homoscedastic. More specifically, the variances between the groups were not equal except for the groups formed from correlation coefficients. Finally, a Dunn–Sidak method was used as a post hoc test to further distinguish which groups were statistically different from each other.~~

~~The regions were defined as Southern Finland, Western and Central Finland, Eastern Finland, Northern Finland, and Lapland as in the other parts of this study. The stations were also divided into open, partly obscured (a few trees nearby or trees on the other side of the road), and obscured (forest on both sides of the road) groups based on the surrounding characteristics (see Table 1). The road maintenance class divided the stations into four groups, 1–4, where class 1 represents high maintenance and class 4 low maintenance (See Appendix A for more detailed explanations of maintenance classes). To determine the statistical significance ($p < 0.05$) of the Kruskal–Wallis test, the differences in the mean ranks of datasets at all analyzed groups were defined using a:~~

~~Null hypothesis (H0): The mean ranks of *k groups are identical, with k =3–5.*~~

~~Alternate hypothesis (H1): At least one mean rank differs from the others.~~

~~Based on the Kruskal–Wallis analysis, the biases were statistically different for the stations located in different regions and for the stations having different maintenance classes (see Table S1 in Supplementary material for *p* values). In particular, the biases were significantly more negative for the stations located in Southern Finland and for the stations having the highest maintenance class. This could be due to the cold bias in the input T_{air} but also due to the lack of snow removal and salting in the model, which might keep the road surface colder than what it would be with the maintenance actions. In addition, traffic is assumed to pack some part of the snow into ice while the remaining part is assumed to be blown away from the road. In Southern Finland, the real traffic amounts are higher than in the other parts of the country, which can also lead to an overestimation of the simulated icy and snowy conditions in the south and, further, to colder road surface conditions than what is observed. However, the surrounding characteristics of the stations did not affect the biases. Also Karsisto et al. (2016) concluded that there were no considerable differences in the biases in simulated T_{road} between the stations having different surrounding characteristics (open, slightly obscured, and obscured).~~

~~The Kruskal–Wallis analysis of the RMSE and *R* values revealed also significant differences between the stations located in different regions and between the stations having different maintenance groups. The RMSE values were significantly better at the stations located in Southern Finland compared to the stations located in Lapland. Similarly, the *R* values were significantly greater for the stations located in southern parts of Finland (Southern and Eastern Finland) compared to the stations located in the northern parts of Finland (Northern Finland and Lapland). The highest RMSE values and the lowest *R* values in northernmost Finland may be explained by the already mentioned warm bias in the input T_{air} over that region during the winter. In addition, significantly smaller RMSE and greater *R* values were obtained for stations having moderate maintenance (class 2) compared to the stations with low maintenance level (class 4). It could be expected that the stations with the lowest maintenance level would have the lowest errors as the maintenance is not taken into account in RoadSurf. But as mentioned before, traffic packs some part of the snow into ice in the model. In real life, the snowpack might actually stay longer than what is simulated by the model: this could especially happen at the stations having low traffic amounts, which is the case for the stations 22 (Saariselkä) and 23 (Sieppijärvi). Low maintenance stations (class 4) did not have the lowest RMSE or the highest *R* values, most likely due to these too fast depleting snowpacks in the model and the biases in input T_{air}. The high maintenance stations (class 1) did not have the smallest RMSE values either, most probably due to the~~

~~negative biases in the simulated T_{air} and T_{road}. As it was the case for the biases, the surrounding characteristics did not have a significant effect on the RMSE values~~. ~~This partly contradicts the results obtained by Karsisto et al. (2016) who found some differences in the RMSE values in October 2013 between the different stations with different surrounding characteristics. In that study, the largest RMSE values were obtained at stations where the Sun was the most obscured. This was hypothesized to be due to the uncertainty in the SW radiation input, which was produced by the NWP model. In the present study, the R values were still significantly lower for the obscured stations compared to the slightly obscured ones, which is in agreement with the findings by Karsisto et al. (2016).~~

3.2.~~3~~2 Zero crossing days

Temperatures close to 0 ºC should be predicted correctly because in these conditions wet road surfaces have a tendency to freeze (e.g. Vajda et al., 2014) and roads are the most slippery in the copresence of ice (Moore, 1975). In this study, a zero crossing day was defined as a day when the road surface temperature had been at least once both below –0.5 ºC and above 0.5 ºC.

Figure ~~8~~11 shows that the monthly amount of zero crossing days and the monthly variation (standard deviation) were well captured ~~well~~ by RoadSurf. This was expected as RoadSurf has been confirmed to simulate T_{road} accurately in the vicinity of zero degrees (Kangas et al., 2015; Karsisto et al., 2016). On average, the correlation coefficient was very high (0.92) and the mean bias was approximately 0.9 days (Fig. ~~8~~11f). The performance of the model differed slightly depending on the analyzed region. Surprisingly, the correlation coefficient was the lowest in Southern Finland and the highest in Northern Finland and Lapland whereas the bias was the lowest in Eastern Finland and the highest in Lapland. The higher biases in Lapland might be explained by the overall overestimation of zero crossing days, which might, in turn, be caused by the warm bias in the simulated T_{road} values as discussed before. Overall, most of the zero crossing days occurred in March, April, and October. However, the number of zero crossing days declined in March and increased in April when moving towards the North. In Lapland, most of the zero crossings occurred in April instead of March. This was also expected as the winter season (and therefore the coldest period) lasts longer in Lapland compared to the southern parts of Finland, leading to less zero crossing days in March. The smallest number of zero crossings took place in January, February, and December. These are usually the coldest months of the year, especially in Lapland (see also Table 1); Thus, 0 ºC is not crossed as often during these months.

3.2.~~4~~3 Road surface classes

The majority of the wintertime and weather-related road accidents in Fenno-Scandia are caused by the snowy and icy road conditions in addition to, for example, the driving habits and worn out tires (Salli et al., 2008). To investigate RoadSurf's skill to correctly predict the road surface classes (e.g. snowy and icy surfaces) ~~correctly~~, the model results and observations were compared by calculating the ~~mean daily~~ fraction of each road surface class occurr~~ed~~ing within a month. The fraction was calculated as a multi-year sum of the occurrence of the surface class in question divided by the multi-year sum of the occurrence of all surface classes and then taking an average between stations falling into the same region.~~-~~ It is good to remember that the observed and modeled road surface classes might not fully match as they are defined differently.

Figure ~~9~~12 shows that overall RoadSurf captured well the prevailing road surface conditions although the observed and modeled fractions differed slightly. For example, the model overestimated the fraction of dry surfaces in all regions (average bias over all regions and all months was ~~3.3~~7 ~~hours~~% as a fraction) and underestimated damp surfaces slightly more (average bias –~~4.2~~16 ~~hours~~%). The model underestimated also wet surfaces (average bias –~~2.3~~6 ~~hours~~%), but the ~~hours accumulated in~~fraction of ~~-~~the partly icy class (~~2.7~~8 ~~hours~~% on average) w~~ere~~as almost equal to this difference between the modeled and observed wet surface fraction. Therefore, these results indicated that wet surfaces tended to be predicted as partly icy, although it has to be remembered that observations do not have a partly icy class. The underestimation of the frost on the road (average bias –~~0.5~~1 ~~hours~~%) and overestimation of ice (~~0.4~~2 ~~hours~~%) were also of ~~the~~a similar magnitude with opposite signs. Moreover, the snow class was slightly overestimated with an average bias of ~~0.6~~2 ~~hours~~%. These results are in line with the study by Kangas et al. (2015) where they encountered an overestimation of ice and snow storages produced by RoadSurf at two stations located in Finland. In addition, they found that sometimes frost predicted by the model was observed as ice in the measurements. In the present study, frosty surfaces were, however, mainly underestimated. On the other hand, both icy and frosty surfaces are slippery, so in that aspect the model behavior (i.e., the tendency of the model to underestimate frost and to overestimate ice with the same magnitude~~ than ice is overestimated~~) is acceptable.

The ~~lack~~absence of road maintenance could be one logical reason why the model overestimate~~s~~d icy and snowy surfaces: In real life, salting prevents roads to become icy and snow is removed from the roads. Accordingly, the observed and modeled fractions of snowy surfaces were very similar to each other in Lapland where ~~much less~~ maintenance, such as snow ploughing, is performed far less frequently compared to the more southern parts of Finland in real life. The icy road fraction was underestimated in Lapland whereas this fraction was overestimated in the other regions: In reality~~ life~~, salting is not performed as often at the stations in Lapland as in Southern Finland and thus icy roads can occur more ~~often~~frequently in the northmost stations. Furthermore, the RoadSurf model takes the effect of traffic into account~~ the effect of traffic~~ in a similar manner regardless of the region. Therefore, the simulated ice and snow storages might deplete too fast in the model considering the substantially lower traffic amounts in the northern parts of Finland compared to the south. For instance, snow storage was slightly underestimated in Lapland although only in January and November (Fig. 12e). The warm bias in Lapland might also have played a role in the underestimation of icy road fraction as icy roads are less likely to occur if the simulated air temperatures are too high. In addition, the underestimated wet and damp surfaces during the winter months (December–February) might be explained by the slightly underestimated wet-day frequency of precipitation over most parts of Finland (see Fig. S6).

