Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

The road weather model RoadSurf driven by the HARMONIE-Climate regional climate model: 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 a high-resolution regional climate model (RCM) HARMONIE-Climate (HCLIM) utilizing ALARO physics (HCLIM-ALARO). 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 bias of –0.3 ºC, RMSE of 2.1 ºC, and Pearson's correlation coefficient of 0.93. The RoadSurf's output bias most probably stemmed from the lack 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 HCLIM-ALARO overestimated precipitation and had a warm bias in simulated air temperatures during the winter season. In turn, these input data biases seemed to result in a warm bias in simulated road surface temperatures. Furthermore, the lack of road maintenance operations in the model might have affected RoadSurf's ability to simulate road surface conditions: 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 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 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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

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 temperatures 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 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 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 conditions are the major cause for the wintertime and weather-related road accidents in Fenno-Scandia (Andersson and Chapman, 2011b; 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. This evaluation is needed in order to build and study future scenarios of road weather in this area with a larger confidence. 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 simulations are evaluated against a standard meteorological dataset, E-OBS, over Fenno-Scandia.

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 of near-surface air temperature and precipitation. This comparison is followed by an evaluation of 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 road weather station's local features, such as location, surrounding characteristics, and road maintenance class, on the model biases.

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, as mentioned before, and a hydrostatic version of the dynamical core. The HCLIM-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 Lindsted et al. (2015).

For this study, HCLIM-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 HCLIM-ALARO simulated domain as well as the regions of Finland that are analyzed in more detail in this study. The lateral boundary conditions of HCLIM-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 HCLIM-ALARO output parameters were produced every full hour.

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. The model takes into account the conditions at the road surface and beneath it, and calculates the vertical heat transfer in 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 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 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}), relative humidity (RH), wind speed (WS), precipitation (Pr) as well as incoming shortwave (SW) and longwave (LW) 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. Instead, RoadSurf was modified so that it utilizes the RCM data, in this case, the output of reanalysis-driven HCLIM-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 HCLIM-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 Gridded daily precipitation and temperature dataset

The HCLIM-ALARO simulated daily precipitation and near-surface air temperatures were compared with the E-OBS dataset (Haylock et al., 2008) which consists of daily precipitation and 2 m air temperature data retrieved from stations located in Europe. The data is available as an interpolated grid which covers the pan-European domain with a resolution of 0.25° (approximately 27.5 km). 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-

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

shielded gauges) for snow (Goodison et al., 1998). Moreover, the accuracy of the E-OBS dataset depends on the number of stations used in the interpolation process: The sparse station density can introduce some errors into the interpolated dataset (e.g. Prein and Gobiet, 2017). 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 HCLIM-ALARO model results for the whole simulated domain covering Fenno-Scandia were thus compared with E-OBS by remapping the modeled values into the E-OBS grid: 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, 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 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 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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

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 HCLIM-ALARO

3.1.1 Near-surface air temperature

The HCLIM-ALARO model accurately captured the seasonal 2 m air temperatures (T_{air}) over the Fenno-Scandian domain between 2002 and 2014. This is confirmed by Fig. 3 which illustrates the multi-year mean seasonal T_{air} from E-OBS as well as the mean biases in the HCLIM-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 were negative in all seasons of which the summer season (June–August) had the smallest domain-averaged bias of –0.25 ºC and the spring season (March–May) the highest domain-averaged bias of –0.68 ºC. The biases were statistically significant mainly over the mountainous areas in Norway where the model had an enhanced cold bias. This error 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, where 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. Lindstedt et al. (2015) encountered similar warm bias in their HCLIM-ALARO simulations 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 the possible biases in the input parameters (ERA-Interim) or from SURFEX's own features. 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 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. However, some regional differences were apparent. For example, in Southern Finland, the biases were mainly negative during the autumn and winter months. 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 biases in Southern Finland. In Eastern Finland, the mean biases resembled Western and Central Finland but were slightly higher for every month except for July,

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

Geoscientific Model Development Discussions

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 summer.

3.1.2 Precipitation

Also multi-year mean seasonal precipitation sums were reliably simulated by HCLIM-ALARO although slight overestimation was evident. Figure 5 depicts both observed multi-year mean seasonal precipitation sums from E-OBS dataset over the model domain in 2002–2014 as well as the differences between HCLIM-ALARO with a reference to E-OBS. Similarly than in 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 (September–October) with a domain-averaged bias of 12.7 % and highest in spring (March–May) with a domain-averaged bias of 31.9 %. The highest biases in simulated precipitation occurred in the Norwegian mountains where the biases were also statistically significant for every season. We stress that E-OBS might suffer from undercatch errors during the winter and spring as well as over the mountainous areas, which may penalize the model in the areas with the most complex topography. The biases were statistically significant over the whole model domain during the summer season. 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 overall overestimation of spring and summertime precipitation in HCLIM-ALARO might be due to too frequent low and moderate intensity precipitation events as Lindstead et al. (2015) pointed out in their study.

Figure 6 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. The mean monthly biases between the regions did not differ substantially from each other. However, the biases were the smallest in the southern parts of Finland during most of the months and, consistently, the largest in the northern parts of Finland. As already seen in Fig. 5, 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).

3.2 Evaluation of RoadSurf-HCLIM

3.2.1 Road surface temperature

The meteorological data from HCLIM-ALARO was used as an input to RoadSurf which 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 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 mean biases, root-mean-square-errors (RMSE) as well as Pearson's correlation coefficients (R) using the average daily road surface temperature values. It is good to keep in mind that the daily and hourly time 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). However, calculating monthly statistics of the above-mentioned metrics using daily data gives us a clear understanding of the model performance during different months during the study period from 2002 to 2014.

Figure 7 makes evident that HCLIM-driven RoadSurf was able to simulate T_{road} with a high accuracy. The mean monthly bias at all 25 stations was –0.3 ºC (range –2.1–2.8 ºC), the average monthly RMSE 2.1 ºC (range 1–4.6 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 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 biases occurred in March and October whereas negative biases occurred in April, November, and December. Eleven stations had negative bias throughout all the analyzed months while the rest of the stations had both negative and positive 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. The highest RMSE values occurred in Lapland 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.

Two probable reasons for the seasonal and regional differences in the model performance are (1) the biases in the HCLIM-ALARO data (mainly T_{air} and precipitation), and (2) the fact that RoadSurf works well in the vicinity of 0 degrees. For example, the comparison of the simulated and observed T_{air} in the wintertime (December–February) revealed a warm bias ranging from 0.2 to 1 ºC in the northern parts of Finland (Northern Finland and Lapland) while Southern Finland had negative biases ranging between –0.5 and –0.1 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 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 HCLIM-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). This good model performance near 0 ºC could partly explain why the RMSE values were actually lower in October compared to 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.

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 and RMSE values between 0.3 and 1.9 ºC at 20 stations in Finland during October and December 2013. 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, one reason for the slightly larger errors obtained in the present study might be the coarser grid resolution of HCLIM-ALARO: Coarser grid resolution implies that not all the local features, such as elevation, are described as in detail as they are in higher resolution NWP models. Increasing the grid resolution of HCLIM-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).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

3.2.3 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 shows that the monthly amount of zero crossing days and the monthly variation (standard deviation) were 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). 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. 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 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 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 surface class occurred within a month.

Figure 9 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 hours) and underestimated damp surfaces slightly more (average bias –4.2 hours). The model underestimated also wet surfaces (average bias –2.3 hours), but the hours accumulated in the partly icy class (2.7 hours on average) were 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 hours) and overestimation of ice (0.4 hours) were also of the similar magnitude with opposite signs. Moreover, the snow class was slightly overestimated with an average bias of 0.6 hours. These results are in line with the study by Kangas et al. (2015)

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

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 to underestimate frost with the same magnitude than ice is overestimated) is acceptable.

The lack of road maintenance could be one logical reason why the model overestimates 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 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 real life, salting is not performed as often at the stations in Lapland as in Southern Finland and thus icy roads can occur more often in the northmost stations. Furthermore, the model takes 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. 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.

3.2.5 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. 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 values of the storages between October and April and, further, setting the daily values to one if the mean value was more than zero and to zero if the mean value was zero. These binary values were used to calculate hits and false alarms (Table 2) and the probability of detection (POD; Eq. (1)) and false alarm ratios (FAR; Eq. (2)) (Roebber, 2009). 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 might be slightly displaced or mistimed. For this reason, the results should be interpreted with care and should be taken as qualitative.

$$POD=\frac{a}{a+c} \tag{1}$$

$$FAR=\frac{b}{a+b} \tag{2}$$

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

The results of the POD-FAR analysis for 11 stations including an optical sensor (see Table 1) are illustrated in Fig. 10 using a categorical performance diagram (Roebber, 2009). 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.

$$Frequency\,bias = \frac{a+b}{a+c} \tag{3}$$

$$CSI = \frac{a}{a+b+c} \tag{4}$$

Figure 10 shows that RoadSurf reliably captured the occurrence of the storage terms as the points locate 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 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. Another interesting fact is that the lack 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. 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 (more 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.