3.2.~~5~~4 Categorical performance of the simulated frequency of water, snow, and ice storages

Rainfall has been considered as one of the main contributing factors in traffic accidents together with snow and ice on the road (e.g. Andersson and Chapman, 2011b). Therefore, the water, snow and ice storages, as well as their frequency, should be simulated accurately. The absolute values of the storages are not discussed here as the modeled values represent areal averages and observations point measurements. In addition, the optical sensor might not sense correctly the exact thickness of the water, snow, or ice layer on the road, but rather it might detect only the upper layer of these storage terms. Thus, RoadSurf's ability to simulate the frequency of the storages was assessed by first calculating the daily ~~mean~~maximum values of the storages between October and April and, further, setting the daily values to one if the daily m~~ean~~aximum value was more than zero and to zero if the daily m~~ean~~aximum value was zero. These binary values were used to calculate hits and false alarms (Table ~~2~~S1) and the probability of detection (POD~~; Eq. (1)~~) and false alarm ratios (FAR~~; Eq. (2)~~) (Roebber, 2009). The details of the POD-FAR analysis are explained in Supplement S1. The number of compared daily cases per station varied between 503 and 1101 days depending on the data availability at each station. However, this method might penalize the model more than it should because the modeled storages were compared with observations using day-to-day values~~might be slightly displaced or mistimed~~. For this reason, we additionally calculated the multi-year sums of all the modeled and observed daily cases with daily maximum more than zero or zero.~~the results should be interpreted with care and should be taken as qualitative.~~

The results of the POD-FAR analysis for 11 stations including an optical sensor (see Table 1) are illustrated in Fig. 1~~0~~3 using a categorical performance diagram (Roebber, 2009; please see Supplement S1 for more details).~~ The POD describes the proportion of the times when the event occurred and was also captured by the model. In contrast, the FAR defines the number of false alarms divided by the number of all cases when the event is modeled. This implies that the performance of the model is the better the closer the POD is to 1 and FAR to 0. Therefore, the best values can be found in the upper-right corner of the diagram as the y-axis shows the POD values and the x-axis the success ratio which means the FAR values in the reversed order (1–FAR). The dotted lines show the frequency bias (Eq. (3)) which indicates overestimation (underestimation) if the values are higher (lower) than 1. The continuous lines represent the critical success index (CSI; Eq. (4)) which in turn represents the hits in relation to the number of cases when the event was either observed or modeled. Ideally, the CSI values should be close to 1. Bootstrapping with 1000 resamples was used to calculate the 95 % confidence intervals for the POD and FAR values in Fig. 10.~~ Figure 1~~0~~3 shows that RoadSurf reliably captured the occurrence of the storage terms as the points located near the upper-right corner of the diagram. However, the model performance varied slightly depending on which storage was simulated. For instance, the modeled water storages had the lowest FAR (highest 1–FAR) values but also the lowest POD values. This means that because the model did not detect water as often as it should, also the false alarm ratio was smaller. The frequency bias values were lower than one indicating an underestimation of the events with water on the surface. The opposite was true for the modeled ice storages: The events were predicted well (POD was high), but false alarms were more frequent (1–FAR was lower). Furthermore, the frequency bias values were greater than one suggesting an overestimation of the events with ice on the road. The POD and FAR values of the modeled snow storages fell somewhere in between the POD and FAR values which were obtained for the water and ice storages. The model underestimated the frequency of the events with snow on the road but to a lesser extent compared to the underestimated frequency of the water storages.

It has to be emphasized once more that the model does not take into account road maintenance measures. Again, the ~~lack~~absence of salting can be one reason for the overestimated occurrence of ice and the underestimated occurrence of water on the road surface. However, the model is thus on the 'safe side', which means that in the operational use the model would give warnings to the road users slightly more often than what would be required. As mentioned before, a part of the underestimated frequency of water might be explained by the slightly underestimated wet-day frequency of precipitation during the winter season. ~~Another interesting fact is that~~On the other hand, the ~~lack~~absence of snow removal in the model did not lead to an overestimated frequency of snow on the road: this frequency was underestimated while the ~~daily~~ fraction of snowy road cover was overestimated as shown in Fig. 12. One possible reason for this discrepancy might be the different amount of road weather stations used in the POD and FAR analysis compared to the road condition analysis (11 vs. 25 stations). Another reason might be that the POD and FAR analysis utilized fewer observations compared to the number of observations used in the analysis of the road surface conditions (due to~~more~~ a higher amount of missing data). ~~In addition, the daily values were given more weight in the POD-FAR analysis compared to the analysis of the road surface classes because the daily fractions of snowy road surface classes represent an average situation within a month.~~ Moreover, the RoadSurf-HCLIM configuration might not capture all the snow events which are observed at the station because the simulated storages represent areal averages. ~~As the majority of the stations having the optical sensor are located in the southern parts of Finland, too fast depleting snowpacks in the model might, however, not be the cause for this underestimation as it could be the case for stations locating more north.~~However, the underestimated frequency of snow cannot be explained by the snowpacks that are depleting too fast in the model. This is because the majority of the stations with an optical sensor utilized in this study are located in the southern parts of Finland where the modeled snowpacks might actually stay longer compared to the measurements as discussed before.

In addition to the POD-FAR analysis, we computed the modeled and observed fractions of the multi-year sums of the daily cases with the daily maximum storage of water, snow, or ice more than zero or zero. The results are shown in Fig. S7 as fractions over all 11 stations. This figure supports the main conclusions from the POD–FAR analysis: The occurrence of water and snow storages were underestimated by the model by a fraction of –18 % and –7 %, respectively. The frequency of ice storage was slightly overestimated by a fraction of 5 %.

4 Conclusions

This study described the performance of the HCLIM38-ALARO regional climate model over F~~enno-Scandia~~inland and, further, evaluated the skill of the HCLIM38-ALARO-driven road weather model RoadSurf to reproduce the present-day road weather conditions in Finland. The HCLIM38-ALARO was forced with the reanalysis product ERA-Interim in the lateral boundaries. This study showed that HCLIM38-ALARO is in good agreement with the gridded daily mean air temperature and precipitation observations: The model reliably reproduced the ~~seasonal and~~ monthly and seasonal temporal and spatial patterns as well as daily variability of these variables over ~~Fenno-Scandia and~~ Finland. Especially daily mean near-surface air temperatures were well represented by HCLIM38-ALARO. On the other hand, daily minimum air temperatures were slightly overestimated and daily maximum temperatures underestimated.~~the p~~ Precipitation was ~~slightly~~ overestimated during all seasons, although some of this overestimation might be caused by the inaccuracy of E-OBS data due to possible undercatch errors and lower station density in the northern parts of ~~the modeled domain~~Finland. Overall, the HCLIM38-ALARO results were found to be in line with other EURO-CORDEX RCMs. The underestimated total cloud fraction in the model led to the overestimated downwelling shortwave and underestimated longwave radiation, which has also been encountered in the previous evaluations of HCLIM over North Europe.

As far as the authors are aware, this may be the first paper that studies the performance of a road weather model which is forced by RCM data. This study revealed that the HCLIM38-ALARO-driven RoadSurf was able to adequately reproduce the daily distributions of road surface temperatures (T_{road}) and accurately simulate ~~road surface temperatures (~~T_{road}~~)~~ with the mean monthly bias of –0.3 ºC~~,~~ and ~~RMSE~~the mean monthly MAE of ~~2.1~~0.9 ºC~~ and Pearson's R of 0.93~~ over Finland. These metrics indicated a slightly poorer performance than what was obtained in the earlier studies of RoadSurf. ~~However, t~~The coarser grid resolution of the HCLIM38-ALARO compared to the NWP model input used in the earlier studies might be the main reason for this outcome as no data assimilation was used for HCLIM38-ALARO or the NWP model. Moreover, the HCLIM38-ALARO simulated air temperature tended to have a warm bias over the northern parts of Finland in the winter. This, in turn, might be the major reason for ~~the significantly~~ the better performance of RoadSurf to simulate T_{road} at the stations located in ~~Southern~~ the southern parts of Finland compared to the stations located in Lapland~~, also confirmed by the Kruskal–Wallis test~~. The variability of the air temperature biases was found to explain the largest part of the variance in the road surface temperature biases as compared to other input variables of RoadSurf.

In addition, RoadSurf adequately captured ~~well~~ the daily zero crossings, which verified the good performance of the model when temperatures approach zero degrees. This is of ~~a~~ great importance as the road surfaces are ~~the~~ most prone to slippery conditions when the road surface temperatures are close to 0 ºC and simultaneous icing occurs. Moreover, the analysis on the road surface classes showed that the model is overall in a good agreement with the observations in terms of the prevailing road conditions. However, the model tended to yield more icy and snowy road surfaces than ~~what~~ indicated by ~~the~~ observations~~ showed~~. The ~~lack~~absence of road maintenance, such as salting and snow ploughing, is very likely the dominant reason for this model behavior as well as for the overestimated occurrence of ice and underestimated occurrence of water on the road surface. On the other hand, the overestimated traffic wear in the model and therefore too fast depletion of ice storages could be the reason for the underestimated fraction of icy surfaces at the northernmost stations.