4 Conclusions

This study described the performance of the HCLIM-ALARO regional climate model over Fenno-Scandia and, further, evaluated the skill of HCLIM-ALARO-driven road weather model RoadSurf to reproduce the present-day road weather conditions in Finland. This study showed that HCLIM-ALARO is in good agreement with the gridded air temperature and precipitation observations: The model reliably produced the seasonal and monthly temporal and spatial patterns over Fenno-Scandia and Finland. Especially near-air temperatures were well represented by HCLIM-ALARO. On the other hand, the 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 undercath errors and lower station density in the northern parts of the modeled domain.

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 HCLIM-ALARO-driven RoadSurf was able to accurately simulate road surface temperatures (T_{road}) with the mean bias of –0.3 ºC, RMSE of 2.1 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, the coarser grid resolution of the HCLIM-ALARO compared to the NWP model input used in the earlier studies might be the main reason for this outcome. Moreover, the HCLIM-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 better performance of RoadSurf to simulate T_{road} at the stations located in Southern Finland compared to the stations located in Lapland, also confirmed by the Kruskal–Wallis test.

In addition, RoadSurf 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 slippery 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 the observations showed. The lack 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 HCLIM-ALARO realistically captured the Fenno-Scandian climate 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, 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 HCLIM-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.

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 (lse):

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 (ls):

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 (lb):

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 HCLIM-ALARO simulations with the help of JPP. JPP did the offline coupling of RoadSurf and HCLIM-ALARO. ET planned and performed the analysis of HCLIM 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) and the data providers in the ECA&D project (http://www.ecad.eu). We are grateful for the help provided by Ari Aaltonen in retrieving the Finnish road weather observations.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

Geoscientific Model Development Discussions
Open Access

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.

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.

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.

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.

Liikennevirasto (Finnish Transport Agency): Teiden talvihoito (The maintenance of the roads during winter), available at: https://www.liikennevirasto.fi/tieverkko/kunnossapito/talvihoito#.Wyt4LRwlGCg, last access: 18 December 2018.

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.

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, 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.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

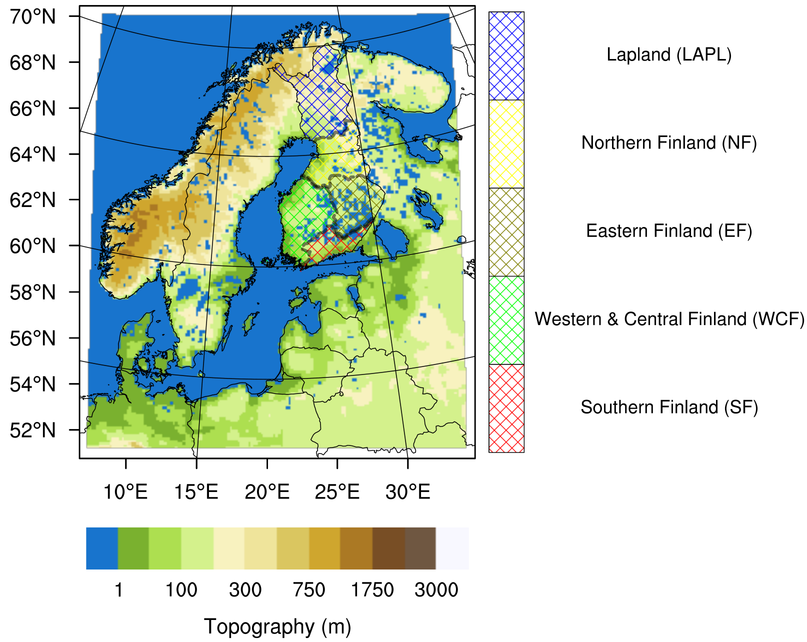

Figure 1: The HCLIM-ALARO model domain and topography at 12.5 km x 12.5 km grid resolution. Colored overlays depict the regions that are evaluated in more detail.

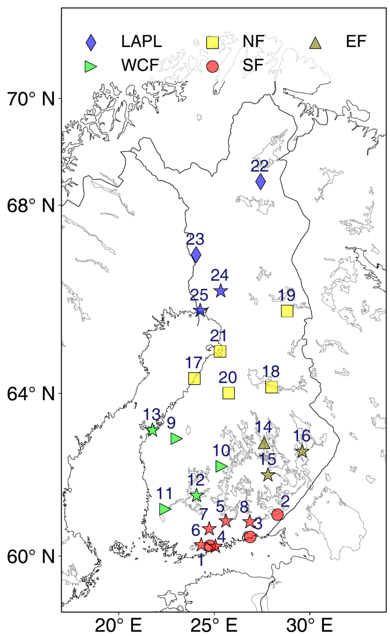

Figure 2: Locations of road weather stations used in this study. The numbers refer to Table 1. The stations with an additional optical sensor are marked as stars. SF stands for Southern Finland, WCF for Western and Central Finland, EF for Eastern Finland, NF for Northern Finland, and LAPL for Lapland.

Geoscientific Model Development Discussions Open Access EGU

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

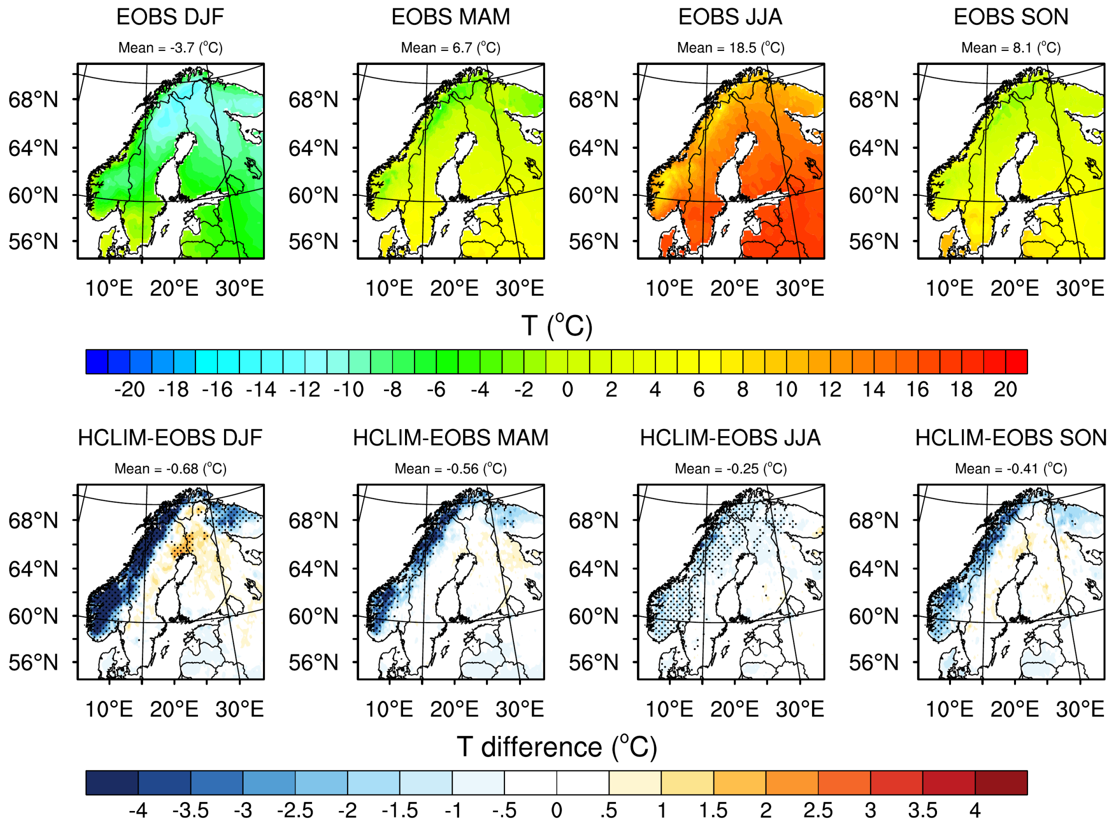

Figure 3: The reference values of 2 m air temperatures (T) from E-OBS data (upper row) and the biases of HCLIM-ALARO modeled T with a reference to E-OBS (lower row). The seasonal means were calculated over the whole model domain for the time period of January 2002–December 2014. Stippled areas represent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