These results were obtained using a limited set of road weather stations in Finland. On the other hand, the 13-year long study period makes the results more robust compared to the earlier studies of RoadSurf which have concentrated only on short verification periods of 1 week to some months. Therefore, the results represented in this study indicated that HCLIM38-ALARO realistically captured the ~~Fenno-Scandian~~ climate over Finland and that this RCM data can be used as an input to RoadSurf in order to produce reliable results of T_{road}, road surface classes, and storage terms. Although RoadSurf represents a 'what-if-nothing-is-done' scenario in terms of road maintenance, it also makes the model ideal to study the relative changes in the road surface conditions due to climate change. Earlier studies of the climate change impacts on road weather have mainly considered the relative changes in air temperature and precipitation. Therefore, the approach presented in this study offers an alternative to these methods: Running the road weather model with HCLIM38-ALARO produced climate projections makes it possible to directly study how the road weather conditions are going to change in the future.

5 Code availability

The ALADIN and HIRLAM consortia cooperate on the development of a shared system of model codes. The HCLIM model configuration forms part of this shared ALADIN-HIRLAM system. According to the ALADIN-HIRLAM collaboration agreement, all members of the ALADIN and HIRLAM consortia are allowed to license the shared ALADIN-HIRLAM codes within their home country for non-commercial research. Access to the HCLIM codes can be obtained by contacting one of the member institutes of the HIRLAM consortium (see links on http://www.hirlam.org/index.php/hirlam-programme-53). The access will be subject to signing a standardized ALADIN-HIRLAM license agreement (http://www.hirlam.org/index.php/hirlam-programme-53/access-to-the-models). The RoadSurf code is not ~~publicly available~~in the public domain and cannot be distributed.

6 Data availability

Due to the very large size of the data files, the data are not publicly available. The data files can be requested from the first author.

Appendix A: The maintenance classes of the roads during wintertime in Finland (Finnish Transport Agency, 2018)

Maintenance class 1 (Ise):

The road is kept bare most of the time. The slipperiness of the roads is prevented beforehand, but mild slipperiness might occur in case of a rapid change in the prevailing weather. Salting is not possible during long-lasting cold periods, which can lead to partially frozen road surfaces. The maintenance is timed so that the harm for the traffic is minimized.

Maintenance class 2 (Is):

The road is kept bare most of the time. The aim is to prevent slipperiness beforehand, but mild slipperiness might occur in case of a rapid change in the prevailing weather. The central and northern parts of Finland, and also the southern part of the country (only during the coldest periods) might have a thin ridge of snow packed on the road, which does not particularly affect driving. Salting is not possible during long-lasting cold periods, which can lead to partially frozen road surfaces.

Maintenance class 3 (Ib):

The road is kept bare most of the time. The aim is to prevent slipperiness beforehand, but mild slipperiness might occur in case of a rapid change in the prevailing weather. During the coldest periods, there might be shallow and narrow ridges of snow packed on the road. Salting is not possible during long-lasting cold periods, which can lead to partially frozen road surfaces.

Maintenance class 4 (l):

The road is maintained at a fairly high standard but mostly without salt. The surface of the road is partially bare depending on the traffic volume and weather. There might be ridges of snow packed on the road and the road might also be fully covered with a snowpack. The road is kept safe enough for the road users. The possible snowpack on the road surface is smoothed. Slipperiness is prevented beforehand only in the autumn and spring and in case of particularly hazardous situations.

Author contribution. ET performed the HCLIM38-ALARO simulations with the help of JPP. JPP did the offline coupling of RoadSurf and HCLIM38-ALARO. ET planned and performed the analysis of HCLIM38-ALARO and HCLIM-driven RoadSurf with the help provided by JPP, MH, HK, and AL. MK and MH assisted with the road weather model RoadSurf and MH with the road weather observations. JPP, HK, and AL initiated the work. ET wrote the paper. All co-authors participated in the paper-writing phase and gave valuable comments regarding the first versions of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This project has partly been funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under a grant agreement No 646857, and by the Academy of Finland project 287440. The leading author would like to thank Maj and Tor Nessling foundation for the financial support of regional climate modeling research. We also acknowledge the E-OBS dataset from the EU-FP6 project ~~ENSEMBLES (http://ensembles-eu.metoffice.com)~~UERRA (http://www.uerra.eu),~~ and ~~the data providers in the ECA&D project (**http://www.ecad.eu**) as well as the ERA5 reanalysis product from ECMWF. We are grateful for the help provided by Ari Aaltonen in retrieving the Finnish road weather observations. In addition, we would like to thank the two anonymous reviewers for their comments and efforts towards improving our manuscript.

References

Andersson, A. K., and Chapman, L.: The impact of climate change on winter road maintenance and traffic accidents in West Midlands, UK, Accident. Anal. Prev., 43, 284–289, https://doi.org/10.1016/j.aap.2010.08.025, 2011a.

Andersson, A. K., and Chapman, L.: The use of a temporal analogue to predict future traffic accidents and winter road conditions in Sweden, Meteorol. Appl., 18, 125–136, https://doi.org/10.1002/met.186, 2011b.

Bénard, P., Vivoda, J., Mašek, J., Smolíková, P., Yessad, K., Smith, C., Brožková, R., and Geleyn, J.: Dynamical kernel of the Aladin–NH spectral limited area model: Revised formulation and sensitivity experiments, Q. J. Roy. Meteor. Soc., 136, 155–169, https://doi.org/10.1002/qj.522, 2010.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K., Lenderink, G., Niemelä, S., Nielsen, K. P., Onvlee, J., Rontu, L., Samuelsson, P., Muñoz, D. S., Subias, A., Tijm, S., Toll, V., Yang, X., and Køltzow, M. Ø.: The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System, Mon. Weather Rev., 145, 1919–1935, https://doi.org/10.1175/MWR-D-16-0417.1, 2017.

Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service Climate Data Store (CDS), https://cds.climate.copernicus.eu/cdsapp#!/home, 2017.

Cornes, R., van der Schrier, G., van den Besselaar, E. J. M., and Jones P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Datasets, J. Geophys. Res. Atmos., 123, https://doi.org/10.1029/2017JD028200, 2018.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge Sanz, B. M., Morcrette, J., Park, B., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J., and Vitart, F.: The ERA Interim reanalysis: configuration and performance of the data assimilation system. Q. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Gerard, L.: An integrated package for subgrid convection, clouds and precipitation compatible with meso gamma scales, Q. J. Roy. Meteor. Soc., 133, 711–730, https://doi.org/10.1002/qj.58, 2007.

Gerard, L., Piriou, J., Brožková, R., Geleyn, J., and Banciu, D.: Cloud and Precipitation Parameterization in a Meso-Gamma-Scale Operational Weather Prediction Model, Mon. Weather Rev., 137, 3960–3977, https://doi.org/10.1175/2009MWR2750.1,2009.

Goodison, B. E., Louie, P. Y. T., and Yang, D.: The WMO solid precipitation measurement intercomparison – final report, World Meteorological Organization Tech. Doc. WMO TD-872, Geneva, Switzerland, Instruments and Observing Methods Report No. 67, 212 pp., 1998.

Hambly, D., Andrey, J., Mills, B., and Fletcher, C.: Projected implications of climate change for road safety in Greater Vancouver, Canada, Climatic Change, 116, 613–629, https://doi.org/10.1007/s10584-012-0499-0, 2013.

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, D20119, https://doi.org/10.1029/2008JD010201, 2008.

~~Helsel, D. R., and Hirsch R. M.: Statistical Methods in Water Resources, in: Techniques of Water - Resources Investigations of the United States Geological Survey, vol. 4, Hydrologic Analysis and Interpretation, U.S. Geological Survey, Reston, VA, chap. A3, 1–510, 2002.~~ Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A. J., Haiden, T., Hólm, E., Haimberger, L., Hogan, R., Horányi, A., Janiskova, M., Laloyaux, P., Lopez, P., Muñoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J.-N., Vitart, F., Yang, X., Zsótér, E., and Zuo, H.: Operational global reanalysis: progress, future directions and synergies with NWP, ECMWF, Reading, UK, ERA ReportSeries 27, 63 pp., 2018.

Hori, Y., Cheng, V. Y. S., Gough, W. A., Jien, J. Y., and Tsuji, L. J. S.: Implications of projected climate change on winter road systems in Ontario's Far North, Canada, Climatic Change, 148, 109–122, https://doi.org/10.1007/s10584-018-2178-2, 2018.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Króner, N., Kotlarski, S., Kriegsmann, A., Martin, E., Van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J. F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Reg. Environ. Chang., 14, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.

Jaroszweski, D., Hooper, E., and Chapman, L.: The impact of climate change on urban transport resilience in a changing world, Prog. Phys. Geog., 38, 448–463, https://doi.org/10.1177/0309133314538741, 2014.

Juga, I., Hippi, M., Moisseev, D., and Saltikoff, E.: Analysis of weather factors responsible for the traffic 'Black Day' in Helsinki, Finland, on 17 March 2005, Meteorol. Appl., 19, 1-9, https://doi.org/10.1002/met.238, 2012.

Juga, I., Nurmi, P., and Hippi, M.: Statistical modelling of wintertime road surface friction, Meteorol. Appl., 20, 318–329, https://doi.org/10.1002/met.1285, 2013.