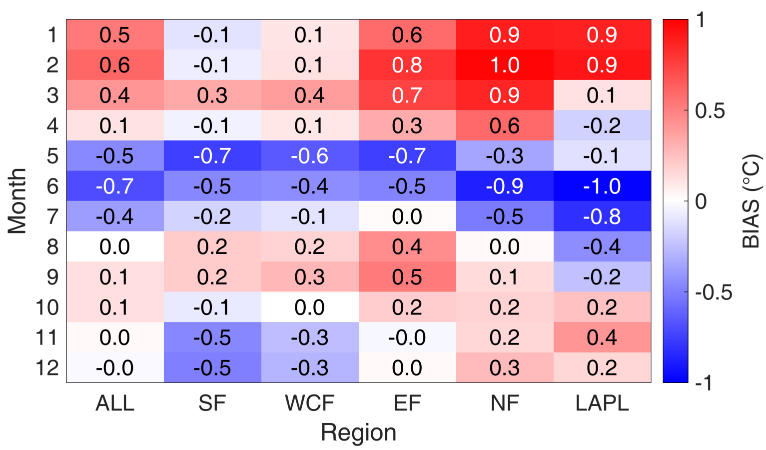

Figure 4: The monthly mean biases of simulated T_{air} in 2002–2014 with a reference to the E-OBS dataset. ALL refers to the results averaged over the whole Finland, SF to Southern Finland, WCF to Western and Central Finland, EF to Eastern Finland, NF to Northern Finland, and LAPL to Lapland.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

Geoscientific Model Development
Discussions

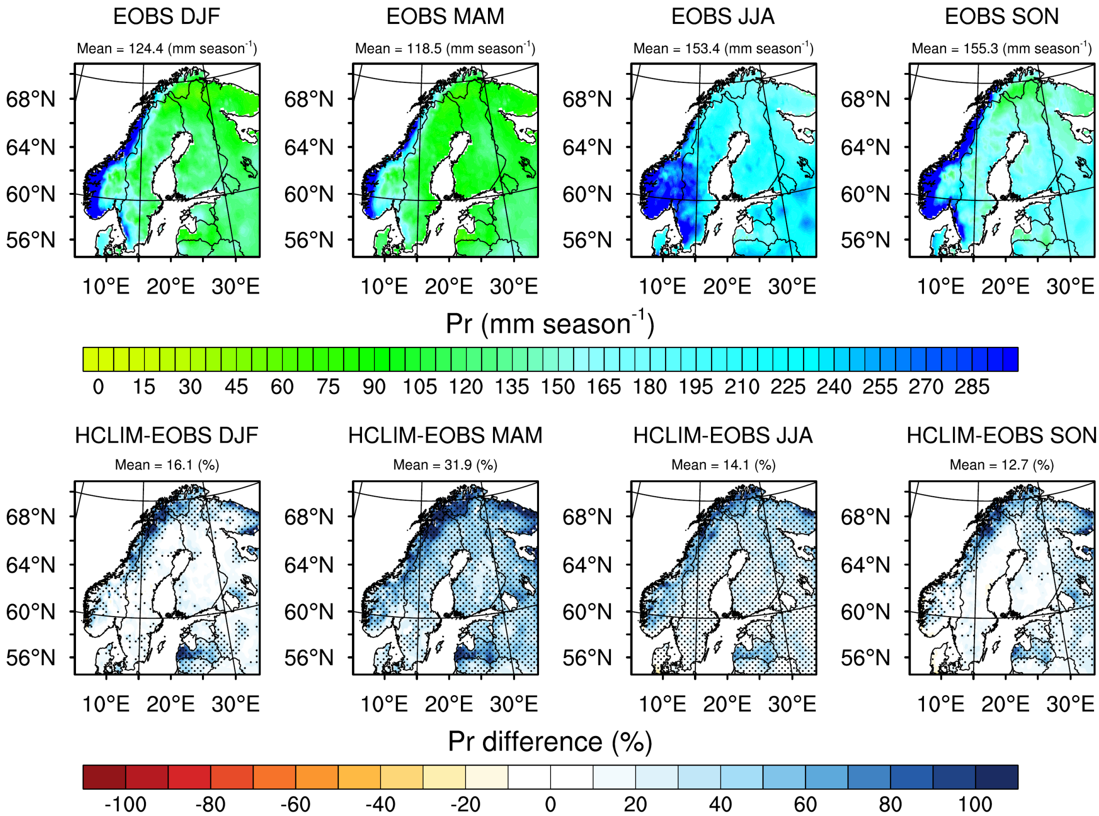

Figure 5: The reference values of precipitation from E-OBS data (upper row) and the biases of HCLIM-ALARO modeled precipitation (*Pr*) with a reference to E-OBS (lower row). The seasonal averages were calculated for the time period of January 2002–December 2014. Stippled areas represent statistically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

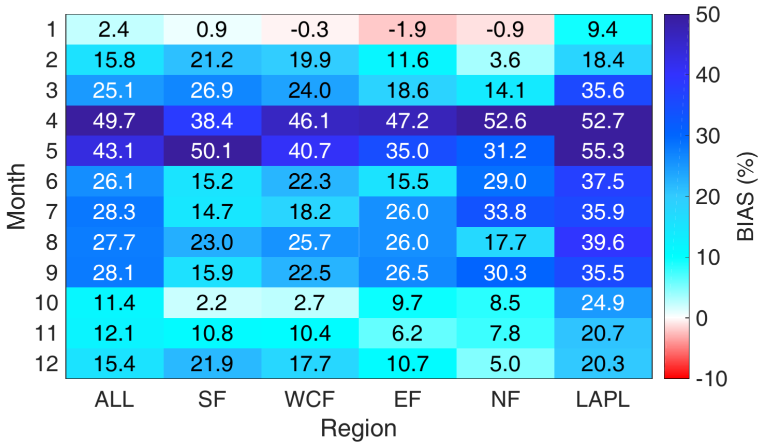

Figure 6: The monthly mean biases of simulated precipitation in 2002–2014 with a reference to the E-OBS dataset. ALL refers to the results averaged over the whole Finland, SF to Southern Finland, WCF to Western and Central Finland, EF to Eastern Finland, NF to Northern Finland, and LAPL to Lapland.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