Kangas, M., Heikinheimo, M., and Hippi, M.: RoadSurf: a modelling system for predicting road weather and road surface conditions, Meteorol. Appl., 22, 544–553, https://doi.org/10.1002/met.1486, 2015.

Karsisto, V., Nurmi, P., Kangas, M., Hippi, M., Fortelius, C., Niemelä, S., and Järvinen, H.: Improving road weather model forecasts by adjusting the radiation input, Meteorol. Appl., 23, 503–513, https://doi.org/10.1002/met.1574, 2016.

Karsisto, V., Tijm, S., and Nurmi, P.: Comparing the Performance of Two Road Weather Models in the Netherlands, Weather Forecast., 32, 991–1006, https://doi.org/10.1175/WAF-D-16-0158.1, 2017.

Koetse, M. J., and Rietveld, P.: The impact of climate change and weather on transport: An overview of empirical findings, Transport Res. D.-Tr. E., 14, 205–221, https://doi.org/10.1016/j.trd.2008.12.004, 2009.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297-1333, https://doi.org/10.5194/gmd-7-1297-2014, 2014.

Liikennevirasto (Finnish Transport Agency): Teiden talvihoito (The maintenance of the roads during winter), available at: ~~https://www.liikennevirasto.fi/tieverkko/kunnossapito/talvihoito#.Wyt4LRwlGCg~~https://vayla.fi/tieverkko/talvihoito#.XQOBDiZRWV4, last access: 1~~8~~4 ~~December~~June 201~~8~~9.

Lind, P., Lindstedt, D., Kjellström, E., and Jones, C.: Spatial and Temporal Characteristics of Summer Precipitation over Central Europe in a Suite of High-Resolution Climate Models, J. Climate, 29, 3501–3518, https://doi.org/10.1175/jcli-d-15-0463.1, 2016.

Lindstedt, D., Lind, P., Kjellström, E., and Jones, C.: A new regional climate model operating at the meso-gamma scale: performance over Europe, Tellus A, 67, 24138, https://doi.org/10.3402/tellusa.v67.24138, 2015.

Makkonen, L., Ylhäisi, J., Törnqvist, J., Dawson, A., and Räisänen, J.: Climate change projections for variables affecting road networks in Europe, Transport Plan. Techn., 37, 678–694, https://doi.org/10.1080/03081060.2014.959352, 2014.

Malin, F., Norros, I., and Innamaa, S.: Accident risk of road and weather conditions on different road types, Accident Anal. Prev., 122, 181–188, https://doi.org/10.1016/j.aap.2018.10.014, 2019.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.- L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, 2013.

Mikkonen, S., Laine, M., Mäkelä, H., Gregow, H., Tuomenvirta, H., Lahtinen, M., and Laaksonen, A.: Trends in the average temperature in Finland, 1847–2013, Stoch. Env. Res. Risk. A., 29, 1521–1529, https://doi.org/10.1007/s00477-014-0992-2, 2015.

Mironov, D. V.: Parameterization of lakes in numerical weather prediction. Description of a lake model.~~, COSMO Tech. Rep. 11~~, Deutscher Wetterdienst, Offenbach am Main, Germany, COSMO Tech. Rep. 11, 41 pp., 2008.

Mironov, D., Heise, E., Kourzeneva, E., Ritter, B., Schneider, N., and Terzhevik, A.: Implementation of the lake parameterisation scheme FLake into the numerical weather prediction model COSMO, Boreal Environ. Res., 15, 218–230, 2010.

Moore, D. F.: The Friction of Pneumatic Tyres, Elsevier Scientific Publishing Company, Amsterdam, Netherlands, 1–220, 1975.

Nurmi, P., Perrels, A., and Nurmi, V.: Expected impacts and value of improvements in weather forecasting on the road transport sector, Meteorol. Appl., 20, 217–223, https://doi.org/10.1002/met.1399, 2013.

Piriou, J., Redelsperger, J., Geleyn, J., Lafore, J., and Guichard, F.: An Approach for Convective Parameterization with Memory: Separating Microphysics and Transport in Grid-Scale Equations, J. Atmos. Sci., 64, 4127–4139, https://doi.org/10.1175/2007JAS2144.1, 2007.

Prein, A. F., and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, Int. J. Climatol., 37, 305–327, https://doi.org/10.1002/joc.4706, 2017.

Roebber, P.: Visualizing Multiple Measures of Forecast Quality, Weather Forecast., 24, 601–608, https://doi.org/10.1175/2008WAF2222159.1, 2009.

Räisänen, J.: Twenty-first century changes in snowfall climate in Northern Europe in ENSEMBLES regional climate models, Clim. Dynam., 46, 339–353, https://doi.org/10.1007/s00382-015-2587-0, 2016.

Salli, R., Lintusaari, M., Tiikkaja, H., and Pöllänen, M.: Wintertime road conditions and accident risks in passenger car traffic, Tampere University of Technology, Department of Business Information Management and Logistics, Transportation Systems, Tampere, Finland, Research Report 68, 70 pp., 2008.

Screen, J.: Arctic amplification decreases temperature variance in northern mid- to high-latitudes, Nat. Clim. Change, 4, 577–582. https://doi.org/10.1038/nclimate2268, 2014.

Vaisala: Vaisala Road and Runway Surface and Depth Sensor DRS511, available at: https://www.vaisala.com/en/products/devices/weather-stations-and-instruments/drs511, last access: 18 December 2018, 2018a.

Vaisala: Vaisala Remote Surface State Sensor DSC111, available at: https://www.vaisala.com/en/products/instruments-sensors-and-other-measurement-devices/weather-stations-and-sensors/dsc111, last access: 18 December 2018, 2018b.

Vajda, A., Tuomenvirta, H., Juga, I., Nurmi, P., Jokinen, P., and Rauhala, J.: Severe weather affecting European transport systems: the identification, classification and frequencies of events, Nat. Hazards, 72, 169–188, https://doi.org/10.1007/s11069-013-0895-4, 2014.

World Meteorological Organization (WMO): Guide to meteorological instruments and methods of observation, 7th ed., Geneva, Switzerland, WMO-No. 8, available at: https://library.wmo.int/pmb_ged/wmo_8_en-2012.pdf, last access: 6 June 2019, 2008.

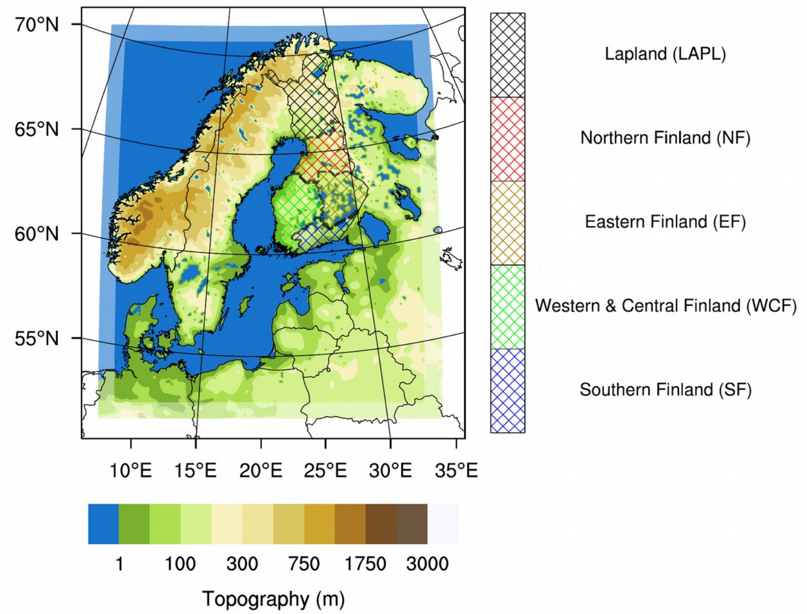

Figure 1: The HCLIM38-ALARO model domain and topography at 12.5 km x 12.5 km grid resolution. Colored overlays depict the regions that are evaluated in more detail. The transparent areas depict the model's 8-point wide relaxation zone.

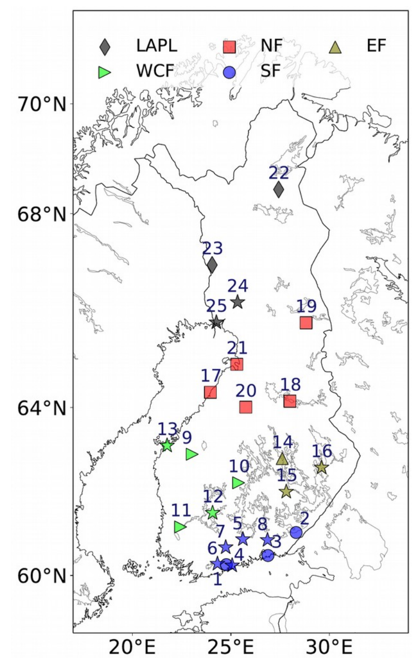

Figure 2: Locations of road weather stations used in this study. The numbers refer to Table 1. The stations with an additional optical sensor are marked as stars. SF stands for Southern Finland, WCF for Western and Central Finland, EF for Eastern Finland, NF for Northern Finland, and LAPL for Lapland.

Table 1. Descriptions of the road weather stations with the mean observed air temperatures (ºC) for the months between October and April in 2002–2014. The stations with an optical sensor are marked with an asterisk (*). The road orientation is defined in the parenthesis. As an example, SE–NW means that the orientation of the road is southeast–northwest. The maintenance classes are described in Appendix A (class 1 means high and class 4 low maintenance).

Region	Number	Station name	Coordinates	Surrounding characteristics and road orientation	Maintenance class	Mean T (ºC) October	Mean T (ºC) November	Mean T (ºC) December	Mean T (ºC) January	Mean T (ºC) February	Mean T (ºC) March	Mean T (ºC) April
Southern Finland	1	Askisto	60.27° N 24.77° E	Open area and a few trees (E–W)	1	5.8	1.6	–2.0	–5.1	–5.5	–2.1	4.4
	2	Lappeenranta	61.07° N 28.31° E	Open area and trees nearby (SW–NE)	2	4.3	0.0	–4.4	–7.6	–7.6	–3.3	3.6
	3	Sutela	60.50° N 26.88° E	Open area and a few trees, river nearby (E–W)	2	5.6	1.1	–2.4	–5.5	–7.0	–2.7	3.8
	4*	Jakomäki	60.25° N 25.06° E	Open area and trees on both sides of the road (SW–NE)	1	6.1	1.8	–1.5	–4.4	–5.2	–1.7	4.4
	5*	Lahti	60.91° N 25.61° E	Open area (SW–NE)	1	4.7	0.8	–3.2	–6.5	–6.6	–2.6	4.0
	6*	Palojärvi	60.29° N 24.32° E	Open area and a few trees, trees on the opposite side (E–W)	1	5.1	1.2	–2.5	–5.5	–5.8	–2.6	3.8
	7*	Riihimäki	60.71° N 24.74° E	Empty lane between the road (SE–NW)	1	4.8	0.4	–3.9	–6.1	–5.9	–2.4	4.2
	8*	Utti	60.89° N 26.86° E	Open area, a few trees, and trees on the opposite side of the road (E–W)	2	4.2	0.3	–3.8	–7.1	–7.1	–2.9	3.7
Western & Central Finland	9	Lapua	62.94° N 23.04° E	Open area and trees on both sides of the road (S–N)	2	3.8	–0.3	–4.0	–6.9	–6.6	–3.0	3.5
	10	Petäjävesi	62.27° N 25.39° E	Open area, a few trees, and trees on the opposite side of the road (E–W)	3	3.5	–0.7	–5.0	–8.2	–8.2	–4.2	2.7
	11*	Seppälänahde	61.21° N 22.45° E	Open area and trees on both sides of the road (SE–NW)	2	4.9	0.9	–2.8	–5.9	–5.7	–2.3	4.0
	12*	Suinula	61.55° N 24.07° E	Open area and trees on both sides of the road (SW–NE)	2	4.3	0.2	–4.0	–7.0	–7.1	–3.5	3.3
	13*	Vaasa	63.14° N 21.76° E	Open area and trees on both sides of the road (SW–NE)	2	4.6	0.3	–3.6	–5.7	–6.6	–3.2	2.9
Eastern Finland	14	Kuopio E	62.84° N 27.61° E	Empty lane between the road (S–N)	1	3.6	–0.7	–5.2	–8.9	–8.8	–4.1	2.8
	15*	Puunkolo	62.06° N 27.81° E	Open area, a few trees, and trees on the opposite side of the road (S–N)	3	3.6	–0.9	–5.6	–8.7	–8.7	–4.3	2.6
	16*	Ylämylly	62.63° N 29.60° E	Open area (SW–NE)	2	3.8	–1.0	–5.7	–9.1	–9.2	–4.5	2.3
Northern Finland	17	Kalajoki	64.34° N 23.96° E	Open area and trees on both sides of the road (SW–NE)	3	4.3	–0.1	–3.8	–7.1	–7.3	–4.1	1.8
	18	Korholanmäki	64.14° N 28.00° E	A few trees and trees on the opposite side of the road (SE–NW)	3	2.3	–2.5	–6.6	–9.6	–9.4	–4.9	2.0
	19	Kuolio	65.83° N 28.82° E	Open area (SW–NE)	4	0.9	–4.4	–8.6	–12.2	–11.6	–7.7	–0.5
	20	Kärsämäki	64.01° N 25.76° E	Open area and trees on the opposite side of the road (S–N)	3	2.7	–1.8	–6.1	–9.4	–9.0	–4.7	2.1
	21	Ouluntulli	64.95° N 25.53° E	Open area and a small hill nearby (SE–NW)	1	3.2	–1.4	–5.6	–9.1	–8.8	–5.1	2.0
Lapland	22	Saariselkä	68.46° N 27.43° E	Open area (SW–NE)	4	–0.5	–6.0	–8.4	–11.2	–11.2	–7.2	–1.4
	23	Sieppijärvi	67.00° N 24.05° E	Open area and trees on both sides of the road (S–N)	4	0.1	–6.7	–9.1	–12.8	–11.9	–6.9	0.5
	24*	Jaatila	66.25° N 25.34° E	Open area and trees on both sides of the road (SW–NE)	3	1.6	–4.0	–7.4	–11.2	–10.6	–6.2	1.0
	25*	Kyläjoki	65.84° N 24.26° E	Open area, at the start of an overpass (E–W)	2	2.6	–2.6	–6.1	–9.8	–9.7	–5.6	0.8

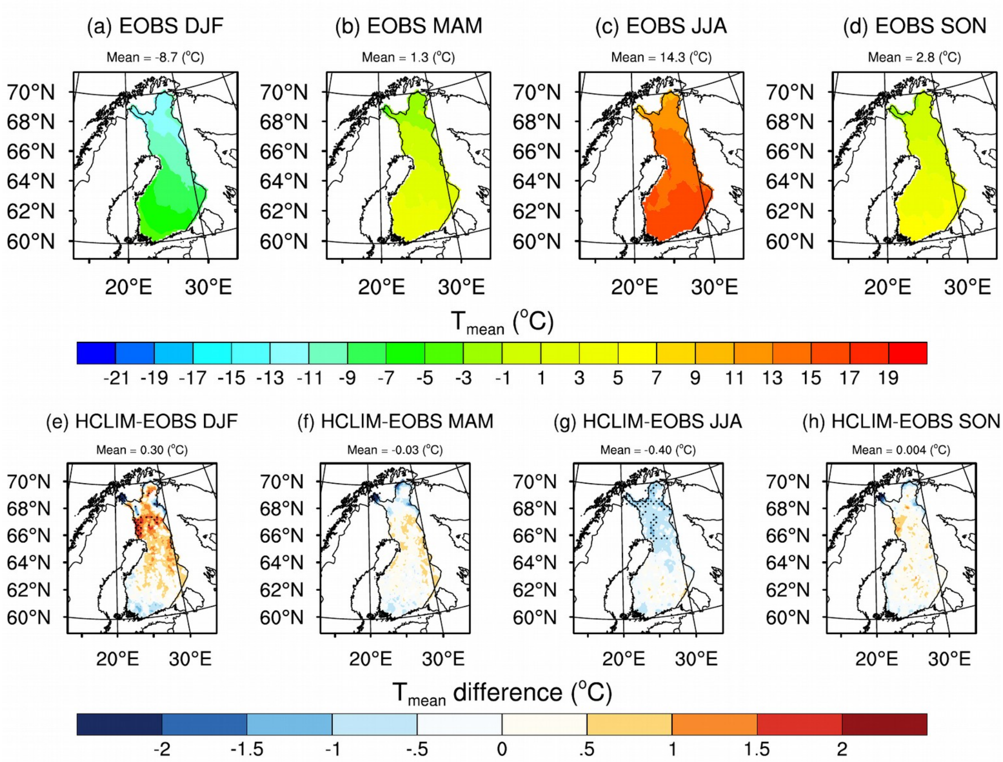

Figure 3: (a–d) The reference values of ~~2 m~~mean near-surface air temperatures (T_{mean}) from E-OBS data~~ (upper row)~~ and (e–h) the biases of HCLIM38-ALARO modeled T_{mean} with a reference to E-OBS~~ (lower row)~~. The seasonal means were calculated over the whole model domain for the time period of January 2002–December 2014. Stippled areas represent statistically significant differences with p values < 0.05.