Geoscientific Model Development
Discussions
Open Access

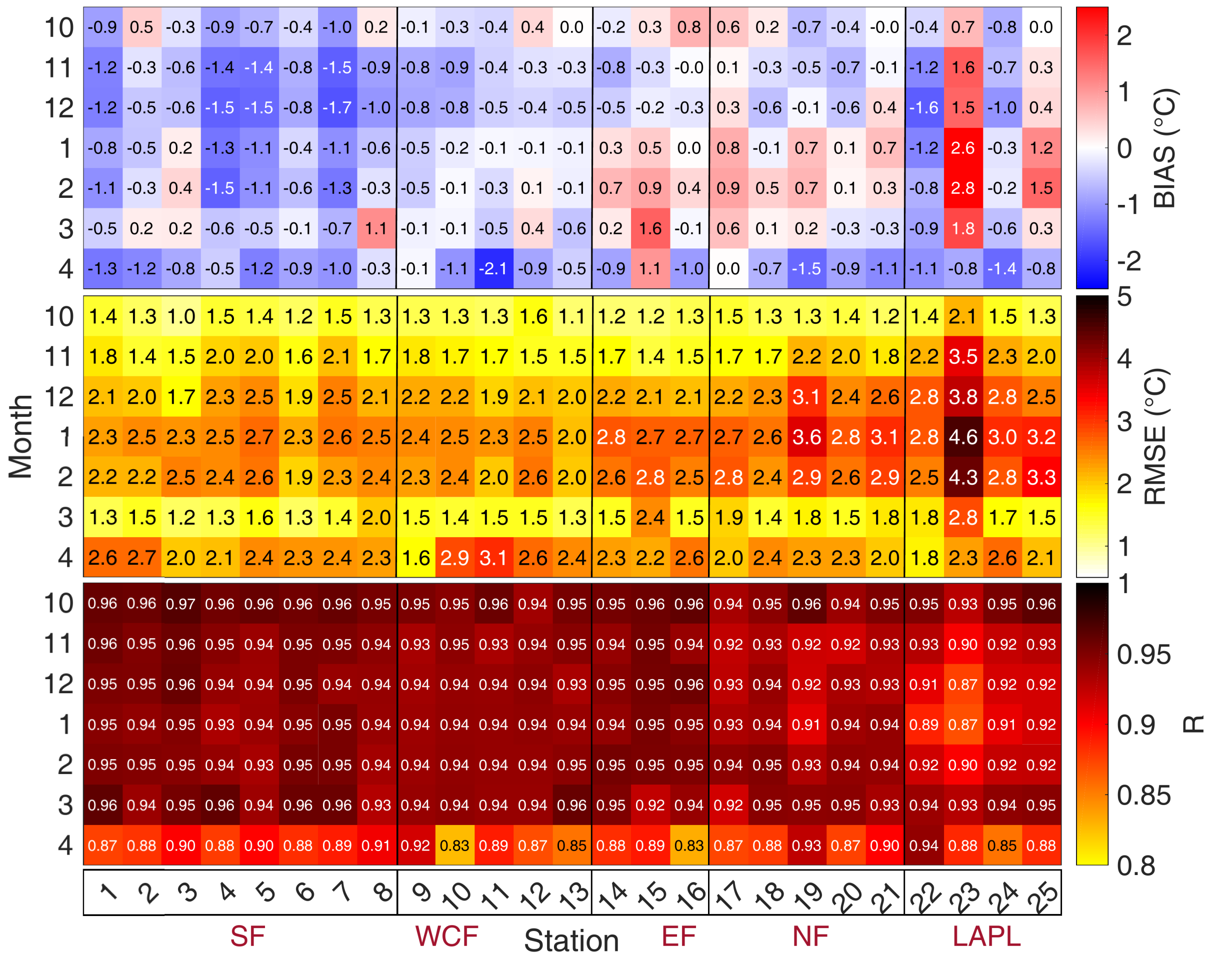

Figure 7: The monthly mean biases (upper row), RMSE (middle row), and R values (lower row) of simulated T_{road} from October to April in 2002–2014. The station indices on the x-axis refer to Table 1. SF refers to Southern Finland, WCF to Western and Central Finland, EF to Eastern Finland, NF to Northern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

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

Geoscientific Model Development
Discussions

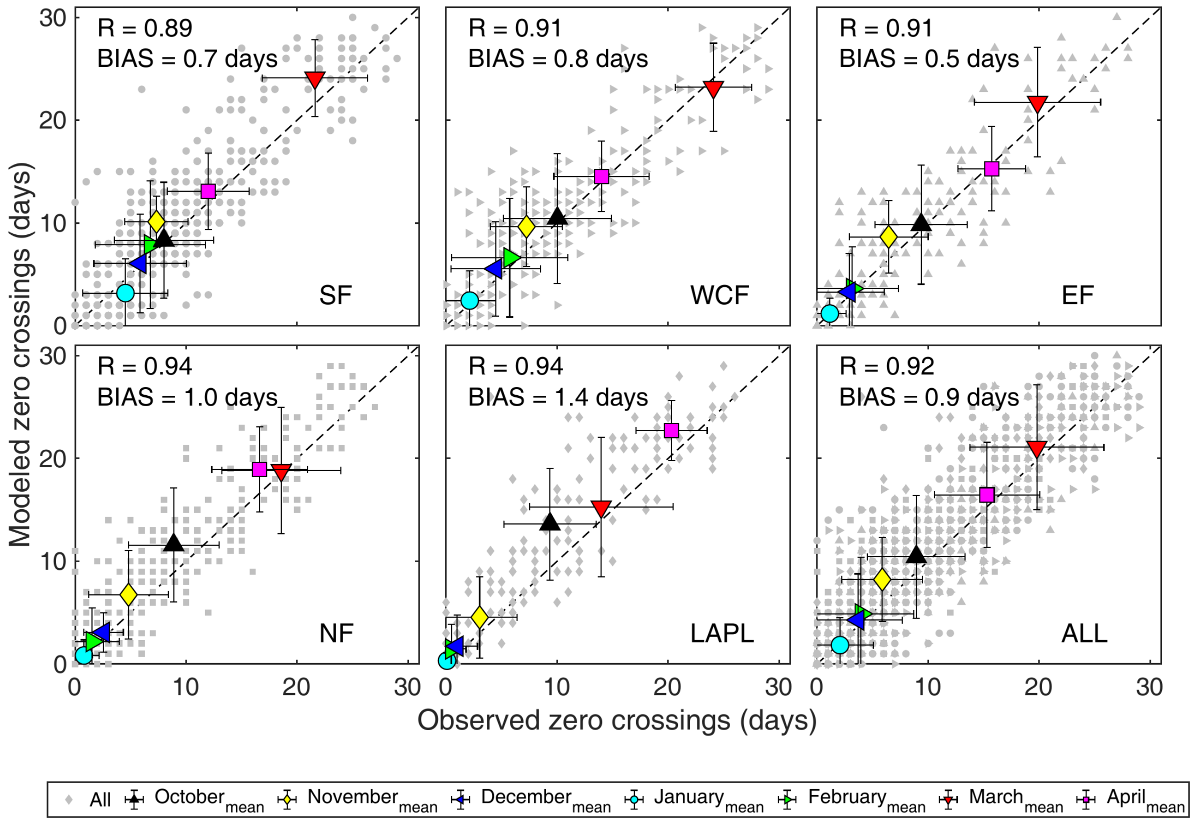

Figure 8: Modeled vs. observed days per month when road temperature had been below –0.5 ºC and above 0.5 ºC (zero crossing day) during October and April in 2002–2014 in Southern Finland (SF), Western and Central Finland (WCF), Eastern Finland (EF), Northern Finland (NF), Lapland (LAPL), and the averages for 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.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