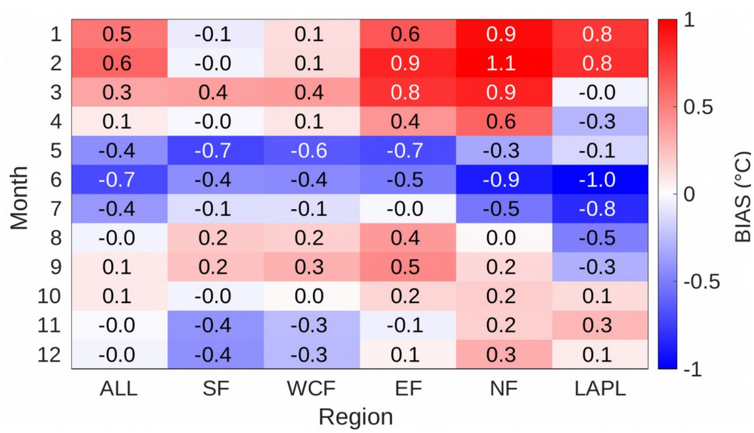

Figure 4: The monthly mean biases of simulated ~~T_{air}~~near-surface air temperature averaged over (a) Southern Finland (SF), (b) Western and Central Finland (WCF), (c) Eastern Finland (EF), (d) Northern Finland (NF), (e) Lapland (LAPL), and (f) the whole Finland (ALL) in 2002–2014 with a reference to~~ the~~ E-OBS~~ dataset~~. ~~ALL refers to the results averaged over the whole Finland, SF to Southern Finland, WCF to Western and Central Finland, EF to Eastern Finland, NF to Northern Finland, and LAPL to Lapland.~~

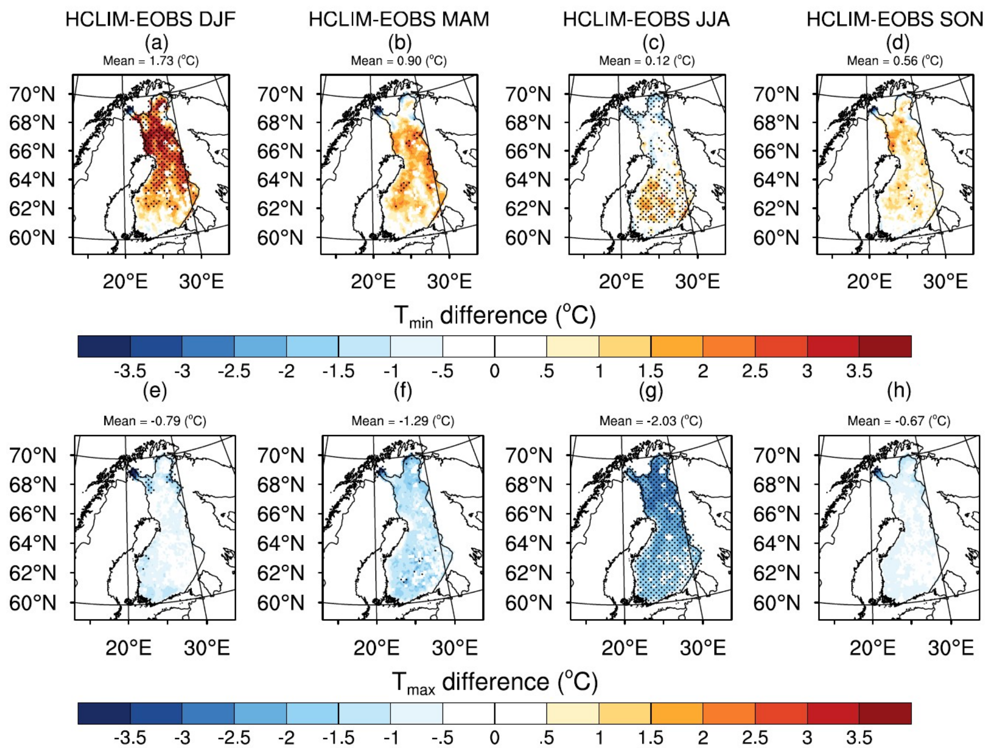

Figure 5: The biases in the simulated seasonal means of (a–d) minimum near-surface air temperature (T_{min}) and (e–h) maximum near-surface air temperature (T_{max}) with a reference to E-OBS. The seasonal mean biases were calculated over Finland for the time period of January 2002–December 2014. Stippled areas represent statistically significant differences with p values < 0.05.

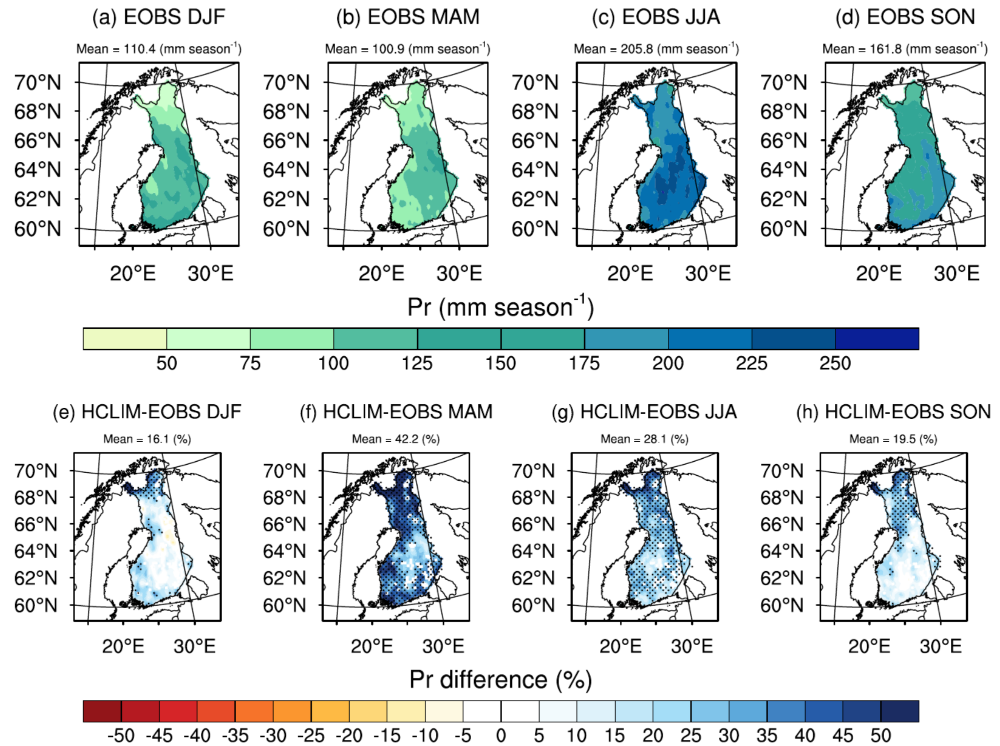

Figure ~~5~~6: (a–d) The reference values of precipitation (*Pr*) from E-OBS data ~~(upper row)~~ and (e–h) the biases of HCLIM38-ALARO modeled ~~precipitation (~~*Pr*~~)~~ with a reference to E-OBS ~~(lower row)~~. The seasonal averages were calculated for the time period of January 2002–December 2014. Stippled areas represent statistically significant differences with *p* values < 0.05.

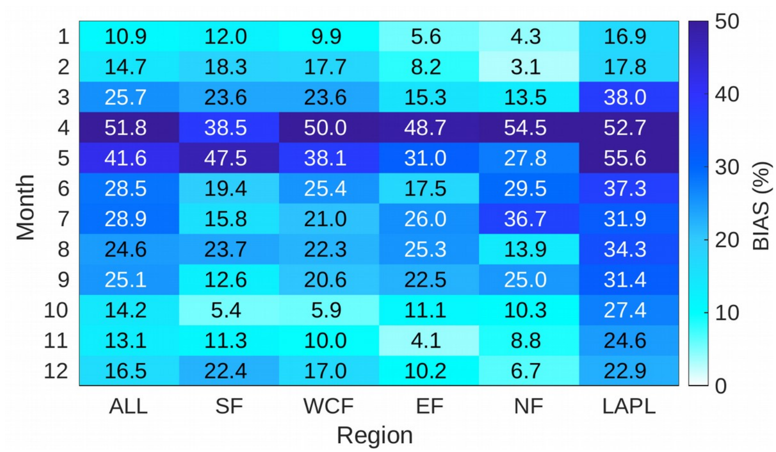

Figure ~~6~~7: The monthly mean biases of simulated precipitation averaged over (a) Southern Finland (SF), (b) Western and Central Finland (WCF), (c) Eastern Finland (EF), (d) Northern Finland (NF), (e) Lapland (LAPL), and (f) the whole Finland (ALL) in 2002–2014 with a reference to ~~the~~ E-OBS ~~dataset~~. ~~ALL refers to the results averaged over the whole Finland, SF to Southern Finland, WCF to Western and Central Finland, EF to Eastern Finland, NF to Northern Finland, and LAPL to Lapland.~~

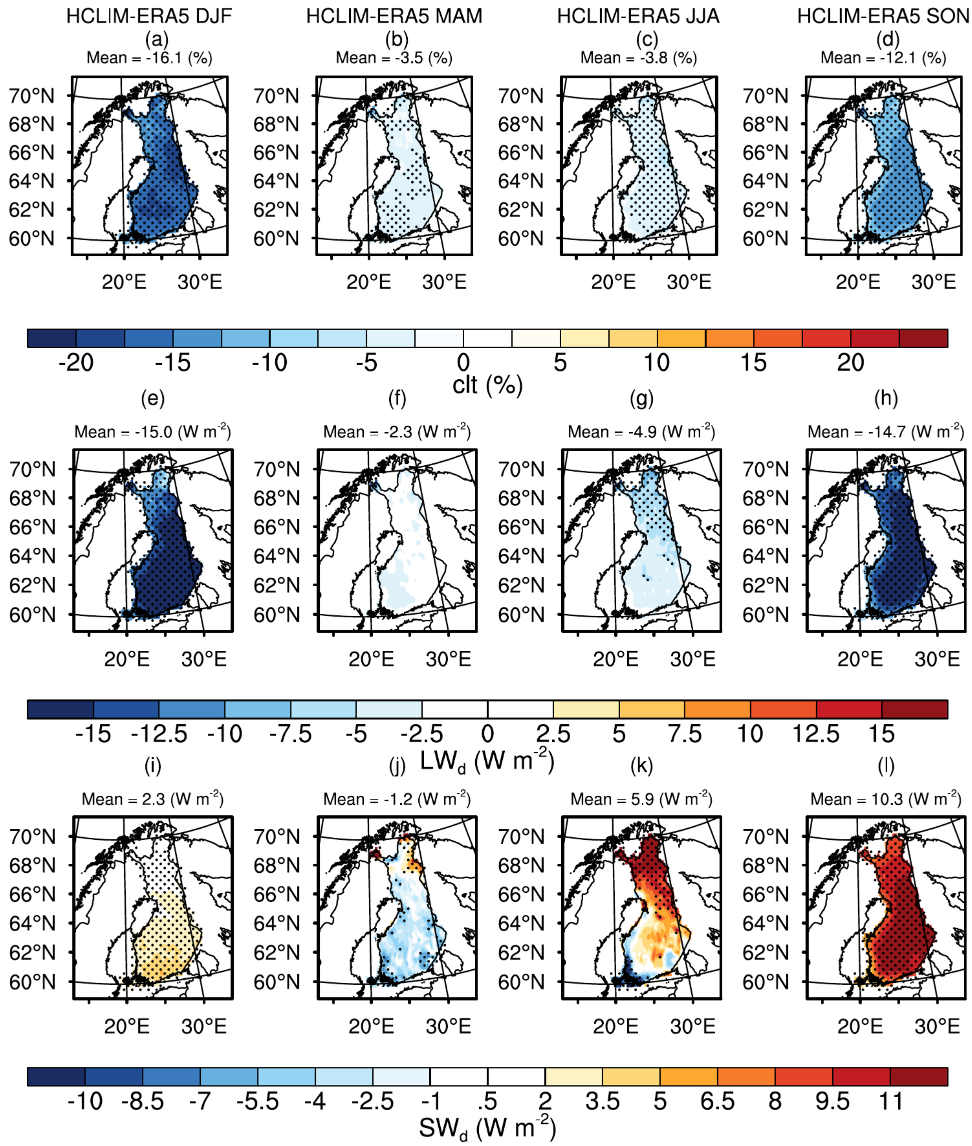

Figure 8: The biases in the simulated seasonal means of (a–d) total cloud fraction (*clt*), (e–h) downwelling longwave and (i–l) shortwave radiation (LW_d and SW_d, respectively) with a reference to ERA5 reanalysis product. The seasonal mean biases were calculated over Finland for the time period of January 2002–December 2014. Stippled areas represent statistically significant differences with p values < 0.05.

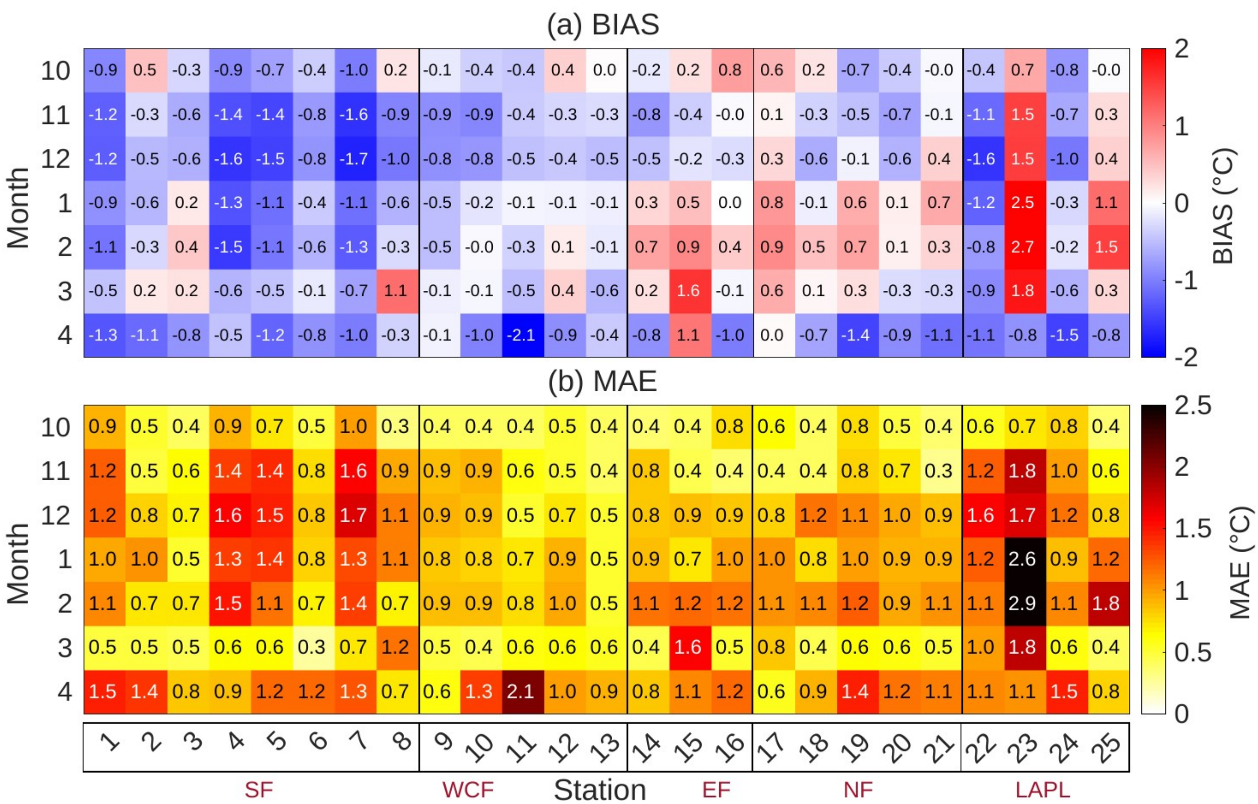

Figure ~~7~~9: (a) The ~~monthly~~ mean monthly biases ~~(upper row), RMSE (middle row),~~ and *~~R~~MAE* values ~~(lower row)~~ of simulated ~~T_{road}~~road surface temperature ~~from~~between October ~~to~~and April in 2002–2014. The station indices on the x-axis refer to Table 1. SF refers to Southern Finland, WCF to Western and Central Finland, EF to Eastern Finland, NF to Northern Finland, and LAPL to Lapland.

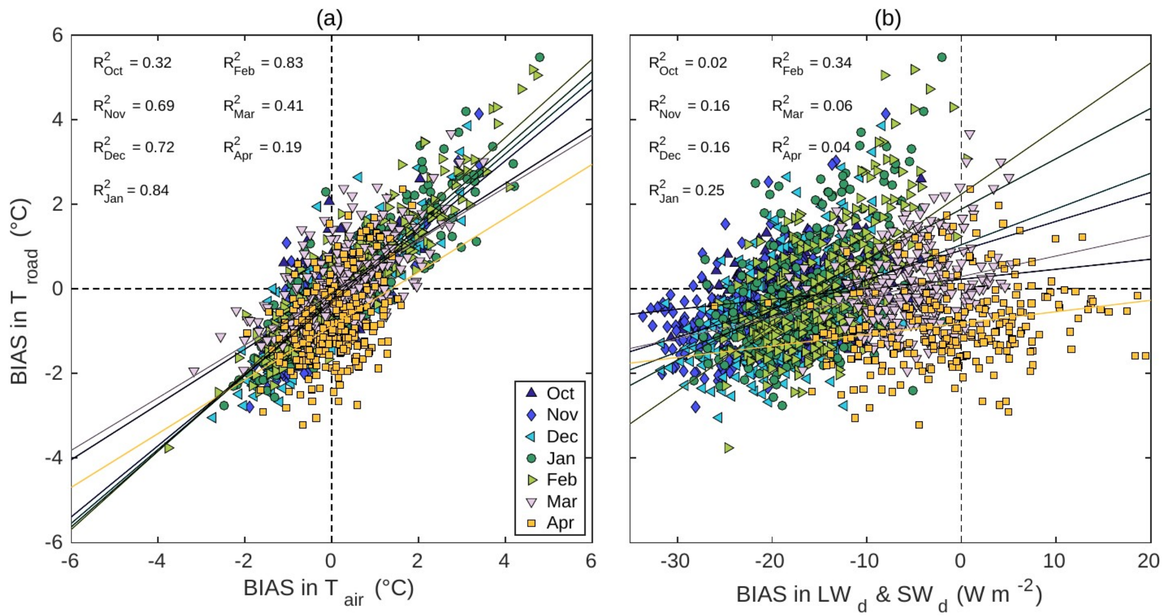

Figure 10: Scatter plots of the mean monthly biases of road surface temperature (T_{road}) against (a) the mean monthly biases of near-surface air temperature (T_{air}) and (b) the mean monthly biases of downwelling longwave (LW_d for October–March) and shortwave radiation (SW_d for April) at the road weather stations. The squared R values represent linear regression for different months with p values < 0.001 (p value for LW_d in October 0.01).