Geoscientific Model Development Discussions
Open Access

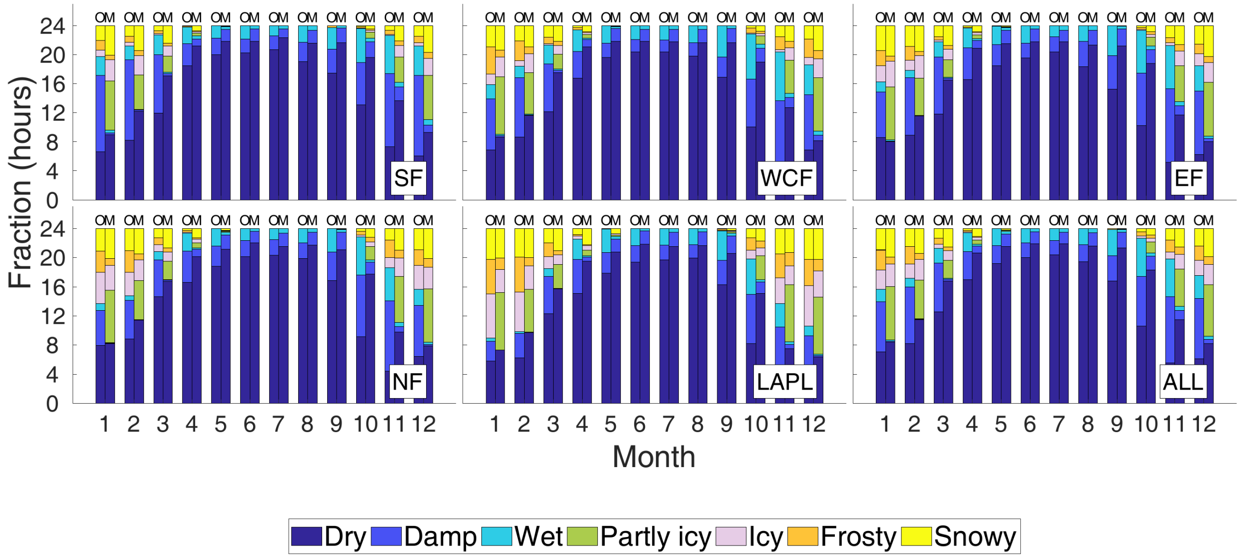

Figure 9: Observed (O) and modeled (M) mean daily fractions of road surface classes (e.g. dry, wet, or icy) within each month in 2002–2014 in Southern Finland (SF), Western and Central Finland (WCF), Eastern Finland (EF), Northern Finland (NF), Lapland (LAPL), and the averages for whole Finland (ALL).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

Table 2. The contingency table.

		Event observed	
		Yes	No
Event modeled	Yes	a	b
	No	c	d

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-330
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 1 February 2019

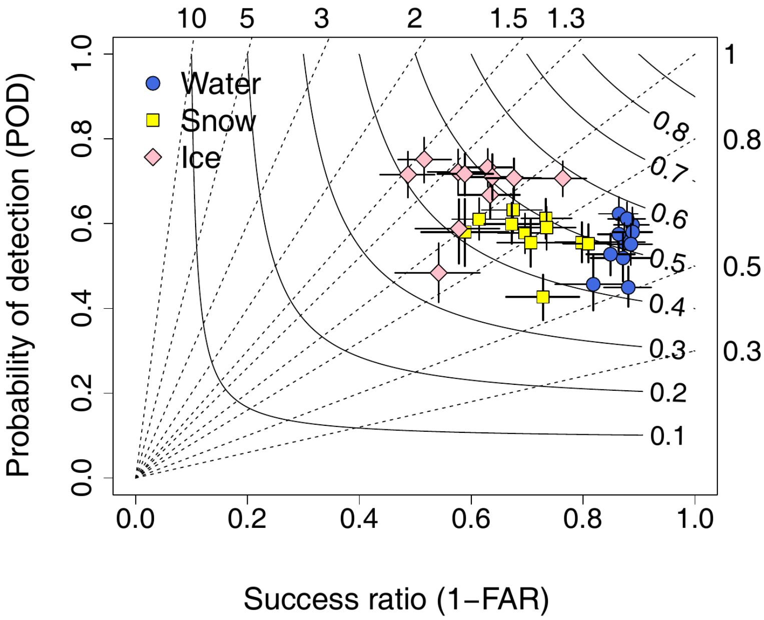

Figure 10: The performance diagram of water, snow, and ice storages modeled for the 11 road weather stations which have an optical sensor (see Table 1). Absolute values of the modeled and observed mean daily storages were not used directly, but instead, the daily value was set to one if the mean value was more than zero and to zero if the mean 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.