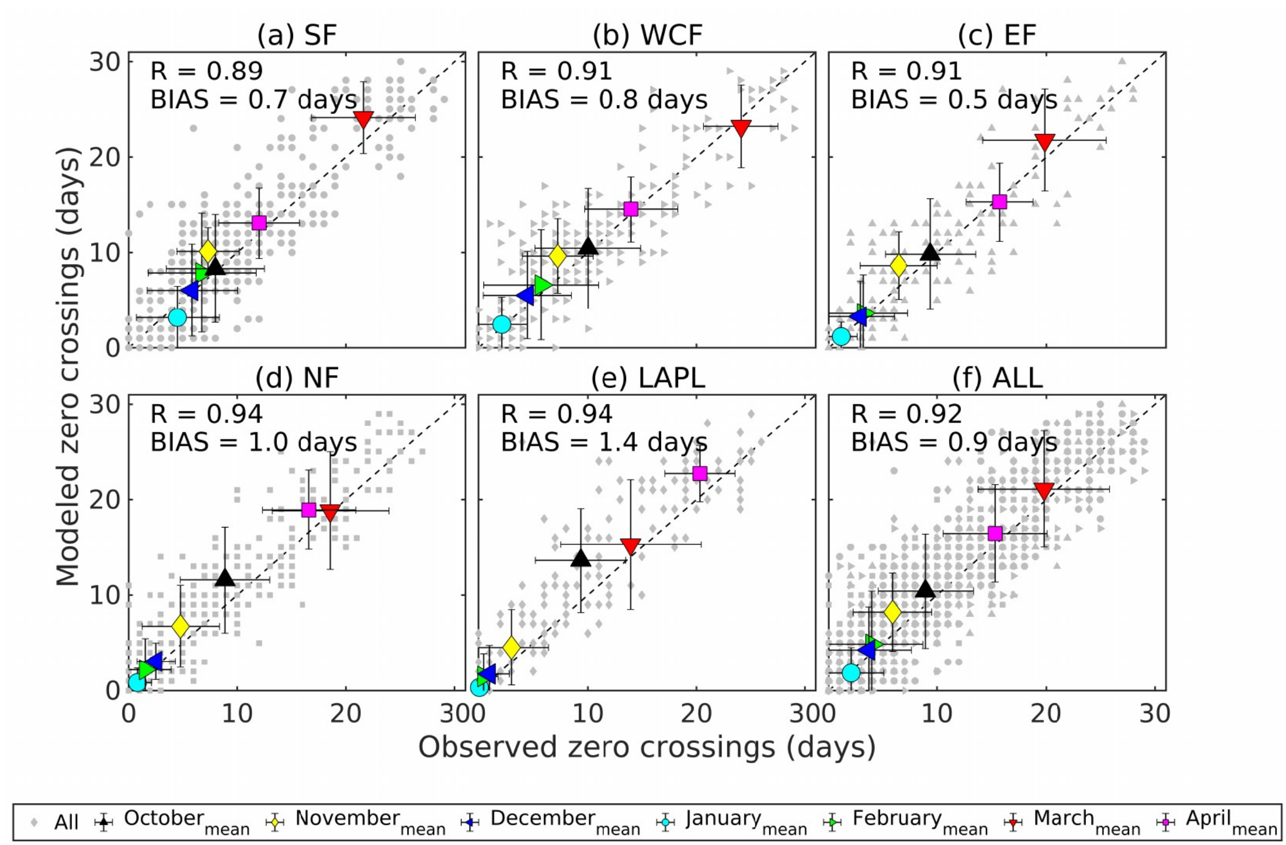

Figure 811: Modeled vs. observed days per month when road temperatures had been both below –0.5 ºC and above 0.5 ºC (zero crossing day) during October and April in 2002–2014 in (a) Southern Finland (SF), (b) Western and Central Finland (WCF), (c) Eastern Finland (EF), (d) Northern Finland (NF), (e) Lapland (LAPL), and (f)~~the averages for~~ the whole Finland (ALL). Grey color represents the monthly values for every year and the multi-year monthly means are illustrated in other colors. The vertical and horizontal bars represent ±1 standard deviation based on 13 years of monthly values from the model and observations, respectively. *R* stands for the Pearson correlation coefficient and BIAS for the mean difference between the modeled and observed values. The dashed black line represents a 1:1 reference line.

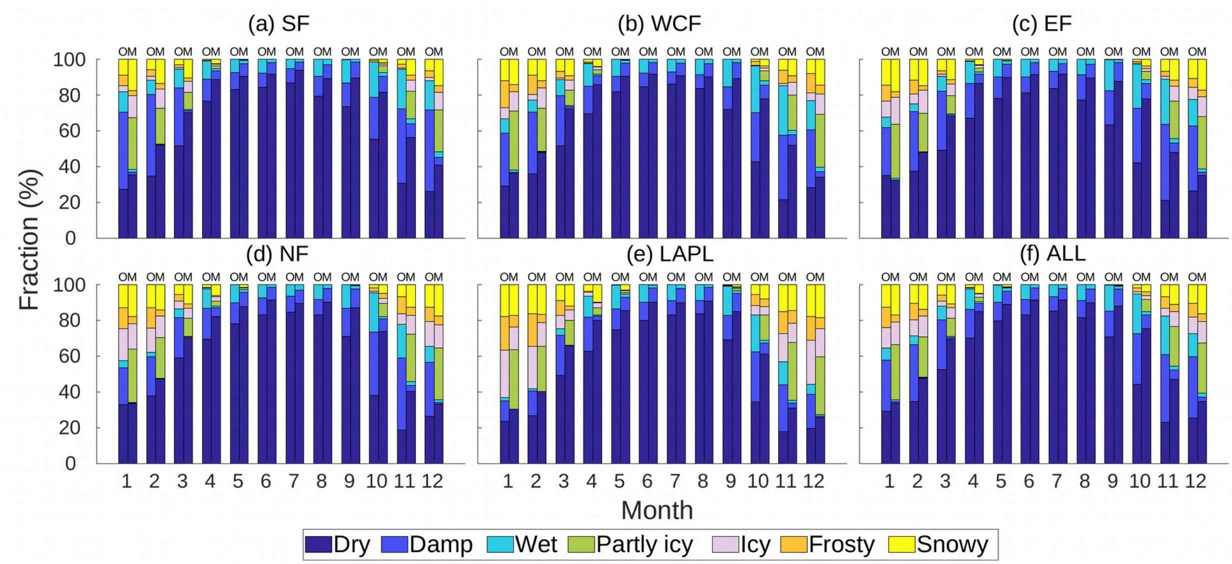

Figure ~~9~~12: Observed (O) and modeled (M) ~~mean daily~~ fractions of road surface classes (e.g. dry, wet, or icy) within each month in 2002–2014 in (a) Southern Finland (SF), (b) Western and Central Finland (WCF), (c) Eastern Finland (EF), (d) Northern Finland (NF), (e) Lapland (LAPL), and (f) the averages for whole Finland (ALL). The definitions of road surface classes differ slightly for the observations and model (e.g. ~~T~~the partly icy class is ~~defined~~included only in the model).

~~**Table 2.** The contingency table.~~

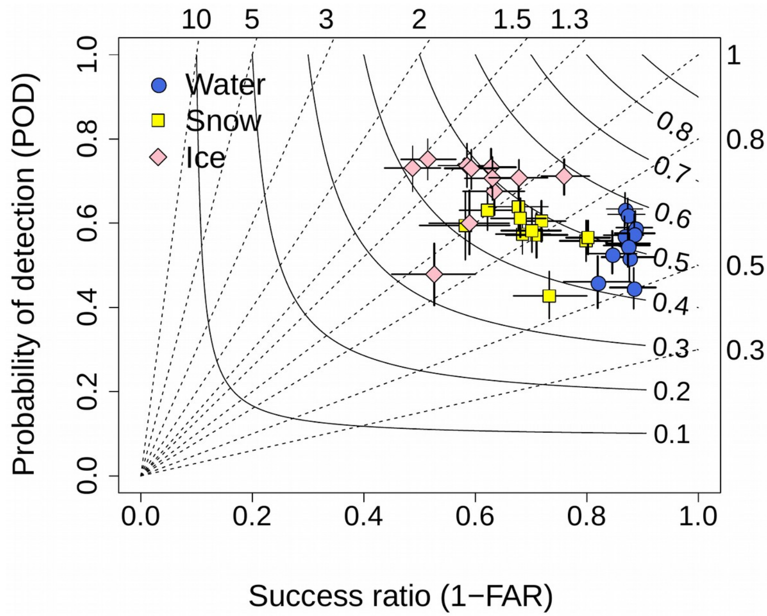

Figure 1~~0~~3: The performance diagram of water, snow, and ice storages modeled ~~for~~at the 11 road weather stations which have an optical sensor (see Table 1). Absolute values of the modeled and observed m~~ean~~aximum daily storages were not used directly, but instead, the daily value was set to one if the m~~ean~~aximum value was more than zero and to zero if the m~~ean~~aximum value was zero. The months between October and April were included in the analysis. Success ratio (1–FAR) runs along the x-axis and POD along the y-axis. Dashed lines represent the frequency bias and continuous lines the CSI. The vertical and horizontal lines represent the 95 % confidence intervals for POD and FAR values, respectively, calculated by using a bootstrap method and 1000 resamples.